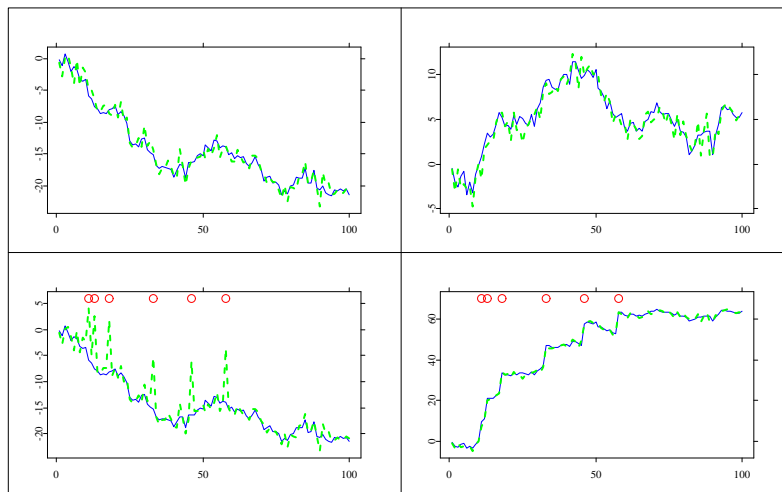


Ansätze zur Robustifizierung des Kalman-Filters



An der Universität Bayreuth
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte Abhandlung

vorgelegt von

Peter Ruckdeschel
geboren am 14. März 1970 in Kulmbach

1. Gutachter: Prof. Dr. Helmut Rieder,
Universität Bayreuth
2. Gutachter: Prof. Dr. Ursula Gather,
Universität Dortmund

Tag der Einreichung: 25. Mai 2001
Tag des Kolloquiums: 5. November 2001

pdfTEX-ed: 16. November 2001

Ich las im Bett mein Werk
[...] und bin im tiefsten auf-
gewühlt über meine große
Kunst. Bei Gott, das ist ein
Werk, das die Jahrhunder-
te überdauert. Wer sich dem
Geiste, der dies schuf, wider-
setzt, der hat das Urteil über
sich schon gesprochen.

*Carl Sternheim (1878–1942) in einem
Brief von 1904*

Der Geist ist das Gegenteil des
Geldes: Je weniger man da-
von hat, um so zufriedener ist
man.

Paul Masson

Einführung und Inhalt

Entstehung dieser Arbeit

Der Anstoß zum Thema dieser Arbeit kam aus den Forschungsinteressen meines Betreuers, Prof. Dr. Rieder, zur Robustheitstheorie für Zeitreihen- und Zustandsraummodelle.

Dementsprechend hatte ich mich mit den asymptotischen Grundlagen in Zeitreihenmodellen bereits in meiner Diplomarbeit (Ruckdeschel (1996)) auseinandergesetzt. Auf Fragen der Robustheit war ich während meines Studiums in zwei Praktika beim Aktiv-/Passivmanagement der Bayerischen Landesbank Girozentrale in München und beim Risiko-Controlling der Deutschen Bank in Frankfurt gestoßen. In diesem Bereich verwendete man Gaußsche Verteilungsannahmen¹ in Gestalt einer treibenden Brownschen Bewegung. Bei den Arbeiten stellte sich heraus, daß in der Praxis diese Verteilungsannahmen keinen Test standhalten; dennoch funktionieren die aus der Theorie abgeleiteten Handelsstrategien im “normalen Tagesgeschäft” gut. Hier könnte man vermuten, daß es vielleicht Ausreißer sind, die an den anderen Tagen die Modellgüte beeinträchtigen, und deren Einfluß man mit robusten Verfahren identifizieren und unter Umständen sogar mildern könnte.

In Bayreuth beteiligten sich Prof. Dr. Huber und Prof. Dr. Rieder zunächst im Rahmen des DFG-Schwerpunktprojekts “Echtzeit-Optimierung großer Systeme” 1995–1996 mit dem Kalman-Filter als rekursivem Verfahren.

In der Fortsetzung dieses DFG-Projekts im Projekt DFG Ri 332/8-1 von Prof. Dr. Rieder im Zeitraum 1997–99, in dessen Rahmen ich zu diesem Thema stieß, sollten dann vor allem die theoretischen Grundlagen der Robustheit für Zustandsraummodelle geklärt werden, in Abgrenzung zu einem Ansatz wie z.B. in der Princeton Studie (Andrews et al. (1972)), der allein mit empirischen Simulationsstudien argumentiert und der in Bayreuth wohl schon aus Kapazitätsgründen schwierig zu verfolgen gewesen wäre.

Vielmehr waren in den bisherigen Minimax-Ansätzen zur AO-Robustifizierung im Zeitbereich von Masreliez und Martin (1977) und Schick (1989) und Schick und Mitter (1994) die theoretischen Grundlagen teilweise noch klärungsbedürftig.

Das Ziel, Robustheit für Zustandsraummodelle begrifflich und theoretisch zu klären, stellte sich zu Beginn des Vorhabens auch im Hinblick auf die robuste Variante des Kalman-Filters und -Glätters von Künstler (1995) und Fahrmeir

¹Selbst in der allgemeineren Definition der Itô-Prozesse, c.f. Karatzas und Shreve (1991) benötigt man zumindest endliche zweite Momente des treibenden Prozesses, was aus Sicht der Robustheit kritisch zu sehen ist.

und Künstler (1999) . Deren Algorithmus, basierend auf dem Posterior–Modus, erschien als das bislang einzige, zur gewünschten Behandlung von IO–Ausreißern fähige Verfahren, wie in den zitierten Referenzen überzeugend in Datenbeispielen illustriert. Eine theoretische Absicherung im anvisierten Robustheitskontext wurde allerdings hierbei nicht verfolgt. Zudem ist dieser Ansatz nur begrenzt “on–line”–tauglich, da er nicht rekursiv ist, und die Problemdimensionen daher mit der Zahl der Beobachtungen wachsen.

Diese Arbeit entstand während meiner Mitarbeit im Projekt DFG Ri 332/8–1 von 1997–99 und meiner Tätigkeit als wissenschaftlicher Assistent am Lehrstuhl Mathematik VII in Bayreuth seit 1999.

Bei einem Forschungsaufenthalt auf Einladung von Prof. Dr. Härdle im Oktober 1999 am SFB 373 (“Quantifikation und Simulation Ökonomischer Prozesse”) der Humboldt Universität zu Berlin hatte ich die Gelegenheit, die rekursiven Verfahren dieser Arbeit in das Programmpaket *XploRe* einzubinden und diese in Form eines Tutorials in Ruckdeschel (2000b) zu dokumentieren.

Auf Einladung von Prof. Dr. Gather an den SFB 475 (“Komplexitätsreduktion in multivariaten Datenstrukturen”) nach Dortmund im November 2000 hatten Prof. Dr. Rieder und ich Gelegenheit, das On–line Monitoring in der Intensivmedizin “live” kennenzulernen, als uns Dr. Imhoff bei einem Besuch der Intensivstation der Städtischen Kliniken Dortmund dieses vorführte und erläuterte.

Bevor wir nun die neu zu definierenden Verfahren einführen, stellt sich zunächst die Frage in Abwandlung des Schiller–Zitats²:

**“Was heißt und zu welchem Ende studiert man
[robuste Kalman–Filter]?”**

Von einem Nicht–Mathematiker nach dem Inhalt dieser Arbeit gefragt, bietet es sich an, diesen zunächst an praktischen Beispielen abzustecken, bei denen bisherige Verfahren an ihre Grenzen stoßen.

On–line monitoring in der Intensivmedizin

Die vielleicht vitalste Anwendung des Kalman–Filters ergibt sich in Gestalt des *On–line monitoring* in der Intensivmedizin.

In der Intensivstation wird der Gesundheitszustand der Patienten mithilfe von automatisch protokollierten *Vitalsignalen*, wie verschiedenen Blutdrücken, Herzfrequenzen, Sauerstoffsättigung des Blutes usw. überwacht. Hierbei stellen sich folgende Forderungen an eine rechner–gestützte Überwachung — vgl. Gather et al. (2000), Fried et al. (2000)

- schnelle Reaktionszeit bei tatsächlichen (dauerhaften?) Zustandsänderungen — sowohl auf erwünschte, etwa nach Medikation, als auch auf unvorhergesehene, etwa bei auftretenden Komplikationen
- geringe Zahl an Fehlalarmen
- Konzentration der einströmenden Information — insgesamt bis zu 200 verschiedene Signale in unterschiedlichen Abstraten — auf wenige Output–Signale
- Erkennung typischer Muster

²c.f. Schiller (1962)

- daneben weitere Ziele außerhalb der Signalverarbeitung im engeren Sinn, wie Ausfallsicherheit, Datenschutz für die überwachten Patienten und Datensicherung zur anschließenden Rechenschaft

Ein Ansatz, an dieses Problem heranzugehen, ist die Formulierung als ein Zustandsraummodell — vgl. Smith und West (1983) und Daumer und Falk (1998), — und die in diesem Zusammenhang entwickelten Techniken zur Schätzung des Zustands, v.a. die des *Kalman-Filter*s zu verwenden.

zwei weitere typische Anwendungen

Beim Walzen von Stahl zu Stahlprofilen wird das Rohmaterial zunächst durch Walztechniken immer weiter verjüngt und schießt schließlich mit Geschwindigkeiten von bis zu 120m/s aus den Pressen. Dabei kommt Hochpräzisionsmeßtechnik mit extrem hohen Abtastraten — bis zu 400s^{-1} zum Einsatz. Trotzdem ist man am Walzende wegen der hohen Geschwindigkeiten nicht in der Lage, spezielle Profilmerekmale völlig zeitsynchron zu messen / kontrollieren, und es müssen statistische Verfahren verwendet werden, um auf die nicht erfaßten Profilstrecken zu schließen. Um große Mengen an Ausschuß zu vermeiden, ist es nötig, schnell auf Störungen auf der Walzstrecke reagieren zu können, wobei aber keinesfalls viele durch Meßstörungen ausgelöste Produktions-Stopps eingelegt werden dürfen. In diesem Kontext werden bereits Methoden aus der robusten Kontrolltheorie mit ihren, von unseren leicht abweichenden Begriffsbildungen verwendet — vgl. Ringwood (2000).

In der Makro-Ökonomie wird bei der Geldmengensteuerung versucht, die Auswirkungen von Interventionen der Notenbank zu identifizieren und gegebenenfalls zu optimieren. Hierzu werden relativ große Zustandsräume zum Einsatz, etwa ein 5-dimensionalen Zustandsraum und ein 3-dimensionalen Beobachtungsraum in Brand (1998) eingesetzt, vgl. auch Pecchi (1993).

Während es sich bei den ersten beiden Anwendungen um sehr zeitkritische Probleme handelt, zeichnet sich das letzte vor allem dadurch aus, daß das System, der Zustandsraum, nicht vollständig bekannt ist und sogenannte Hyper-Parameter aus den Daten mitgeschätzt werden müssen. Im zweiten Beispiel ist die Kenntnis der Hyper-Parameter weniger das Problem als vielmehr die extrem schnelle Verarbeitung. Im ersten Beispiel schließlich kommt als Herausforderung neben einem nicht vollständig spezifizierten System und der möglichst schnellen Reaktion noch hinzu, daß wir es bei einigen beobachteten Größen klar mit nicht-normalverteilten Variablen zu tun haben, weil sie über längere Zeiträume (fast) konstant sind, um dann größere Ausschläge zu produzieren.

Interventionsanalyse und robuste Verfahren³

Den Interventionen und unvorhergesehenen Zustandsänderungen in den vorigen Beispielen ist gemeinsam, daß sie nicht die "Regel" sind, es sich also in einem heuristischen Sinn um Ausreißer von den normalen Bedingungen handelt. Dies legt nahe, sie auch im mathematischen Sinn als Ausreißer aufzufassen und solche Verfahren zu verwenden, die bei Ausreißern stabil bleiben. Es stellt sich also das

³Weitere Ansätze zur *Interventionsanalyse* finden sich in Box und Tiao (1975), Wei (1990), Kapitel 9, pp. 184–204 und de Jong und Penzer (1998).

Problem, robuste Verfahren im Zustandsraummodell zu entwickeln, zu bewerten und zu realisieren. Dabei ergeben sich folgende wünschenswerte Ziele

- möglichst genaue Zustandsschätzung
- Unempfindlichkeit gegenüber systemfremden Störungen
- schnelle Reaktionszeit
- “on-line”-Fähigkeit
- eine für den Anwender einfach interpretierbare Form der Zustandsschätzung
- exakte Kenntnis des Systems nicht erforderlich

Dem entsprechen auf statistischer Seite

- hohe Effizienz bei der Zustandsschätzung im idealen Modell und akzeptable Effizienz bei Abweichungen
- Unverzerrtheit im idealen Modell
- Unempfindlichkeit gegenüber AO's, SO's⁴
- schnelle Reaktionszeit bei IO's, LO's⁴
- Rekursivität (evtl. m -Rekursivität)
- leicht interpretierbare/kontrollierbare Form der Zustandsschätzung
- Mitschätzung der Hyper-Parameter ohne Beeinträchtigung der Zustandsschätzung — Adaptivität im Sinn der Semiparametrik / robusten Statistik

Inhalte dieser Arbeit

Wie eingangs angekündigt, werden wir uns in dieser Arbeit dem Problem von der theoretischen Seite her nähern:

Ziel dieser Arbeit ist es, robuste Verfahren aus einem entscheidungstheoretischen Standpunkt heraus im Zustandsraummodell zu entwickeln, die den oben genannten Anforderungen genügen.

Während wir uns in Teil I der Arbeit — Kapitel 1–8 — mit der Zustandsschätzung in einem vollständig bekannten System beschäftigen, entwickeln wir in Teil II — Kapitel 9 und 10 — Verfahren zur robusten Schätzung der Hyper-Parameter in einem zeitinvarianten Zustandsraum. Ein ausführlicher Anhang vervollständigt und vertieft die Darstellung.

In Kapitel 1 werden wir das Zustandsraummodell und die sich darin ergebenden Schätzprobleme formal definieren. Dabei werden wir die klassischen Lösungen in Gestalt des Kalman-Filters und -Glätters und ihre Optimalitätseigenschaften erläutern. Anhand eines Beispiels werden die Defizite der klassischen Lösung unter Modellabweichungen dargestellt und anschließend bisherige Ansätze zur Robustifizierung skizziert.

In den nächsten drei Kapiteln werden wir mit Analogieschlüssen Ansätze, die sich in anderen Kontexten der robusten Statistik als optimal erwiesen haben, ins Zustandsraummodell übertragen:

⁴vgl. Abschnitt 1.4

In Kapitel 2 nehmen wir die Eigenschaft des klassischen Kalman-Filters als bestes lineares Verfahren zum Ausgangspunkt einer rekursiven Robustifizierung, dem rLS-Filter. Duncan und Horn (1972) haben den Kalman-Filter als Lösung eines gewichteten KQ-Problems gewonnen. In dem so entstehenden "Regressionsmodell" mit stochastischem Parameter definieren wir, basierend auf der robusten Regression wie in Kapitel 7 in Rieder (1994), in Kapitel 3 dieser Arbeit ein strikt rekursives — den rIC-Filter — und in Kapitel 4 ein m -Schritt-rekursives, robustes Verfahren — den mIC-Glätter und -Filter.

In Kapitel 5 vergleichen wir dann die rekursiven, robusten Verfahren rLS- und rIC-Filter in einer Mini-Studie im ersten Schritt eines Steady-State Modells mit weiteren in der Literatur als Robustifizierung vorgeschlagenen Verfahren.

Die statistischen Eigenschaften unserer Verfahren untersuchen wir in Kapiteln 6 und 7; in Kapitel 6 gehen wir dabei den Fragen der Unverzerrtheit, der Absolutstetigkeit sowie der asymptotischen Normalität nach und widmen uns in Kapitel 7 der Frage der (asymptotischen) Stationarität der Verfahren.

In Kapitel 8 rechtfertigen wir — zumindest numerisch — die mit Analogieschlüssen hergeleiteten Verfahren durch robuste Optimalität im Zustandsraummodell. Dazu verwenden wir einen geeigneten Ausreißertyp, den SO^5 , und verallgemeinern ein Optimalitätsresultat von Birmiwal und Shen (1993). In einem vereinfachten Modell untersuchen wir die robuste Optimalität im Minimax-Ansatz sowie in einem an das Lemma 5 von Hampel (1968) angelehnten Ansatz sowohl unter SO 's als auch unter AO 's. Im ursprünglichen Modell weisen wir [numerisch] nach, daß in einem noch zu definierenden Sinn der rLS-Filter optimal-robust ist. Für den rIC-Filter zeigen wir anhand einer verallgemeinerten Cramér-Rao-Schranke (fast-)Optimalität. Eine Optimalität des mIC-Verfahrens im IO-Fall wird nicht betrachtet.

Um die Verfahren dieser Arbeit auch in Situationen zugänglich zu machen, in denen das System erst aus den Daten ermittelt werden muß, wenden wir uns in den Kapiteln 9 und 10 der Frage der robusten Hyper-Parameter-Schätzung zu. Hierzu legen wir in Kapitel 9 zunächst die theoretischen Grundlagen, indem wir erst die lokal-asymptotische Theorie à la Le Cam und Hájek im Kontext unvollständiger Beobachtungen bereitstellen und dann den EM-Algorithmus in diese Theorie einordnen und robustifizieren.

In Kapitel 10 verallgemeinern wir den Ansatz aus Kapitel 9 auf Modelle mit Zeitabhängigkeit, um so das Zustandsraummodell erfassen zu können. In einem Exkurs beantworten wir dabei die Frage der Identifizierbarkeit der Hyper-Parameter mit einem lokal asymptotischen Ansatz. Der Spezialfall des normalen Zustandsraummodells ist ausführlich dargestellt, und in diesem Kontext werden wir einen klassisch-optimalen und einen robusten EM-Algorithmus entwerfen.

Untermauert werden diese Resultate durch einen ausführlichen Anhang:

In Anhang A wiederholen wir die wahrscheinlichkeitstheoretischen Begriffe der schwachen und vagen Konvergenz, der gleichgradigen Integrierbarkeit, sowie der Stationarität und Ergodizität und geben eine asymptotische Charakterisierung der Normalverteilung, die wir in Kapitel 6 verwenden. Ebenso stellen wir in einem Abriß die wesentlichen Optimalitätsresultate des lokal-asymptotischen Ansatzes im unabhängig identisch verteilten Setup dar.

Für die Untersuchung der Glattheit des Zustandsraummodells bei der Schätzung der Hyper-Parameter benötigen wir einige multivariate Verallgemeinerun-

⁵vgl. Abschnitt 8.1.3

gen der univariaten Theorie, die wir in Anhang B bereitstellen.

Bei der Frage der Stationarität unserer Verfahren können wir einige Schlüsse recht einfach aus dem Sylvesterschen Trägheitssatz ableiten, indem wir eine leicht verallgemeinerte Definition von *positiv definit* bei Matrizen verwenden. Dies findet sich ebenso in Anhang C wie ein “Quick-and-Dirty-Algorithmus” zur globalen Optimierung und ein Lemma zur konvexen Optimierung, welche wir beide verschiedentlich in der Arbeit verwenden.

Anhang D stellt einige numerische Verfahren zur Bestimmung optimal-robuster Influenzkurven für die Situation zur Verfügung, daß die klassischen Scores [multivariat] normalverteilt sind.

In Anhang E schließlich geben wir ein numerisches Verfahren zum Handling der block-tridiagonalen Matrizen an, wie sie bei der Verwendung einer blockgestützten Variante des mIC-Glätters und -Filters auftreten.

Danksagung

An dieser Stelle gilt es allen zu danken, die mir bei der Erstellung der Arbeit behilflich waren. Zu allererst ist hier mein Betreuer, Prof. Dr. Rieder, zu nennen, der mir jederzeit mit Rat und Tat zur Seite stand, von dem ich wesentliche Anregungen zum Inhalt der Arbeit erhielt, und bei dem ich die Begriffe der asymptotischen und robusten Statistik erlernte. In meiner Zeit als Assistent gab er mir die Gelegenheit, mit dem Thema verwandte Vorlesungen mit Übungen zu betreuen. Gerade in der Endphase des Zusammenschreibens hielt er mir den Rücken frei und ließ mir genügend Zeit für diese Aufgabe. Vor allem aber sei ihm für seine Geduld gedankt — und auch für das Einschreiten, als aus dem Projekt “Doktorarbeit” wegen meiner Vorstellung, ein möglichst umfassendes Werk abliefern zu wollen, eine enzyklopädische Abhandlung zu werden drohte. Von diesem Drang war nun auch die Zweitgutachterin, Prof. Dr. Gather, betroffen, die unter der eigenwilligen Umwandlung des olympischen Mottos in “amplius, longius, gravius” zu leiden hatte und bei der ich mich an dieser Stelle für das Koreferat bedanken möchte.

Bei einem Besuch in Dortmund auf ihre Einladung hin konnte ich — nicht zuletzt auch dank der detaillierten Informationen von Dr. Imhoff eine tiefere Einsicht in die Notwendigkeit gewinnen, “online”-taugliche robuste Verfahren zu entwickeln.

Ein herzlicher Dank geht auch nach München, von wo aus mich Dr. Rita Künstler in der Anfangsphase mit Erläuterungen zu ihrer Dissertation unterstützte. Ebenfalls bedanken möchte ich mich bei Prof. Dr. Härdle, der mir mit einer Einladung nach Berlin die Gelegenheit gab, zumindest die rekursiven Verfahren dieser Arbeit einer größeren Öffentlichkeit zugänglich zu machen.

Dank sei auch Dr. habil. Marlene Müller, die mir neben der Betreuung ihrer Vorlesung Stochastik I genug Freiräume für die Dissertation ließ.

Mein (fast-)Büronachbar Dr. habil. Christian Zillober hat sich dankenswerterweise die Zeit genommen, mich bei der numerischen Lösung des diskretisierten Optimierungsproblems in Abschnitt 8.6.8 zu beraten und hat das Problem mit den von mir gestellten Daten dann auch gelöst.

Im Zusammenhang mit der Charakterisierung der Endlichkeit der multivariaten Lokations-/Skalen-Fisher-Information in Abschnitt B.2 bedanke ich mich bei PD. Grunau für den Hinweis auf Maz’ya (1985).

Ullrich Deichsel, Kulmbach / Bamberg, verdanke ich das Beispiel aus der Stahlindustrie in dieser Einleitung, sowie Claus Greiber, Lehrstuhl VWL I in Bayreuth, das makro-ökonomische Beispiel.

Besonderer Dank geht auch an meinen Kollegen Matthias Kohl, der sich seinen wohlverdienten Italienurlaub mit dem Gegenlesen dieser Arbeit sicherlich nicht versüßt hat. Für einige Tips zu L^AT_EX sowie für die administrativen “Guide-Lines” bedanke ich mich bei Hans-Christian von Bothmer.

Außerhalb der Universität möchte ich mich bei meinen Freunden und Verwandten, besonders aber bei meinen Eltern für ihre Geduld bedanken, wenn die eine oder andere Rückmeldung etwas länger auf sich warten ließ oder ich am Telefon etwas abwesend war.

Abschließend geht mein Dank aber an Kerstin Hartmann, ohne die ich sicher nie fertig geworden wäre, die sich — als Soziologin — beim Korrekturlesen an die spezielle Sprache der Mathematik mit ihren vielen fremden Zeichen und Formeln erst hat gewöhnen müssen und mich nicht nur einmal, als es aufs Ende dieser Arbeit zugeht, nach einer langen Sitzung am Institut erst einmal wieder hat “erden” müssen. Daß das alles sicherlich ihrem eigenen Promotionsvorhaben gelinde gesagt nicht gerade förderlich war, macht meinen Dank nur noch größer.

I apologize for not making this thesis available to a broader scientific public, writing in English, but as is provided for in the examination regulations of the faculty for Mathematics and Physics at Bayreuth University a PHD-thesis is to be written in German except for very special cases.

At this stage I provide the reader with at least an English introduction starting next page as well as a conclusion starting on p. 236.

Introduction and Contents

Genesis of this work

The motivation for this subject emerged from the research interests of my advisor, Prof. Dr. Rieder, regarding robustness theory for time series and state space models.

According to this I had already studied the asymptotic statistics of time series models in the framework of my diploma thesis (Ruckdeschel (1996)). Robustness questions first arose to me during practical stages as a student at the at the Bayerischen Landesbank Girozentrale in Munich and Deutschen Bank in Frankfurt, where I worked on risk control. The theoretical framework used in this area is essentially⁶ based on Gaussian distributional assumptions in form of a driving Brownian motion.

Working with the data it became apparent that in real life — at least in banking — these distributional assumptions do not withstand any tests. Nevertheless the trading strategies derived from this theory work well in “common day trading”. This suggests that it might be outliers that at the few exceptional days lead to a minor model quality. By means of robust statistics it ought to be possible to quantify the influence and perhaps even attenuate the effects of these outliers.

At Bayreuth, Prof. Dr. Huber and Prof. Dr. Rieder studied the Kalman filter as a recursive procedure, initially taking part in the in the DFG Research Center “Real-Time Optimization of Large Systems” in 1995–1996.

The framework in which I came into contact with this topic was Prof. Dr. Rieder’s project DFG Ri 332/8-1 in 1997–99, continuing the studies for the Research Center. The aim of this second project was, above all, to settle the theoretical basis of robustness in state space models, in contrast to an approach like the Princeton Robustness Study (Andrews et al. (1972)), which argues solely with empirical simulation studies. Such an approach would probably have proven difficult to realize in Bayreuth considering the available capacities.

Moreover, up to then, in the main contributions to a robust minimax-approach against AO’s in the time domain, mainly due to Masreliez and Martin (1977) and Schick (1989) and Schick and Mitter (1994), the theoretical foundations had still not been settled in a completely satisfactory way.

The aim of settling robustness in the state space model conceptually and theoretically also arose in the beginning of this project with respect to the ro-

⁶Even for the definition of the more general Itô-processes, c.f. Karatzas and Shreve (1991), we at least need second moments; this assumption is to be seen critically from the point of view of robustness.

bust variant of the Kalman filter and smoother of Künstler (1995), Fahrmeir and Künstler (1999). Up to then, their algorithm based on the principle of posterior mode seemed to be the only one, capable of treating IO's in a desired way, as illustrated by convincing data examples in the cited references. A theoretical justification in the envisaged robustness context however was not pursued. As this approach is not recursive, the dimensions of the problem grow with the number of observations, and thus the use of this procedure “on-line” is restricted.

This thesis is the result of research begun during my employment at the DFG-project DFG Ri 332/8-1 from 1997-99 and continued after I became assistant at the chair Mathematics VII at Bayreuth University in 1999.

Invited by Prof. Dr. Härdle to Berlin to the SFB 373 (“Quantifikation und Simulation Ökonomischer Prozesse”) of Humboldt University in October 1999, I was given the opportunity to implement the recursive procedures of this thesis within the `XploRe`-package and to document them in form of a tutorial in Ruckdeschel (2000b), thus making them accessible to a broad public.

Together with Prof. Dr. Rieder, I was invited by Prof. Dr. Gather to the SFB 475 (“Komplexitätsreduktion in multivariaten Datenstrukturen”) at Dortmund in November 2000. There we were given the possibility of seeing on-site how on-line monitoring in intensiv care works in praxis, as Dr. Imhoff showed and explained to us in detail at a visit to the surgical department of Community Hospital Dortmund.

Before we propose some new procedures in this field, in this slight modification of Friedrich Schiller's⁷ headline for his course on universal history, we should perhaps first answer the question

“What is and to which end do we study [robust Kalman filtering]?”

Asked this question regarding the content of this thesis by a non-mathematician, it should be worthwhile starting with some practical examples showing up the limitations of procedures used up to now.

On-line monitoring in intensive care

The most vital application of Kalman filtering is probably the task of *on-line monitoring* in intensive care:

For this the state of health of a patient is monitored, automatically protocolling *vital signals* such as different blood pressures, heart frequencies, pulsoximetry, etc. In a natural way the following requirements to a computer-based monitoring arise — c.f. Gather et al. (2000), Fried et al. (2000)

- short response times to actual (lasting) changes of state — as well w.r.t. desired ones, e.g. after medication, as w.r.t. unforeseeable ones, e.g. after spontaneous complications
- small number of false alarms
- concentration of incoming information to a few output signals — up to 200 different signals with different sampling rates are used as input

⁷c.f. Schiller (1962)

- detection of typical patterns
- further goals beyond the scope of signal processing in a narrow sense such as failure safety, protection of privacy for the patients monitored and consistent and complete book-keeping of the records for subsequent accountance

One approach to comply with these requirements is to formulate the problem as a state space model — c.f. Smith and West (1983) and Daumer and Falk (1998), — and to use the techniques for state estimation developed in this context, especially those of *Kalman filtering*.

Two more applications

In the process of rolling steel to steel profiles or to wire, the raw material is progressively thinned out by big rolls, and finally shoots out with a speed of up to 120m/s at the end of the production line. In this framework high precision laser measurement equipment is being used with extremely high sample rates — up to 400s^{-1} to control the process. Nevertheless it is technically not possible to measure / control particular profile features in a completely time synchronous manner due to the high speed of the out-coming steel. So statistical procedures are employed to infer on the unobserved profile range. In order to avoid large amounts of junk it is necessary to be able to react quickly on malfunctions in the production line, where by no means many production breaks are to be incurred due to measurement failures. In this context methods from robust control theory are being applied already using notions slightly differing from ours — c.f. Ringwood (2000).

An important task in money supply control in the field of macro-economics is to identify effects of interventions of the central bank and if possible to optimize them. In this area relatively large state space models are used, e.g. of dimensions 5 in the state space and 3 in the observation space are proposed in Brand (1998), also c.f. Pecchi (1993).

While the first two applications are very time-critical problems, the last one is special because only the dimensions of the assumed state space model and the observation mechanism are known and so-called hyper parameters have to be estimated from the data. In the second example knowledge of the hyper parameters is not so much a problem as the speed of processing. In the first example, we additionally have the challenge of some obviously non-Gaussian variables among the observed data, being (practically) constant over long periods of time and then giving rise to some sharp peaks.

Intervention analysis and robust procedures⁸

The interventions and unforeseen changes of state discussed in the previous examples have in common that they are not the regular behavior, loosely spoken they are outliers. This suggests interpreting them as outliers in a mathematical sense as well and therefore just admitting procedures that remain stable in the presence of outliers. So we are confronted with the problem of designing,

⁸Further approaches to *intervention analysis* may be found in Box and Tiao (1975), Wei (1990), chapter 9, pp. 184–204 and de Jong and Penzer (1998).

judging and realizing robust procedures in the framework of state space models. Desirable goals in this context would be

- state estimation as accurate as possible
- insensitivity w.r.t. system–exogenous disturbances
- short response time
- “on–line”–capability
- a state estimator easily interpretable by the user
- no need of exact knowledge of the system

This corresponds to the following requirements on side of the statistician

- high efficiency for state estimation in the ideal model and acceptable efficiency in the presence of outliers
- unbiasedness in the ideal model
- insensitivity w.r.t. AO’s, SO’s⁹
- short response times w.r.t. IO’s, LO’s⁹
- recursiveness (maybe m –recursiveness)
- an easy–to–interpret/control form of the state estimator
- joint estimation of hyper parameters without affecting state estimation — adaptivity in the sense of semi–parametric / robust statistic

Contents of this thesis

In contrast to the practical applications illustrated above, we will approach the problem theoretically: This thesis intends to develop robust procedures for the state space model arising from a decision theoretic point of view that comply with the above–mentioned requirements.

While in part I of this thesis — chapters 1–8 — we are dealing with state estimation assuming a completely specified model, in part II — chapters 9 and 10 — we will develop procedures to robustly estimate the hyper parameters in a time–invariant state space model. A detailed appendix completes and deepens this elaboration.

In chapter 1 we define the state space model and pose the resulting estimation problems in a rigorous manner. Doing so we will recapitulate the classical solutions in form of the Kalman filter and smoother and their optimality properties. By means of an example we illustrate the defects of the classical solution under model deviances, followed by review of approaches to robustification in this area so far.

In the subsequent three chapters, we will transfer approaches by analogy to the state space model framework that have proven optimal in different contexts of robust statistics:

In chapter 2 we start with the property of the classical Kalman filter to be optimal among all linear procedures: This property is used to define a recursive robustification, the rLS filter. Duncan and Horn (1972) have shown that the Kalman filter can be interpreted as the solution to weighted least squares

⁹c.f. section 1.4

problem. In the “regression model” with a stochastic parameter arising in this approach we design a strictly recursive, robust procedure in chapter 3 — the rIC filter — and an m -step-recursive robust procedure in chapter 4 — the mIC smoother and filter — both based on robust regression as to be found in chapter 7 in Rieder (1994).

In chapter 5 we conduct a mini-study on the behavior of the rLS and rIC filters at the first step of a steady state model. In this study we compare these recursive, robust procedures to further procedures discussed in the literature.

The statistical properties of our procedures are examined in chapters 6 and 7; in chapter 6 we discuss the questions of unbiasedness, absolute continuity and asymptotic normality. After that we concentrate on the question of (asymptotic) stationarity of our procedures in chapter 7.

In chapter 8 we justify — at least numerically — the procedures within the state space model, which up to then were just based on formal analogy to robustness problems in different contexts. To do so we use a suitable type of outlier, SO^{10} , and generalize an optimality result by Birmiwal and Shen (1993). In a simplified model we study robust optimality for both AO’s and SO’s with a minimax approach as well as with an approach based on Lemma 5 of Hampel (1968). In the original model we show [numerically], that in a sense yet to be defined the rLS filter is robustly optimal. For the rIC filter we show (near) optimality by means of a generalized Cramér–Rao–bound. Optimality of the mIC–procedure for the IO–case is not considered.

In order to make these procedures accessible also in the situation of a system yet to be estimated from the data, we address the question of robust hyper parameter estimation in chapters 9 and 10. For this we build up the theoretical fundament in chapter 9, providing the local–asymptotic theory à la Le Cam and Hájek in the context of incomplete observations. We then cast the EM algorithm into the framework of this theory, proposing a robustification.

In chapter 10 we generalize the approach from chapter 9 to models with time–dependencies, thus covering the state space model. In an excursus we answer the question of identifiability of the hyper parameters by means of a local asymptotic approach. The special case of a Gaussian state space model is illustrated in detail, and in this context we design both a classically–optimal and a robust EM algorithm.

These results are supported by a detailed appendix:

In appendix A we recapitulate the notions of weak and vague convergence, of uniform integrability as well as stationarity and ergodicity in probability theory. We furthermore give an asymptotic characterization of the normal distribution, that is used in chapter 6. Finally, in a short abridge, we summarize the essential optimality results of the local–asymptotic approach in the setup of i.i.d. observations.

For the study of smoothness of the state space model in the estimation of the hyper parameters, we need some multivariate generalizations of univariate theory, which are provided in appendix B.

Using Sylvester’s law of inertia we can quite easily draw some conclusions on stationarity of our procedures, if we work with a slightly generalized notion of *positive definiteness* of matrices. This can be found in appendix C as well as a “quick–and–dirty algorithm” for global optimization and a lemma for convex

¹⁰c.f. section 8.1.3

optimization, both of which are used at different places within this thesis. Appendix D contains some numerical procedures for determining optimal–robust influence curves for the situation that the law of the classical scores is a [multi-variate] normal distribution. Eventually, in appendix E we present a numerical procedure for the handling of block–tridiagonal matrices as occurring in the mIC smoother and filter when using a blockwise–clipped version.

Inhaltsverzeichnis

Einführung und Inhalt	ii
Introduction and Contents [Engl.]	ix
Notation	xxv
I robuste Zustandsschätzung bei bekannten Hyper- Parametern	1
1 “klassischer” Rahmen	3
1.1 das Zustandsraummodell	3
1.1.1 Definition	3
1.1.2 Verteilungsannahmen	3
1.1.3 Abgrenzung	4
1.2 Problemstellung und klassische Lösung	6
1.2.1 Filter-/Glättungs-/Vorhersage-Problem	6
1.2.2 der klassische Kalman-Filter	7
1.2.3 Eigenschaften des klassischen Kalman-Filters	8
1.3 Charakterisierung der Linearität	10
1.4 Nichtrobustheit des klassischen Verfahrens	13
1.4.1 AO-Ausreißer	13
1.4.2 IO-Ausreißer	14
1.4.3 andere Ausreißer-Typen	14
1.4.4 Beispiele für AO's und IO's	14
1.5 Robustifizierungsansätze	15
1.5.1 Robustheit in der Kontrolltheorie	16
1.5.2 Robustheit durch “Hard Rejection”	16
1.5.3 Robustheit durch v_t, ε_t mit stärkeren Flanken als in (V3) oder (V4)	16
1.5.4 Robustheit durch Verwenden parallel laufender Filter	18
1.5.5 Robustheit mit robusten Verfahren aus anderen Kontexten	19
1.5.6 Minimax-Robustheit	19
2 der rLS-Filter	21
2.1 Motivation	21
2.2 Robustifizierung	21
2.2.1 robustes Optimierungsproblem	22

2.2.2	der rLS-Filter	22
2.2.3	algorithmische Eigenschaften des rLS-Filters	22
2.2.4	Wahl von b und K	23
2.3	Berechnung der Tuningkonstanten b und K	23
2.3.1	vereinfachte Bestimmung von b	23
2.3.2	numerische Bestimmung der Tuningkonstanten	24
2.3.3	Resultat für $t = 1$	24
2.4	Varianten	24
2.4.1	Hybridfilter	25
2.4.2	der Filter von Masreliez/Martin	25
2.4.3	optimierter MM-Filter	26
2.5	Ist M^{rLS} wirklich besser als M^{KK} ?	27
2.5.1	einige analytische Resultate	27
2.5.2	Bestimmung der optimalen Matrix M^{rLS}	28
2.6	Beispiele	29
2.7	Bewertung	32
2.8	Verfügbarkeit	33
3	der rIC-Filter	34
3.1	das Filterproblem als Regressionsproblem	34
3.1.1	der klassische Kalman-Filter als KQ-Regressions-Schätzer	34
3.1.2	das Zustandsraummodell als Regressionsmodell	35
3.1.3	Influenzkurven in diesem Regressionsmodell	36
3.1.4	der Kalman-Filter als One-Step-Schätzer	37
3.2	der rIC-Filter	37
3.2.1	Definition	37
3.2.2	Kriterium für b	38
3.2.3	eine Variante dieser Robustifizierung	39
3.3	Äquivarianz beim rIC und andere Normen	39
3.3.1	Fall $p = 1$	39
3.3.2	informations-standardisierte IC	40
3.3.3	selbststandardisierte IC	41
3.4	rICsep mit anderen Normen	41
3.4.1	informations-standardisierte IC	41
3.4.2	selbststandardisierte IC	42
3.5	Beispiele	43
3.6	Bewertung	44
3.7	Verfügbarkeit	45
4	der mIC-Glätter und -Filter	46
4.1	Problemstellung	46
4.1.1	Innovationsausreißer	46
4.1.2	ein großes Regressionsproblem	46
4.2	klassische Lösung	47
4.2.1	die Kleinste-Quadrate-Lösung	47
4.2.2	die $(m + 1)p$ -dimensionale Scores-Funktion	47
4.2.3	Influenzkurven in diesem Regressionsmodell	48
4.2.4	verschiedene Konstruktionen im idealen Modell	48
4.3	Robustifizierung	49

4.3.1	Definition von ψ^{mIC}	49
4.3.2	feste und variable Fensterweite	49
4.3.3	Posterior–Modus für andere Eichverteilungen als die Normalverteilung	50
4.3.4	separates Stützen	51
4.4	blockweises Stützen	52
4.4.1	Vorarbeiten	52
4.4.2	blockweise definierte Scores, IC’s und Fisher–Information	52
4.4.3	Kalibrierung	53
4.4.4	Verwendung numerischer Verfahren für Blockmatrizen	54
4.5	Varianten: Standardisierung in anderen Normen	57
4.6	Beispiele	57
4.7	Bewertung	62
4.8	Verfügbarkeit	62
5	Mini-Studie	63
5.1	Motivation	63
5.2	Notation und Definitionen	63
5.2.1	Modell	63
5.2.2	Verteilungen	64
5.3	Schätzer	65
5.3.1	der klassische Kalman–Filter in diesem Zusammenhang	65
5.3.2	geeichter Posterior–Modus	65
5.3.3	geeichter bedingter Erwartungswert	67
5.3.4	rLS	68
5.3.5	rIC	69
5.4	Berechnungen	69
5.4.1	numerische Verfahren	69
5.4.2	geschichtete Monte–Carlo Simulation für den MSE	71
5.5	Ergebnisse	71
5.5.1	Tabellen	71
5.5.2	graphische Darstellung	73
6	einige statistische Eigenschaften	76
6.1	Symmetrie und Unverzerrtheit der rob. Verfahren	76
6.1.1	ein Hilfssatz zu symmetrischen Verteilungen	76
6.1.2	Folgerungen für rekursive Filter	77
6.2	Dichten der rekursiven Filter	78
6.2.1	Dichte von $\Delta\beta_{t t}^{\text{LS}}$	79
6.2.2	Dichte von $\Delta\beta_{t t}^{\text{LS}}$ bei normalen, 1–dim. v_t, ε_s	79
6.2.3	Dichte von $\Delta\beta_{t t}^{\text{IC}}$	81
6.3	as. Normalität der robusten Verfahren	86
6.3.1	Resultate als Folgerung von Proposition A.2.4	86
6.3.2	Folgerungen aus den Dichte–Darstellungen	88
7	Stabilität und Stationarität der Filter	90
7.1	Folgerungen aus Sylvesterschem Trägheitssatz für den Kalman–Filter	90
7.2	Begriffe aus der Kontrolltheorie	92

7.2.1	Lyapunov–Stabilität	92
7.2.2	Zustandsraummodelle in der Kontrolltheorie	93
7.2.3	Kontrolltheorie im zeitinvarianten Fall	93
7.2.4	aus der Kontrolltheorie im allgemeinen Fall	97
7.3	Markovketten mit bel. Zustandsraum	98
7.3.1	robuste Verfahren als Realisation einer Markovkette	98
7.3.2	Harris–Rekurrenz	99
7.3.3	Begriffe der stochastischen Kontrolltheorie	101
7.4	as. Stabilität und Stationarität: rLS	102
7.5	as. Stabilität und Stationarität: rIC	106
7.6	as. Stabilität und Stationarität: mIC	109
8	optimal–robuste Filter	112
8.1	Problemstellung	112
8.1.1	verfolgte Ansätze	112
8.1.2	Problemstellung im vereinfachten Zustandsraummodell	113
8.1.3	SO– und AO–Ausreißer	113
8.1.4	Abgrenzung zu anderen Ansätzen	115
8.2	“Lemma 5”–Ansatz bei SO’s	116
8.2.1	Lösung im reduzierten Modell	116
8.2.2	der L5–Filter	117
8.2.3	Wahl von b	118
8.2.4	Beispiele zum L5–Filter	118
8.2.5	Bewertung — L5–Filter vs. rLS–Filter	118
8.2.6	Verfügbarkeit	120
8.3	Minimaxproblem für SO	120
8.3.1	Problemstellung	120
8.3.2	Lösung	120
8.4	Lösung im ursprünglichen Modell	126
8.4.1	das ursprüngliche Problem im reduzierten Modell	126
8.4.2	sequentielles vs. rekursives Spiel	127
8.4.3	Optimalität des rLS–Filters auch für $t > 1$	129
8.4.4	numerische Auswertung	131
8.5	Parametrisierung des Grades der Robustheit	135
8.6	“Lemma 5”–Ansatz für AO’s im reduzierten Modell	136
8.6.1	Problemstellung	136
8.6.2	Lösung	136
8.6.3	Existenz und Eindeutigkeit einer Lösung	137
8.6.4	Lösungsansatz mit Lagrangetechniken	138
8.6.5	Formulierung mit Dichten	138
8.6.6	Eigenschaften der Lösung	138
8.6.7	numerische Lösung	138
8.6.8	ein Beispiel	144
8.7	Minimax–Resultat für AO’s im reduzierten Modell	145
8.7.1	Problemstellung	145
8.7.2	Zusammenhang zur Bayes–Schätzung	147
8.7.3	Zusammenhang zum “Lemma 5”–Ansatz	151
8.7.4	numerische Berechnungen	151

8.8 Optimalität des rIC 152
 8.8.1 eine Cramér–Rao–Schranke für stochastische Parameter . 153
 8.8.2 “fast”–Optimalität des rIC 155

II robuste Schätzung der Hyper–Parameter 157

9 der EM–Algorithmus aus Sicht der as. Stat. 160

9.1 Motivation 160
 9.2 Glattheit von parametrischen Modellen 160
 9.2.1 Wurzeldichtenkalkül 161
 9.2.2 L_2 –Differenzierbarkeit in glatten Modellen 161
 9.3 Glattheit von Modellen für gröbere σ –Algebren 162
 9.3.1 Wurzeldichtenkalkül für gröbere σ –Algebren 162
 9.3.2 L_2 –Differenzierbarkeit für gröbere σ –Algebren 162
 9.3.3 L_2 –Differenzierbarkeit für Dreiecksschemata 163
 9.4 Folgerungen für unvollständige Beobachtungen 164
 9.4.1 L_2 –Differenzierbarkeit im m –Stichprobenproblem 164
 9.4.2 Situation mit Missings als m –Stichprobenproblem 166
 9.4.3 Umsetzung der allgemeinen asymptotischen Theorie 167
 9.5 EM–Algorithmus 169
 9.5.1 “klassischer” EM–Algorithmus 169
 9.5.2 statistische Eigenschaften des klassischen EM–Algorithmus’ 171
 9.5.3 “One–Step” EM–Algorithmus 172
 9.5.4 “klassischer” vs. “One–Step” EM–Algorithmus 175
 9.5.5 robuster EM–Algorithmus 175
 9.6 Beispiel: bivariate, normales Skalenmodell 177
 9.6.1 Modell 177
 9.6.2 klassischer EM–Algorithmus für Beispiele 9.6.3 und 9.6.4 180
 9.6.3 statistische Eigenschaften. 181
 9.6.4 One–Step–EM–Algorithmus 186

10 Glattheit des Zustandsraummodells 187

10.1 Motivation 187
 10.2 lokal–asympt. Theorie für rekursive Modelle 188
 10.2.1 Wurzeldichtenkalkül für Martingaldifferenzen 188
 10.2.2 Log–Likelihood–Entwicklung bei Martingaldifferenzen . . 190
 10.2.3 L_2 –Diff’barkeit bei Abhängigkeit 191
 10.2.4 Umsetzung der allgemeinen asymptotischen Theorie 197
 10.3 LAN–Eigenschaft für Zustandsraummodelle 200
 10.3.1 Skizze der Vorgehensweise 200
 10.3.2 Modell 200
 10.3.3 Glättungs–Filterierungen 200
 10.3.4 die LAN–Eigenschaft 202
 10.4 Exkurs: lokale Identifizierbarkeit 205
 10.4.1 Skizze des Vorgehens 207
 10.4.2 hinreichende Bedingungen für Regularität von \mathcal{I} 208
 10.4.3 asymptotische Kovarianz der empirischen ACF $\hat{\gamma}^{(Y)}$ 209
 10.4.4 Rangbetrachtung von \mathcal{T} 211

10.4.5	Berechnung der Hyper-Parameter aus $\hat{\gamma}^{(Y)}$	213
10.4.6	Differenzierbarkeit von G und Ableitung	214
10.4.7	Beweis von Theorem 10.4.6	215
10.5	Beispiel: normales Zustandsraummodell	215
10.5.1	L_2 -Differenzierbarkeit des normalen Modells	215
10.5.2	L_2 -Ableitung	216
10.5.3	Fisher-Information	220
10.5.4	bedingter ML-Schätzer	221
10.5.5	Algorithmus von Shumway und Stoffer	222
10.5.6	Identifizierbarkeit und ein Startschätzer	222
10.5.7	One-Step-EM-Algorithmus	223
10.5.8	robuster EM-Algorithmus	224
10.5.9	Adaptivität	226
	Schluß	227
	Conclusion [Engl.]	236
	Anhang	245
A	W-Theorie und as. Statistik	245
A.1	schwache Konvergenz von Maßen	245
A.2	Eigenschaften der Normalverteilung	246
A.2.1	bedingte Verteilungen bei Gaußschen Vektoren	246
A.2.2	elliptische Symmetrie	247
A.2.3	ein asymptotisches Cramér-Lévy-Theorem	247
A.3	Stationarität und Ergodizität	248
A.4	gleichgradige Integrierbarkeit	250
A.5	Abriß der as. Optimalitätstheorie der Statistik	251
B	multivariate Statistik	254
B.1	multivariate Statistik / multilineare Algebra	254
B.2	multiv. Lok-/Skalen-Fisher-Information	255
B.3	Glattheit des multiv. Lokations-/Skalen-Modells	262
B.4	generalisierte Fisher-Information	265
B.5	Kovarianzen von Kovarianzen	266
C	weitere verwendete Resultate	269
C.1	der Sylvestersche Trägheitssatz und Folgerungen	269
C.2	Optimierung	272
C.2.1	ein globaler Optimierungsalgorithmus	272
C.2.2	ein Hilfssatz aus der konvexen Analysis	273
C.3	Funktionalanalysis	273
C.3.1	einige Tatsachen	273
C.3.2	schwache Unterhalbstetigkeit von $\ \cdot\ _\infty$ in L_2	274
D	opt.-rob. IC's bei normalvert. Scores	275
D.1	Problemstellung	275
D.1.1	allgemeiner Fall	275
D.1.2	Relevanz für das Filter-Problem	275

D.2	Berechnung von A	276
D.2.1	multiplikative Struktur im Regressionsmodell	276
D.2.2	polare Darstellung der normalen Scores	276
D.2.3	technische Hilfsresultate	277
D.2.4	Berechnung der Integrale	279
D.2.5	Konvergenz der Algorithmen D.2.2 und D.2.3	280
D.3	Berechnung von b	280
D.4	b_{\min} in Abhängigkeit der Dimension	281
D.5	Verfügbarkeit	282
E	Numerik für Blockmatrizen	284
E.1	Cholesky-Zerlegung von Blockmatrizen	284
E.1.1	Cholesky-Zerlegung für allgemeine Block-Matrizen	284
E.1.2	Cholesky-Zerlegung für Block-tridiagonale Matrizen	285
E.2	LDM-Zerlegung für Block-indizierte Matrizen	286
	Literaturverzeichnis	287
	Personenregister	300
	Index	304

Abbildungsverzeichnis

1.1	AO's und IO's im Steady State Modell	15
2.1	$M^{\text{MM}}, M^{\text{rLS}}$ vs M^{KK} — großer Ausschnitt	30
2.2	$M^{\text{MM}}, M^{\text{rLS}}$ vs M^{KK} — kleiner Ausschnitt	31
2.3	der rLS-Filter in Beispiel 1 von P. Franěk	31
2.4	der rLS-Filter in Beispiel 2 von P. Franěk	32
2.5	der rLS-Filter in Beispiel 3 von P. Franěk – gefilterte $y_{t t}$ -Werte	33
3.1	der rIC-Filter in Beispiel 1 von P. Franěk	43
3.2	der rIC-Filter in Beispiel 2 von P. Franěk	44
3.3	der rIC-Filter in Beispiel 3 von P. Franěk – gefilterte $y_{t t}$ -Werte	45
4.1	Zustände im Steady-State Modell bei LO's	58
4.2	$\beta_{t t}^{\text{KK}}$ und $\beta_{t t}^{\text{mIC}}$ im idealen Modell und unter Kontamination	59
4.3	$\beta_{t-m t}^{\text{KK}}$ und $\beta_{t-m t}^{\text{mIC}}$ im idealen Modell und unter Kontamination	59
4.4	$\beta_{t-s t}^{\text{KK}}$ und $\beta_{t-s t}^{\text{mIC}}$, $s = 0, \dots, m$ im idealen Modell und unter Kontamination	60
4.5	Zustände in Beispiel 4.6.2 bei artifiziellem Ausreißer in $t = 199$	60
4.6	$\beta_{t t}^{\text{KK}}$ und $\beta_{t t}^{\text{mIC}}$ in Beispiel 4.6.2 im idealen Modell und unter Kontamination	60
4.7	$\beta_{t-m t}^{\text{KK}}$ und $\beta_{t-m t}^{\text{mIC}}$ in Beispiel 4.6.2 im idealen Modell und unter Kontamination	61
4.8	$\beta_{t-s t}^{\text{KK}}$ und $\beta_{t-s t}^{\text{mIC}}$, $s = 0, \dots, m$ in Beispiel 4.6.2 im idealen Modell und unter Kontamination	61
5.1	Schätzer für β_1 auf Basis der Beobachtung von y_1	73
5.2	$E[\beta_1 y_1]$ unter verschiedenen Verteilungsannahmen für ϵ_1	74
5.3	Posterior-Modus und bedingter Erwartungswert	75
6.1	Dichten von $\Delta\beta_{t t}^{\text{KK}}$ und $\Delta\beta_{t t}^{\text{rLS}}$ im Steady-State Modell für $\delta = 17.8\%$	82
8.1	der L5-Filter in Beispiel 8.2.2	119
8.2	die Dichten p^Y , \hat{p}_0^Y und \tilde{p}_0^Y	126
8.3	$r_0(r)$ im Steady-State Modell aus Beispiel 1.4.1, sowie aus Beispielen 8.2, 8.3	132
8.4	$r_0(r)$ im Zeitverlauf	133
8.5	$\hat{f}^{\text{rLS}}(y)$ und $\hat{f}^{\text{SO}}(y)$ in $t = 50$	134
8.6	$r_{50}(r)$ im Steady-State Modell aus Beispiel 1.4.1	134

8.7	Zusammenhang $r_{\text{so}}, \delta, b^{\text{so}} = 1/\tilde{\rho}^{\text{so}}$	136
8.8	f^{so} und f^{ao}	145
8.9	\tilde{P}_0^Y und \tilde{P}_1^Y	152
9.1	schematische Darstellung der Herleitung der asymptotischen Normalität des MLE	172

Tabellenverzeichnis

2.1	$M^{\text{rLS}}, M^{\text{MM}}$ vs M^{KK} im Steady-State Modell	29
2.2	weitere untersuchte Modelle	29
5.1	MSE bei der Schätzung von β_1 wissend y_1	71
5.2	$\text{MSE}_{P^\varepsilon}(\mathbf{E}_{\tilde{P}^\varepsilon}[\beta_1 y_1])$	72
5.3	relative Abweichung von $\text{MSE}_{P^\varepsilon}(\mathbf{E}_{P^\varepsilon}[\beta_1 y_1])$ in Tabelle 5.1	72
5.4	relative Abweichung von $\text{MSE}_{P^\varepsilon}(\mathbf{E}_{\tilde{P}^\varepsilon}[\beta_1 y_1]) : \text{MSE}_{P^\varepsilon}(\mathbf{E}_{P^\varepsilon}[\beta_1 y_1])$	72
8.1	untersuchte Modelle	131
D.1	b_{\min} und δ_{\max} in Abhängigkeit der Dimension der IC	282
D.2	Umsetzung der Algorithmen in XploRe	283

Notation

Abkürzungen

AO	additiver Ausreißer
as.	asymptotisch
f.s.	fast sicher
l.s.c.	unterhalb stetig
IO	Innovations-Ausreißer
KK	klassischer Kalman-Filter
L5	“Lemma 5”-Filter
LO	Level-Ausreißer
LS	Kleinste-Quadrate
LSE	Kleinste-Quadrate-Schätzer
MC	Monte-Carlo
mIC	mIC-Glätter und -Filter
ML	Maximum-Likelihood
MLE	Maximum-Likelihood-Schätzer
MM	Masreliez/Martin
MSE	mittlerer quadratischer Fehler
NB	(unter der) Nebenbedingung
OS	One-Step
rIC	rIC-Filter
rLS	rLS-Filter
schließlich	für alle hinreichend großen Indizes
SO	substitutiver Ausreißer
sto.u.	stochastisch unabhängig
u.i.v.	unabhängig identisch verteilt
////	QED

Mengen und Funktionen

\mathbb{N}	die natürlichen Zahlen $\{1, 2, \dots\}$
\mathbb{N}_0	die natürlichen Zahlen mit 0 $\{0, 1, 2, \dots\}$
\mathbb{Z}	die ganzen Zahlen
\mathbb{Q}	die rationalen Zahlen
\mathbb{R}	die reellen Zahlen
\mathbb{R}	die erweiterten reellen Zahlen $[-\infty; \infty]$, homeomorph zu identifizieren mit $[-1; 1]$ über die Isometrie $z \mapsto z/(1 + z)$

\mathbb{C}	die komplexen Zahlen
\mathcal{S}_k	die k -dimensionale Einheits-Sphäre im \mathbb{R}^k
ω_{k-1}	Oberfläche von \mathcal{S}_k
$I_A, I(A)$	Indikatorfunktion der Menge A

 σ -Algebren

$\mathbb{B}, \bar{\mathbb{B}}$	Borel- σ -Algebra von \mathbb{R} bzw. $\bar{\mathbb{R}}$
$\sigma(\mathcal{E})$	kleinste σ -Algebra (auf der Grundmenge Ω), die \mathcal{E} umfaßt
$A \otimes B$	Produkt- σ -Algebra von A und B
B^k	$\bigotimes_{1 \leq i \leq k} B$

Maße

$\mathcal{M}(\mathcal{A})$	die Menge aller (nicht-negativen) Maße auf der σ -Algebra \mathcal{A}
$\mathcal{M}_\sigma(\mathcal{A})$	die Menge aller (nicht-negativen) σ -finiten Maße auf der σ -Algebra \mathcal{A}
$\mathcal{M}_b(\mathcal{A})$	die Menge aller (nicht-negativen) beschränkten Maße auf der σ -Algebra \mathcal{A}
$\mathcal{M}_1(\mathcal{A})$	die Menge aller Wahrscheinlichkeitsmaße auf der σ -Algebra \mathcal{A}
$\text{supp}(P)$	kleinste abgeschlossene Teilmenge A von Ω (separabel, metrisch), so daß $P(\Omega \setminus A) = 0$ [Parthasarathy (1967), II Def. 2.1]
$\mu * \nu$	Faltung von μ und ν
$\mu \otimes \nu$	Produktmaß von μ und ν
μ^k	$\bigotimes_{1 \leq i \leq k} \mu$
$\mu \ll \nu$	μ wird von ν dominiert
$(\mu_n) \ll (\nu_n)$	(μ_n) ist benachbart zu (ν_n)
$\mu_n \xrightarrow{w} \mu$	schwache Konvergenz von μ_n gegen μ
$\mu_n \xrightarrow{v} \mu$	vage Konvergenz von μ_n gegen μ

Zufallsvariablen und Erwartungswerte

$X \sim F$	X ist verteilt nach F
$\sigma(X)$	die von X erzeugte Urbild- σ -Algebra
$X \circ P$	Bildmaß von X unter P
$P^X, \mathcal{L}(X)$	Bildmaß von X unter einem nicht näher spezifizierten Maß Pr auf dem Definitionsbereich
$\varphi_X(t)$	charakteristische Funktion von $\mathcal{L}(X)$
$E[X]$	Erwartungswert von X
$\text{Var}[X]$	Varianz von X
$\text{Cov}[X]$	Kovarianz(matrix) von X
$\text{Cov}[X, Y]$	Kovarianz(matrix) von X und Y
$E[X \mathcal{A}]$	bedingter Erwartungswert von X gegeben die σ -Algebra \mathcal{A}
$E[X Y]$	bedingter Erwartungswert von X gegeben $\sigma(Y)$

$E[X Y = y]$	faktorisierter bedingter Erwartungswert von X gegeben $Y = y$
$\kappa^{(X)}, \kappa(X)$	vierte Kumulante von X — vgl. Abschnitt B.1
$\gamma^{(X)}$	Autokovarianzfunktion von X
$X_n \xrightarrow{P_n} X$	stochastische Konvergenz von X_n unter P_n gegen X
$X_n = O_{P_n}(r_n)$	Straffheit von X_n/r_n unter P_n
$X_n = o_{P_n}(r_n)$	stochastische Konvergenz von X_n/r_n unter P_n gg. 0
$X = Y [\mu]$	$\mu(X \neq Y) = 0$
$X \geq Y [\mu]$	$\mu(X - Y < 0) = 0$

Verteilungen und Erwartungswerte unter Kontamination

$P^{[\cdot]}$	ideales (Bild)Maß (von $[\cdot]$)
$\hat{P}^{[\cdot]}$	“reales”, kontaminiertes (Bild)Maß (von $[\cdot]$)
$\tilde{P}^{[\cdot]}$	kontaminierendes (Bild)Maß (von $[\cdot]$)
$\check{P}^{[\cdot]}$	Eich-(Bild)Maß (von $[\cdot]$)
E_{id}	Erwartungswert, ausgewertet im idealen Modell
E_{cont}	Erwartungswert, ausgewertet unter einer nicht näher spezifizierten, kontaminierten Verteilung

spezielle Verteilungen

I_x	(Dirac) Einpunktmaß auf x
λ^k	Lebesguemaß auf dem \mathbb{R}^k
$\text{ufo}(A)$	Gleichverteilung auf A
$\mathcal{N}_p(\mu, \Sigma)$	p -dimensionale Normalverteilung mit Mittelwert μ und Kovarianz Σ , $\mathcal{N}(\mu, \sigma^2) = \mathcal{N}_1(\mu, \sigma^2)$
$\chi_n^2, \chi_n^2(\delta^2)$	zentrale bzw. nichtzentrale χ^2 -Verteilung mit n Freiheitsgraden [und Nichtzentralität δ^2]
t_n	t -Verteilung mit n Freiheitsgraden
Cauchy	Cauchy- oder t -Verteilung mit einem Freiheitsgrad
Φ, φ	Verteilungsfunktion und Lebesgue-Dichte von $\mathcal{N}(0, 1)$

Verteilungsannahmen — (vgl. Abschnitt 1.1.2)

(V1)	Verteilungen mit endlicher [multivariate] (Lokations-) Skalen-Fisher-Information
(V2)	treibende Prozesse als unkorrelierte, weiße Rauschen
(V3)	Gaußsches Zustandsraummodell
(V4)	Gaußsches, invariantes Zustandsraummodell

mathematische Symbole

$\text{ggT}(a, b)$	größter gemeinsamer Teiler von a und b
$\text{kgV}(a, b)$	kleinstes gemeinsames Vielfaches von a und b
f_+, f_-	$ f \mathbb{I}_{\{f>0\}}$ bzw. $ f \mathbb{I}_{\{f<0\}}$
$a \wedge b, a \vee b$	$\min(a, b), \max(a, b)$
\inf, \sup	punktweises Infimum / Supremum
\inf_P, \sup_P	P -wesentliches Infimum / Supremum
\uparrow, \downarrow	monotone Konvergenz
$a \rightsquigarrow b$	a ersetzt durch b
o, O	die üblichen Landausymbole
A°, \bar{A}	offener Kern bzw. abgeschlossen Hülle von A
$f', \partial f, \nabla f$	Ableitung von f
$\text{lin}(A)$	abgeschlossene, lineare Hülle von A
$\text{lin}_k(A)$	abgeschlossene, lineare Hülle der Vektoren $Ma + b$, $a \in A$, $M \in \mathbb{R}^{k \times p}$, $b \in \mathbb{R}^k$, p die Dimension der Elemente aus A — vgl. (1.18)
A^\perp	orthogonales Komplement von A
$\text{op}(x A)$	orthogonale Projektion von x auf $\text{lin}_k(A)$, k die Dimension von x
$\min!, \max!$	$\hat{=}$ "... ist zu minimieren / maximieren"
$\stackrel{!}{=}$	$\hat{=}$ "... ist so zu bestimmen, daß ... " oder "Es ist zu zeigen, daß Gleichheit herrscht.", entsprechend für andere Relationssymbole
$::$	definiert als

Matrix- und Vektornotation

$a_{s:t}$	für Vektoren $(a_s^\tau, a_{s+1}^\tau, \dots, a_t^\tau)^\tau$
$A^{i,j}$	für Matrizen $a_s \cdot a_{s+1} \cdot \dots \cdot a_t$
$a_{(i)}, A_{(i),(j)}, A^{(i),(j)}$	$A^{i,j} := (A^{-1})_{i,j}$ entsprechende Blöcke bei blockweise definierten Matrizen/Vektoren; wenn aus dem Zusammenhang klar auch ohne (\cdot)
\mathbb{I}_p	Einheitsmatrix in p Dimensionen
$\text{GL}(p)$	die Menge aller invertierbaren $\mathbb{R}^{p \times p}$ -Matrizen
$\mathcal{O}(p), \mathcal{U}(p)$	die Menge aller orthogonalen [unitären] $\mathbb{R}^{p \times p}$ - [$\mathbb{C}^{p \times p}$] Matrizen
A^τ, A^H	die Transponierte [Hermitesche] von A
$\text{rk } A$	Rang von A
$\ker A$	Nullraum von A
$\text{tr } A$	Spur von A
$\sigma(A)$	Menge der Eigenwerte von A
$\rho(A)$	Betrag des betragsmäßig maxim. Eigenwertes von A
$\text{maxev}(A),$ $\text{minev}(A)$	maximaler (minimaler) Eigenwert von A , wenn A nur reelle Eigenwerte besitzt
$A \preceq B, A \prec B$	$B - A$ positiv (semi)definit im Sinn von Definition C.1.1

$A^{\frac{1}{2}}$	eindeutige, positiv (semi)definite Wurzel von $A = A^T \succeq 0$
$ x $	euklidische Norm von $x \in \mathbb{R}^p$
$ x _A$	$ x^T A^{-1} x ^{\frac{1}{2}}$ für $A = A^T \succ 0$
$\text{vec}(A)$	in einen Vektor umgewandelte Matrix A — vgl. Abschnitt B.1
$\text{svec}(A)$	in einen Vektor umgewandelte Matrix $A = A^T$, wobei man die Hälfte unterhalb der Diagonalen wegläßt — vgl. Abschnitt B.1
$A \otimes B$	Kroneckerprodukt von A und B
$A^{\otimes k}$	$A \otimes \dots \otimes A$ (k mal)
$K_{p,q}$	Kommutatormatrix — vgl. Abschnitt B.1
$\text{diag}(a_1, \dots, a_n)$	(Block-)Diagonalmatrix mit a_i als (Block-)Einträgen auf der Diagonale
$\text{ndiag}_1(a_1, \dots, a_n)$	Matrix mit a_i als (Block-)Einträgen auf der ersten oberen Neben-(Block-)Diagonalen

Funktionsräume

$\mathcal{C}^k(\Theta)$	Raum der beschränkten stetigen Funktionen $\Theta \rightarrow \mathbb{R}^k$ versehen mit der sup-Norm, $\mathcal{C}(\Theta) = \mathcal{C}^1(\Theta)$
$\mathcal{C}_c^k(\Theta)$	der Unterraum von $\mathcal{C}^k(\Theta)$ der stetigen Funktionen mit kompaktem Träger
$\mathcal{D}_k(\Theta)$	der Unterraum von $\mathcal{C}^k(\Theta)$ der beliebig oft differenzierbaren Funktionen mit kompaktem Träger
$L_2^k(P), L_p^k(P)$	Hilbert-/Hölder-Raum der (Äquiv.-klassen der) \mathbb{R}^k -wertigen, P -meßb. Funktionen, so daß $\int f ^p dP < \infty$; $L_p(P) = L_p^1(P)$, ($p \in [1, \infty]$)
$\mathcal{L}_2^k(\mathcal{A}), \mathcal{L}_p^k(\mathcal{A})$	Hilbert-/Hölder-Raum der (Äquivalenzklassen der) $\xi dP^{\frac{1}{p}}$ mit $\xi \in L_p^k(P)$, $P \in \mathcal{M}_b$ — vgl. Abschnitt 9.2.1; $\mathcal{L}_p(\mathcal{A}) = \mathcal{L}_p^1(\mathcal{A})$, ($p \in [1, \infty]$)
$\sqrt{dP}, dP^{\frac{1}{p}}$	Element von $\mathcal{L}_2(\mathcal{A}), \mathcal{L}_p(\mathcal{A})$
L, ℓ	die Menge aller zugelassenen Verlustfunktionen / eine spezielle Verlustfunktion aus L , vgl. Definition A.5.7

Variablenamen

p	Dimension des Zustandsraums
q	Dimension des Beobachtungsraums
Z_t, Z	Beobachtungsmatrix
F_t, F	Zustandsübergangsmatrix
v_t	Innovationen
ε_t	Beobachtungsfehler
β_t	Zustand zum Zeitpunkt t
y_t, y_t^h	Beobachtung bzw. deren Realisation zum Zeitpunkt t
Q_t, Q	Varianz der Innovationen

V_t, V	Varianz der Beobachtungsfehler
$\beta_{t s}$	$E[\beta_t y_{1:s}]$
$\hat{\beta}_{t s}^{[\cdot]}$	Filter/Glätter/Vorhersage gemäß Verfahren $[\cdot]$ — wenn aus Zusammenhang klar auch ohne $[\cdot]$
$\Delta\beta_{t s}^{[\cdot]}$	$\beta_t - \hat{\beta}_{t s}^{[\cdot]}$, für $s = t - 1$ auch $\Delta\beta_t^{[\cdot]}$
$\Sigma_{t s}^{[\cdot]}$	$\text{Cov}[\Delta\beta_{t s}^{[\cdot]}]$
M_t^{KK}	Kalman-Gain zum Zeitpunkt t
J_t	$\Sigma_{t t}^{\text{KK}} F_{t+1}^{\text{KK}} (\Sigma_{t+1 t}^{\text{KK}})^{-1}$
$\Delta y_{t s}^{[\cdot]}$	$y_t - Z_t \hat{\beta}_{t s}^{[\cdot]}$, für $s = t - 1$ auch $\Delta y_t^{[\cdot]}$
$\Delta_t^{[\cdot]}$	$\text{Cov}[\Delta y_t^{[\cdot]}]$
$v_{t s}^{[\cdot]}$	$\beta_{t s}^{[\cdot]} - F_t \beta_{t-1 s}^{[\cdot]}$
$\varepsilon_{t s}^{[\cdot]}$	$y_{t s}^{[\cdot]} - Z_t \beta_{t s}^{[\cdot]}$
$\Sigma^{[\cdot]}$	$\lim_t \Sigma_{t t-1}^{[\cdot]}$ oder falls t klar $\Sigma_{t t-1}^{[\cdot]}$
$\mathcal{T}^{[\cdot]}$	$\lim_t \Sigma_{t t}^{[\cdot]}$ oder falls t klar $\Sigma_{t t}^{[\cdot]}$
$\Delta^{[\cdot]}$	$\lim_t \text{Cov}[\Delta y_t^{[\cdot]}]$ oder falls t klar $\Delta_t^{[\cdot]}$
$M^{[\cdot]}$	$\lim_t M_t^{[\cdot]}$ oder falls t klar $M_t^{[\cdot]}$
$b^{[\cdot]}$	Clipping-/Stutzhöhe
δ	relativer Effizienzverlust
$r^{[\cdot]}$	Kontaminationsradius
X_t	unbeobachtbarer Zustand im reduzierten Modell
$\sigma(y_{1:t}^0)$	um die Kenntnis von $\Delta\beta_{1:t}^{\text{KK}}$ augmentierte σ -Algebra $\sigma(y_{1:t})$
$\sigma(y_{1:t}^{[0]})$	$\sigma(y_{1:t}^0)$ resp. $\sigma(y_{1:t})$
$D(y)$	Abkürzung für $E_{\text{id}}[X Y=y] - E_{\text{id}}[X]$
f_ε	Abkürzung für $E_{P^\varepsilon}[f(X+\varepsilon) \varepsilon]$
$\hat{f}^{\text{SO}}, \hat{f}^{\text{AO}}$	optimale Korrektur im "Lemma 5-Ansatz" für SO/AO
f_0, f_1	optimale Korrektur im Minimax-Ansatz für SO/AO
π_1, π_2	Projektion auf die entsprechenden (Blöcke von) Randverteilungen
\mathcal{P}	statistisches Modell
Λ, \mathcal{I}	L_2 -Ableitung, Fisher-Information
ψ	Influenzkurve
$\hat{\psi}$	klassisch optimale Influenzkurve
\mathcal{J}	Skalen-Fisher-Information
$\mathcal{P}^{\mathcal{U}}$	größeres Modell längs der Unter- σ Algebra \mathcal{U}

$\Lambda^{\mathcal{U}}, \mathcal{I}^{\mathcal{U}}$	entsprechende L_2 -Ableitung, Fisher-Information
G, g	in Kapitel 10: Verteilung und λ^k -Dichte der Innovationen in (V1)
H, h	in Kapitel 10: Verteilung und λ^k -Dichte der Beobachtungsfehler in (V1)
\bar{Q}, \bar{V}	in Kapitel 10: symmetrische Wurzeln aus Q, V
\mathcal{F}_t	Filtrierung
$\mathcal{F}_{T,s,k_{T,s}}$	Glättungsfiltrierung $\mathcal{F}_{T,s,k_{T,s}} := \sigma(\beta_{0:s}, Y_{1:k_{T,s}})$ — vgl. (9.58)
\mathcal{G}_T^m	m -Schritt Glättungsfiltrierung $(\mathcal{F}_{T,0,\min\{s+1,T\}})$

Teil I

robuste Zustandsschätzung
bei bekannten
Hyper-Parametern

Kapitel 1

“klassischer” Rahmen

1.1 das Zustandsraummodell

1.1.1 Definition

Der Rahmen der vorliegenden Untersuchung ist das Zustandsraummodell, das sich ergibt, wenn wir auf einen \mathbb{R}^p -wertigen Zustand β_t nur indirekt in Form von \mathbb{R}^q -wertigen Beobachtungen y_t schließen können.

Genauer ergibt sich folgendes Modell für Matrizen $F_t \in \mathbb{R}^{p \times p}$, $Z_t \in \mathbb{R}^{q \times p}$,

$$\beta_t = F_t \beta_{t-1} + v_t \quad (1.1)$$

$$y_t = Z_t \beta_t + \varepsilon_t \quad (1.2)$$

mit *Innovationen* v_t und *Beobachtungsfehlern* ε_t und einem Anfangszustand β_0 . Dabei sind die Matrizen F_t , Z_t deterministisch, die Innovationen v_t und die Fehler ε_t , mithin auch die Zustände β_t und die Beobachtungen y_t zufällig.

Bemerkung 1.1.1 Synonyme für Zustandsraummodelle sind *Hidden Markov Models (HMM)* (vgl. Künsch (2001)) und *Dynamische Lineare Modelle (DLM)* (vgl. Harrison und Stevens (1976), West und Harrison (1989)).

1.1.2 Verteilungsannahmen

Über die Verteilung dieser Zufallsgrößen im idealen Modell werden wir in den verschiedenen Abschnitten der Arbeit je nach Anforderungen folgende Annahmen benutzen:

(V1) Verteilungen mit endlicher Fisher-Information

$$\begin{aligned} v_t &\stackrel{\text{u.i.v.}}{\sim} G, & G &\text{ mit endlicher Lokations-/Skalen-Fisher-Information} \\ \varepsilon_t &\stackrel{\text{u.i.v.}}{\sim} H, & H &\text{ mit endlicher Skalen-Fisher-Information} \\ \beta_0 &\sim G_0 & \beta_0, \{\varepsilon_t\}, \{v_t\} &\text{ als Prozesse sto. u.,} \end{aligned} \quad (1.3)$$

wobei die Begriffe [Lokations-]/Skalen-Fisher-Information im Sinn von Abschnitt B.2 zu verstehen sind.

(V2) treibende Prozesse als unkorrelierte, weiße Rauschen

$$\begin{aligned} E[\beta_0] &= a_0, & E[v_t] &= 0, & E[\varepsilon_t] &= 0, \\ \text{Cov}[\beta_0] &= \Sigma_0, & \text{Cov}[v_t] &= Q_t, & \text{Cov}[\varepsilon_t] &= V_t, \\ \text{Cov}[\varepsilon_s, \varepsilon_t] &= 0 \quad \forall s \neq t, & \text{Cov}[v_s, v_t] &= 0 \quad \forall s \neq t, & \text{Cov}[\varepsilon_s, v_t] &= 0 \quad \forall s, t \end{aligned} \quad (1.4)$$

(V3) Gaußsches Zustandsraummodell

$$\begin{aligned} v_t &\overset{\text{sto.u.}}{\sim} \mathcal{N}_p(0, Q_t), & \varepsilon_t &\overset{\text{sto.u.}}{\sim} \mathcal{N}_q(0, V_t), \\ \beta_0 &\sim \mathcal{N}_p(a_0, \Sigma_0), & \beta_0, \{\varepsilon_t\}, \{v_t\} &\text{ als Prozesse sto. u.,} \end{aligned} \quad (1.5)$$

(V4) Gaußsches, zeitinvariantes Zustandsraummodell

$$\begin{aligned} v_t &\overset{\text{u.i.v.}}{\sim} \mathcal{N}_p(0, Q), & \varepsilon_t &\overset{\text{u.i.v.}}{\sim} \mathcal{N}_q(0, V), \\ \beta_0 &\sim \mathcal{N}_p(a_0, \Sigma_0), & \beta_0, \{\varepsilon_t\}, \{v_t\} &\text{ als Prozesse sto. u.,} \end{aligned} \quad (1.6)$$

Im ersten Teil der Arbeit gehen wir dabei mindestens von (V2) aus, in weiten Teilen sogar von (V3) oder (V4).

Dabei nehmen wir stets an, daß wir die *Hyper-Parameter* F_t , Z_t , Q_t , V_t kennen, wobei wir in Kapitel 7 die (asymptotische) Stationarität unserer Verfahren in dem Fall untersuchen, daß die Hyper-Parameter zeitinvariant sind.

Von dieser Zeitinvarianz gehen wir auch im zweiten Teil der Arbeit aus, nehmen aber nur noch an, daß wir $Z_t = Z$ kennen, daß die restlichen Hyper-Parameter aber aus den Daten mitzuschätzen sind. Dazu benutzen wir dann (V1), vgl. Abschnitt 10.5.

1.1.3 Abgrenzung

Diese Spezifikation eines Zustandsraummodells ist keinesfalls die einzig mögliche, und so grenzen wir uns in diesem Abschnitt von weiteren Zustandsraumdefinitionen ab.

deterministisch / stochastisch

Auch in der Steuerungs-/Kontrolltheorie beschäftigt man sich mit Zustandsräumen, wenn auch mit anderer Akzentuierung. Untersuchungsobjekt ist¹ ein dynamisches System $[(T; \Sigma; \Omega; U), x, g]$ mit Zeitbereich T , Zustandsraum Σ , Steuerungs-Raum U , mit U einer Menge von Funktionen $u : T \rightarrow \Omega$, einem Zustand $x(t)$, der sich mit einer Übergangsfunktion ϕ als $x(t) = \phi(t, u|_{(t_0;t]}, x(t_0))$ entwickelt, und einem Beobachtungsmechanismus g , der zu einem Zeitpunkt $t \in T$ die Beobachtung $y(t) = g(x(t), u(t), t)$ generiert.

In diesem Zusammenhang stellt man sich beim Problem der optimalen Steuerung / Kontrolle die Aufgabe, zu einem Anfangswert $(t_0, x(t_0)) \in T \times \Sigma$ und einer reellwertigen Verlustfunktion², die zu einer Steuerung $u \in U$ den Wert $J(x(t_0), t_0, u, x, t)$ liefert, diejenige Steuerung $\hat{u} \in U$ zu finden, die J minimiert.

Aus dieser Theorie können wir viele Begriffe übernehmen, vgl. Abschnitt 7.2;

¹vgl. Athans und Falb (1966), pp. 165–194

²im Kontext der Kontrolltheorie oft auch Performance-Index

identifizieren wir $T \rightsquigarrow \mathbb{N}_0$, $\Sigma \rightsquigarrow \mathbb{R}^p$, $\Omega \rightsquigarrow \mathbb{R}^q$, $x(t) \rightsquigarrow \beta_t$, $u(t) \rightsquigarrow v_t$, $y(t) \rightsquigarrow Z\beta_t$, wobei wir $\varepsilon_t \equiv 0$ und

$$\phi(t, u|_{(t_0;t]}, x(t_0)) = v_t + \sum_{j=1}^{t-1} F_{t:t-j+1} v_{t-j} + F_{t:1} \beta_0$$

mit $F_{s:t} = F_s \cdot \dots \cdot F_t$ setzen, so werden die Zusammenhänge mit unserer Definition klar.

Darauf aufbauend hat sich eine stochastische Variante der Kontrolltheorie entwickelt, bei der es zusätzlich zu der Zustandsraummodellierung in (1) und (2) einen externen Steuerungsprozeß u_t gibt, der von $y_{1:t-1}$ abhängen darf, was sich in einer Modifikation von (1) zu

$$\beta_t = F_t \beta_{t-1} + v_t + u_t \quad (1.7)$$

niederschlägt. Dieser Erweiterung entspricht auf der Zeitreihenseite die Verwendung von ARMAX-Modellen — X für exogene Variable, vgl. etwa Lütkepohl (1993), 13.2.1.c, Caines (1988), pp. 97, Hannan und Deistler (1988).

Den aus der deterministischen Kontrolltheorie zur Verfügung stehenden Begriffen der Beobachtbarkeit, Kontrollierbarkeit und ihre Abschwächungen, vgl. Abschnitt 7.2, stehen entsprechende Begriffsbildungen der stochastischen Kontrolltheorie gegenüber, die wir in Abschnitt 7.3 einführen und die wir für Stationaritätsüberlegungen unserer Verfahren nutzen.

Ohne hier auf weitere Literatur aus diesem Bereich einzugehen, verweisen wir stattdessen auf die entsprechenden Rubriken in der (MSC2000)-Klassifikation. Man wird hierzu fündig unter 49-XX, 34H05, 34K35, 65Kxx, 90Cxx und 93-XX.

Zeit: diskret / kontinuierlich

Wie schon in (1) klar wird, ist der Setup, in dem wir auch während der gesamten Arbeit bleiben werden, zeitdiskret, während in den meisten Lehrbüchern zur Kontrolltheorie vorrangig der zeitkontinuierliche Fall behandelt wird.

Führt der zeitdiskrete Ansatz zwanglos zu — möglicherweise stochastischen — Differenzgleichungen, so kommt man im zeitkontinuierlichen Setup unweigerlich zu gewöhnlichen [stochastischen] Differentialgleichungen.

Ohne Robustheitsaspekte zu berücksichtigen, wird dieser Ansatz im stochastischen Bereich etwa in den Lehrbüchern von Kushner (1967; 1971) verfolgt, entsprechende Kapitel finden sich auch in Anderson und Moore (1990) und Chen (1985).

Zustandsraum: diskret / kontinuierlich

Eine weitere Unterscheidung ergibt sich aus der Frage, ob der Zustandsraum diskret oder — wie im Kontext dieser Arbeit — beliebig ist.

Für ersteren kann man — insbesondere, wenn es nur endlich viele Zustände x_i gibt, — ohne Mehraufwand alle Zustandsübergänge $\beta_t \rightarrow \beta_{t+1}$ durch die Angabe der Übergangverteilung $p_{i,j} = P(\beta_{t+1} = x_i | \beta_t = x_j)$ zulassen und muß sich nicht auf lineare Übergänge beschränken.

Dies führt auf die Theorie der Hidden Markov Models, wie sie sich zum Beispiel in Elliott et al. (1995) dargestellt findet. Allgemeinere Zustandsräume werden dabei zwar erfaßt, sind aber wesentlich schwieriger zu behandeln, weil man dort

die Theorie der Markov-Prozesse auf beliebigen Zustandsräumen³ benötigt. So ist auch ein Theorem über die LAN-Eigenschaft von HMM's auf beliebigen Zustandsräumen von Jensen und Petersen (1999) in unserem Kontext nicht anwendbar, da dort ein kompakter Zustandsraum gefordert wird.

Übergänge: linear / nichtlinear

Wie in (1) und (2) ersichtlich ist, beschränken wir uns auf lineare Zustandsräume. In der Literatur werden aber — durchaus auch für beliebige Zustandsräume — auch nichtlineare Übergänge und Beobachtungen betrachtet, etwa als *physikalisches Modell* in Anderson und Moore (1979), pp. 195–204, bei dem man

$$\beta_t = \tilde{F}_t(\beta_{t-1}) + v_t, \quad y_t = \tilde{Z}_t(\beta_t) + \varepsilon_t \quad (1.8)$$

mit differenzierbaren \tilde{F}_t , \tilde{Z}_t unter (V3) betrachtet. Für diese Modelle verwendet man dann etwa den *Extended Kalman-Filter*, $\beta_{t|t}^{\text{eKK}}$, in dem man im Korrekturschritt \tilde{F}_t und \tilde{Z}_t jeweils an der Stelle $\beta_{t|t-1}^{\text{eKK}}$ linearisiert,

$$F_t \rightsquigarrow \tilde{F}_t|_{\beta_{t|t-1}^{\text{eKK}}}, \quad Z_t \rightsquigarrow \tilde{Z}_t|_{\beta_{t|t-1}^{\text{eKK}}}$$

setzt und dann den klassischen Kalman-Filter verwendet, und als Prädiktionschritt $\beta_{t+1|t}^{\text{eKK}} = F_t \beta_{t|t-1}^{\text{eKK}}$ setzt. Weitere Beispiele für nichtlineare Erweiterungen des Zustandsraummodells finden sich in West und Harrison (1989), Kap. 13.

Historischer Abriss

Ursprünglich kommt die Idee des Filterns laut Anderson und Moore (1979) aus dem Bereich des Elektro-Ingenieurwesens und war zunächst primär frequenzorientiert. Von statistischer Seite kamen dann wesentliche Anstöße von Kolmogoroff (1941), sowie von Wiener (1948; 1949), allerdings waren diese Ansätze an die Stationarität des Zustandsprozesses geknüpft. Ein entscheidender Durchbruch auch für nichtstationäre Zustandsprozesse waren schließlich die Arbeiten von Kalman (1960) und Kalman und Bucy (1961). Die dort vorgestellten Verfahren haben sich in der Folge in mehrerer Hinsicht als optimal herausgestellt, was Morris (1976) zu der These veranlaßt hat, der Kalman-Filter sei ein robuster Schätzer (c.f. Schick und Mitter (1994)).

In Wahrheit ist der Kalman-Filter allerdings nicht robust — was wir in dieser Arbeit auch nachweisen werden. Das eigentliche Ziel dieser Arbeit wird es daher sein, effizient berechenbare, robuste Alternativen zum Kalman-Filter zu entwickeln.

1.2 Problemstellung und klassische Lösung

1.2.1 Filter-/Glättungs-/Vorhersage-Problem

Wenn wir für den Vektor der Beobachtungen

$$y_{1:t} := (y_1^\top, \dots, y_t^\top)^\top \quad (1.9)$$

³vgl. auch Abschnitt 7.3

schreiben, so stellt sich das Problem, eine Schätzung für β_t auf Basis $y_{1:s}$ zu finden. Dabei gehen wir davon aus, daß wir $a_0 = E[\beta_0]$ kennen. Je nachdem, ob $s < t$, $s = t$ oder $s > t$ ist, sprechen wir vom *Vorhersage-*, vom *Filter-* oder vom *Glättungs-Problem*.

Verwenden wir die quadratische Verlustfunktion, so gilt es, den im Sinne des mittleren quadratischen Fehlers — im folgenden MSE — besten, $y_{1:s}$ -meßbaren Schätzer f_t für β_t zu finden, also

$$E |\beta_t - f_t(y_{1:s})|^2 = \min_{f_t} !. \quad (1.10)$$

Dieses Problem wird bekanntlich durch den bedingten Erwartungswert gelöst:

$$\hat{f}_t(y_{1:s}) = E[\beta_t | y_{1:s}] =: \beta_{t|s}. \quad (1.11)$$

Dieser ist aber im allgemeinen kaum praktikabel auszuwerten, und so reduziert man die Schwierigkeit, indem man nur lineare Schätzer zuläßt. Dies führt auf kleinste-Quadrate-optimale Verfahren, auf deren Herleitung wir noch näher eingehen.

Zur Ermittlung dieser optimalen, linearen Schätzer gibt es effiziente rekursive Verfahren, die unter dem Namen *Kalman-(Bucy)-Filter* bzw. *Kalman-Glätter* bekannt sind und auf die Arbeiten von Kalman (1960) und Kalman und Bucy (1961) zurückgehen.

1.2.2 der klassische Kalman-Filter und seine Glättungs- und Vorhersagependants

Kalman-Filter

Der klassische Kalman-Filter ist definiert als

$$\begin{aligned} \text{Initialisierung: } \beta_{0|0}^{\text{KK}} &= a_0, & \Sigma_{0|0}^{\text{KK}} &= Q_0 \\ \text{Prädiktion: } \beta_{t|t-1}^{\text{KK}} &= F_t \beta_{t|t-1}^{\text{KK}}, & \Sigma_{t|t-1}^{\text{KK}} &= F_t \Sigma_{t-1|t-1}^{\text{KK}} F_t^T + Q_t \\ \text{Korrektur: } M_t^{\text{KK}} &= \Sigma_{t|t-1}^{\text{KK}} Z_t^T [Z_t \Sigma_{t|t-1}^{\text{KK}} Z_t^T + V_t]^{-1} \\ \beta_{t|t}^{\text{KK}} &= \beta_{t|t-1}^{\text{KK}} + M_t^{\text{KK}} (y_t - Z_t \beta_{t|t-1}^{\text{KK}}) \\ \Sigma_{t|t}^{\text{KK}} &= \Sigma_{t|t-1}^{\text{KK}} - M_t^{\text{KK}} Z_t \Sigma_{t|t-1}^{\text{KK}} \end{aligned} \quad (1.12)$$

Dabei heißt M_t^{KK} auch *Kalman-Gain*, und $\Sigma_{t|t-1}^{\text{KK}}$ resp. $\Sigma_{t|t}^{\text{KK}}$ sind die Kovarianzen des Vorhersagefehlers resp. des Filterfehlers

$$\Delta \beta_t^{\text{KK}} := \beta_t - \beta_{t|t-1}^{\text{KK}}, \quad \Delta \beta_{t|t}^{\text{KK}} := \beta_t - \beta_{t|t}^{\text{KK}}. \quad (1.13)$$

Für die Größen M_t^{KK} und $\Sigma_{t|t}^{\text{KK}}$ ergeben sich auch die Darstellungen

$$(\Sigma_{t|t}^{\text{KK}})^{-1} = (\Sigma_{t|t-1}^{\text{KK}})^{-1} + Z_t^T V_t^{-1} Z_t, \quad (1.14)$$

$$M_t^{\text{KK}} = \Sigma_{t|t}^{\text{KK}} Z_t^T V_t^{-1} \quad (1.15)$$

Kalman–Vorhersage

Als Vorhersage für β_t aufgrund von $y_{1:s}$, $t - s > 1$ und deren Fehlerkovarianz ergeben sich folgende Terme, basierend auf dem Kalman–Filter

$$\begin{aligned}\beta_{t|s}^{\text{KK}} &= F_t \beta_{t-1|s}^{\text{KK}} \\ \Sigma_{t|s}^{\text{KK}} &= F_t \Sigma_{t-1|s}^{\text{KK}} F_t^\tau + Q_t\end{aligned}\tag{1.16}$$

Kalman–Glätter

Der klassische Kalman–Glätter und seine Fehlerkovarianz basieren ebenfalls auf dem Kalman–Filter und sind für $t = T, \dots, 1$ definiert als

$$\begin{aligned}\beta_{t-1|T}^{\text{KK}} &= \beta_{t-1|t-1}^{\text{KK}} + J_{t-1}(\beta_{t|T}^{\text{KK}} - \beta_{t|t-1}^{\text{KK}}) \\ J_{t-1} &= \Sigma_{t-1|t-1}^{\text{KK}} F_t^\tau (\Sigma_{t|t-1}^{\text{KK}})^{-1} \\ \Sigma_{t-1|T}^{\text{KK}} &= \Sigma_{t-1|t-1}^{\text{KK}} + J_{t-1}(\Sigma_{t-1|T}^{\text{KK}} - \Sigma_{t|t-1}^{\text{KK}}) J_{t-1}^\tau\end{aligned}\tag{1.17}$$

Wenn aus dem Zusammenhang klar, lassen wir das Superskript KK weg.

1.2.3 Eigenschaften des klassischen Kalman–Filters

zentrale algorithmische Eigenschaften des Kalman–Filters

An dieser Stelle halten wir schlagwortartig einige der unserer Ansicht nach zentralen Punkte des Kalman–Filters fest

- eine einfach verständliche Struktur, die in einen Initialisierungs–Schritt, einen Vorhersage–Schritt und einen Korrektur–Schritt zerfällt
- der Korrektur–Schritt läßt sich als lineare Funktion einfachst auswerten
- die gesamte Information aus der Vergangenheit, die für eine Schätzung eines zukünftigen Zustandes genutzt wird, ist im Wert von $\beta_{t|t-1}^{\text{KK}}$ zusammengefaßt
- der Korrektur–Schritt ist linear und daher nicht robust, denn y_t geht unbeschränkt in den Wert von $\beta_{t|t}^{\text{KK}}$ ein;

Diese Eigenschaften — bis auf die letzte — wollen wir versuchen, in unseren robusten Filtern so weit wie möglich beizubehalten.

Optimalität des Kalman–Filters unter allen linearen Schätzern

Für spätere Zwecke wiederholen wir an dieser Stelle die wesentlichen Argumente, die zeigen, daß der klassische Kalman–Filter optimal unter allen linearen Verfahren ist, wobei wir (V2)⁴ verlangen.

⁴für eine Darstellung bei korrelierten v_s , ε_t siehe Anderson und Moore (1979), Kap. 11

Sei dazu $L_2^k = L_2^k(P^{\{y_t\}, \{\beta_t\}})$, $k = p, q$ der Hilbertraum der $(\{y_t\}, \{\beta_t\})$ -meßbaren, \mathbb{R}^k -wertigen Funktionen f mit $\int |f|^2 dP < \infty$, und $\text{lin}_k(y_{1:t})$ der durch lineare Transformationen der Gestalt

$$a_0 + \sum_{j=1}^t A_j y_j, \quad A_j \in \mathbb{R}^{k \times q}, \quad a_0 \in \mathbb{R}^k, \quad (1.18)$$

aufgespannte, abgeschlossene Untervektorraum des L_2^k , sowie entsprechend $\text{lin}_k(y_{1:0})$ der durch die Konstanten im \mathbb{R}^k erzeugte, abgeschlossene Untervektorraum. Dann erhalten wir die orthogonale Zerlegung

$$\text{lin}_q(y_{1:t}) = \text{lin}_q(y_{1:t-1}) \oplus \text{lin}_q(\Delta y_t^{\text{KK}}), \quad (1.19)$$

wobei Δy_t^{KK} sich als orthogonaler Rest bei der Projektion von y_t auf $\text{lin}_q(y_{1:t-1})$ ergibt. Schreiben wir nun $\text{oP}(x|A)$ für die orthogonale Projektion von x auf $\text{lin}_k(A)$ mit k der Dimension von x , und setzen bereits voraus, daß sich $\beta_{t|t-1}^{\text{KK}}$ als $\text{oP}(\beta_t|y_{1:t-1})$ ergibt, so erhalten wir

$$\text{oP}(y_t|y_{1:t-1}) = Z_t \beta_{t|t-1}^{\text{KK}},$$

denn nach (2) ist $y_t = Z_t \beta_t + \varepsilon_t$, und nach (V2) $\varepsilon_t \perp \text{lin}_q(y_{1:t-1})$. Daher ist $\Delta y_t^{\text{KK}} = y_t - Z_t \beta_{t|t-1}^{\text{KK}}$, und die $(\Delta y_t^{\text{KK}})_t$ bilden ein Orthogonalsystem in L_2^q . Somit ergibt sich rekursiv

$$\beta_{0|0}^{\text{KK}} := \text{oP}(\beta_0^{\text{KK}}|y_{1:0}) = \text{E}[\beta_0] = a_0 \quad (1.20)$$

$$\beta_{t|t-1}^{\text{KK}} := \text{oP}(\beta_t^{\text{KK}}|y_{1:t-1}) = F_t \beta_{t-1|t-1}^{\text{KK}} \quad (1.21)$$

$$\begin{aligned} \beta_{t|t}^{\text{KK}} &:= \text{oP}(\beta_t^{\text{KK}}|y_{1:t}) = \text{oP}(\beta_t|y_{1:t-1}) + \text{oP}(\beta_t|\Delta y_t) = \\ &= \beta_{t|t-1}^{\text{KK}} + \underbrace{\text{oP}(\beta_t - \beta_{t|t-1}^{\text{KK}}|\Delta y_t)}_{=: \Delta \beta_t^{\text{KK}}} \end{aligned} \quad (1.22)$$

Gleichung (21) erhält man dabei, weil nach (1) $\beta_t = F_t \beta_{t-1} + v_t$ und nach (V2) $v_t \perp \text{lin}_p(y_{1:t-1})$ ist.

Um $\text{oP}(\Delta \beta_t^{\text{KK}}|\Delta y_t^{\text{KK}})$ zu bestimmen, betrachten wir das lineare Modell

$$\Delta y_t^{\text{KK}} = Z_t \Delta \beta_t^{\text{KK}} + \varepsilon_t. \quad (1.23)$$

Dann ergibt sich $\text{oP}(\Delta \beta_t^{\text{KK}}|\Delta y_t^{\text{KK}})$ mit den Momenten

$$\begin{aligned} \text{E} \Delta \beta_t^{\text{KK}} &= 0, \quad \text{E} \Delta y_t^{\text{KK}} = 0, \quad \text{Cov} \Delta \beta_t^{\text{KK}} = \Sigma_{t|t-1}^{\text{KK}} \\ \text{Cov}(\Delta y_t^{\text{KK}}, \Delta \beta_t^{\text{KK}}) &= Z_t \Sigma_{t|t-1}^{\text{KK}}, \quad \text{Cov} \Delta y_t^{\text{KK}} = Z_t \Sigma_{t|t-1}^{\text{KK}} Z_t^T + V_t \end{aligned}$$

als Lösung des Kleinste-Quadrate-Problems

$$\text{E} |\Delta \beta - M \Delta y|^2 = \min_{M \in \mathbb{R}^{p \times q}} ! \quad (1.24)$$

Nutzen wir nun die Normalengleichung aus, die durch die Minimalität im L_2 -Sinn impliziert wird, erhalten wir:

$$\begin{aligned} \Delta \beta - \hat{M} \Delta y &\perp M \Delta y \quad \forall M \\ \text{i.e.} \quad \text{E}[(\Delta \beta - \hat{M} \Delta y)^\tau (M \Delta y)] &= 0 \quad \forall M \end{aligned} \quad (1.25)$$

$$\begin{aligned} \Leftrightarrow \text{E}[(\Delta \beta - \hat{M} \Delta y)(\Delta y)^\tau] &= 0 \\ \Leftrightarrow \hat{M} &= \text{E}[(\Delta \beta)(\Delta y)^\tau] \text{E}[(\Delta y)(\Delta y)^\tau]^{-1} \\ &= \Sigma_{t|t-1} Z^\tau \{Z \Sigma_{t|t-1} Z^\tau + V\}^{-1} \end{aligned} \quad (1.26)$$

Bemerkung 1.2.1 Eine andere Art, den Ausdruck optimal im KQ-Sinn zu rechtfertigen, ergibt sich aus Lemma 3.1.1.

weitere Optimalitäten unter (V3), (V4)

- Es gilt $\beta_{t|s}^{\text{KK}} = \beta_{t|s}$:

Proposition 1.2.2 *Im Gaußschen Fall, also unter Verteilungsannahme (V3) bzw. (V4), ist der so bestimmte Kalman-Filter sogar optimal unter allen $y_{1:t}$ -meßbaren Verfahren.*

BEWEIS: Im Gaußschen Fall ist der bedingte Erwartungswert linear — vgl. Lemma A.2.1. ////

- $\beta_{t|t}^{\text{KK}}$ ist Posterior-Modus:

Proposition 1.2.3 *Im Gaußschen Fall, also unter Verteilungsannahme (V3) bzw. (V4), gilt*

$$\beta_{t|t}^{\text{KK}} = \underset{\beta}{\operatorname{argmax}} p^{\beta_t | y_{1:t} = y_{1:t}^h}(\beta; y_{1:t}^h) \quad (1.27)$$

BEWEIS: Im Gaußschen Fall fallen nach Proposition 1.2.2 Kalman-Filter und bedingter Erwartungswert zusammen. Andererseits ist die Dichte von $\beta_{t|t}^{\text{KK}}$ als Dichte einer Normalverteilung unimodal und symmetrisch um den (bedingten) Erwartungswert verteilt, so daß dieser mit dem Modalwert zusammenfällt. ////

- $\beta_{t|t}^{\text{KK}}$ ist ML-Schätzer in einem Regressionsmodell
Wie wir in Kapitel 3 zeigen werden, läßt sich der Kalman-Filter auch als ML-Schätzer eines gewissen Regressionsmodells gewinnen.

1.3 Charakterisierung der Linearität des bedingten Erwartungswertes

In Proposition 1.2.2 haben wir gesehen, daß der klassische Kalman-Filter als bester linearer Filter im Gaußschen Kontext sogar optimal unter allen $y_{1:t}$ -meßbaren Verfahren ist, weil der bedingte Erwartungswert $E[(\Delta)\beta | (\Delta)Y]$ in diesem Fall linear ist. Diese Eigenschaft wird sich auch als wesentlich für die Optimalität des rLS-Filters herausstellen, den wir in Kapitel 2 einführen werden.

Es stellt sich die Frage, ob diese Linearität an die Normalität gebunden ist, oder auch außerhalb von Normalverteilungsannahmen erreicht werden kann. Hierzu kann man folgende Antwort geben:

Theorem 1.3.1 *Seien X, ε unabhängige \mathbb{R}^p - bzw. \mathbb{R}^q -wertige Zufallsgrößen, $E[X] = 0$, $E[XX^\tau] = \Sigma$, $E[\varepsilon] = 0$, $E[\varepsilon\varepsilon^\tau] = V \succ 0$. Wir betrachten $Y := ZX + \varepsilon$ mit $0 \neq Z \in \mathbb{R}^{q \times p}$ bekannt, und es gelte $\operatorname{rk}(Z\Sigma Z^\tau + V) = q$.*

- (a) *Ist $E[X|Y] = MY$ mit einem $M \in \mathbb{R}^{q \times p}$, so gilt*

$$M = \Sigma Z^\tau (Z\Sigma Z^\tau + V)^{-1} \quad (1.28)$$

(b) Sind die Verteilungen $\mathcal{L}(\varepsilon + t)$, $t \in \text{supp}(\mathcal{L}(ZX))$ mit

$$\text{supp}(\mathcal{L}(ZX)) := \bigcap \{A \in \mathbb{B}^p : A = \bar{A}, P(ZX \in A) = 1\}$$

durch ein translationsinvariantes, σ -finites Maß $\mu_2 \in \mathcal{M}_\sigma(\mathbb{B}^q)$ dominiert, d.h. gibt es ein σ -finites Maß $\mu_1 \in \mathcal{M}_\sigma(\mathbb{B}^p)$, so daß $\forall A \in \mathbb{B}^p, B \in \mathbb{B}^q$

$$P(X \in A, Y \in B) = \int I_A(x) I_B(y) p^\varepsilon(y - Zx) p^X(x) \mu_1(dx) \mu_2(dy) \quad (1.29)$$

mit einer μ_1 -Dichte p^X und einer μ_2 -Dichte p^ε , und besitzt die charakteristische Funktion von ε keine Nullstellen in \mathbb{R}^q , so gibt es höchstens eine Verteilung $F_0 = F_0(\mathcal{L}(\varepsilon))$ von ZX , so daß der bedingte Erwartungswert linear ist.

Hinreichend für die Existenz einer solchen Verteilung ist die Existenz einer differenzierbaren, positiv semidefiniten Abbildung $f : \mathbb{R}^p \rightarrow \mathbb{R}$, so daß für alle $t \in \mathbb{R}^q$ und $i = 1, \dots, p$

$$\sum_{j=1}^q ((\mathbb{I}_p - MZ)^{-1} M)_{i,j} \frac{\partial}{\partial y_j} \log p^\varepsilon \Big|_t = \frac{\partial}{\partial x_i} \log f \Big|_{Z^\tau t}, \quad (1.30)$$

Dann kann P^X als Verteilung mit charakteristischer Funktion $\hat{p}^X := f$ gesetzt werden.

(c) Für $p = q = 1$ ist die Existenz einer Abbildung f mit (30) für beliebiges Σ äquivalent mit der Tatsache, daß $\varepsilon \sim P_\infty^\varepsilon$, P_∞^ε eine unendlich teilbare Verteilung mit endlicher Varianz, also (vgl. Loève (1960), p.292)

$$\log \hat{p}_\infty^\varepsilon(u) = c \int (e^{iux} - 1 - iux)/x^2 G(dx), \quad (1.31)$$

wobei $\hat{p}_\infty^\varepsilon$ die charakteristische Funktion von P_∞^ε ist, $0 < c < \infty$ eine Konstante, G irgendein Wahrscheinlichkeitsmaß auf \mathbb{B} .

In diesem Fall muß für einen linearen bedingten Erwartungswert X der Verteilung P_0^X mit charakteristischer Funktion $\hat{p}_0^X(t) = (\hat{p}_\infty^\varepsilon(t))^{(Z^2 \Sigma / V)}$ folgen.

(d) Gilt $\varepsilon \sim \mathcal{N}_q(0, V)$, so ist (29) erfüllt, und es gilt

$$E[X|Y] \text{ ist linear} \quad \iff \quad ZX \sim \mathcal{N}_q(0, Z\Sigma Z^\tau). \quad (1.32)$$

Bemerkung 1.3.2 (a) Wir haben nichts über die Verteilung von X eingeschränkt auf $\ker Z$ gesagt. Diese kann auch nicht eindeutig bestimmt sein, denn Y kann nicht unterscheiden zwischen $X_1 = X_{1,1} + X_{1,2}$ und $X_2 = X_{1,1} + X_{2,2}$ mit $X_{i,2} \in \ker Z$, $X_{1,1} \in (\ker Z)^\perp$.

(b) Daß es tatsächlich Verteilungen P^ε gibt, die keine linearen Erwartungswerte produzieren können, zeigt folgendes Beispiel: Sei $p = q = 1$, $Z = V = 1$, $\Sigma = 1/2$, $\varepsilon = \pm 1$ mit Wahrscheinlichkeit jeweils $1/2$. Dann müßte $q(t) = p^\varepsilon(t)^{\frac{1}{2}} = \cos(t)^{\frac{1}{2}}$ positiv semidefinit sein, mit der Folge, daß sich $\mathcal{L}(\varepsilon)$ auffassen lassen müßte als die Verteilung der Summe zweier unabhängiger, identisch verteilter Zufallsvariablen $\varepsilon_1, \varepsilon_2$. Dies kann nicht sein:

Zunächst müßte $\mathcal{L}(\varepsilon_1)$ rein diskret sein — sonst pflanzte sich ein zum Zählmaß singulärer Teil bei der Faltung fort. Der Träger von $\mathcal{L}(\varepsilon_1)$ müßte andererseits mindestens 2 Punkte umfassen; dann aber erhöht sich die Zahl der Trägerpunkte bei einer Faltung um mindestens einen.

BEWEIS: (a): Wegen der vorausgesetzten Linearität können wir M als Kleinste-Quadrate-Lösung ermitteln, und erhalten wie bei der entsprechenden Herleitung des klassischen Kalman-Filters das behauptete M als Lösung von

$$\mathbb{E} |X - MY|^2 = \min_M!$$

(b): Wegen der unterschiedlichen Dimensionen von X und Y benötigen wir eine verallgemeinerte Faltung:

Seien $f : \mathbb{R}^q \rightarrow \mathbb{R}$, $g : \mathbb{R}^p \rightarrow \mathbb{R}$, $f \in L_1(\mu_2)$, $g \in L_1(\mu_1)$. Dann definieren wir die Funktion $h : \mathbb{R}^q \rightarrow \mathbb{R}$, $y \mapsto h(y) := \int f(y - Zx)g(x) \lambda^p(dx)$ als Z -Faltung, in Zeichen $f *_Z g$. Es gilt wie üblich ein Faltungssatz für die Fouriertransformierten, die wir durch ein $\hat{\cdot}$ kennzeichnen:

$$\widehat{f *_Z g}(t) = \hat{f}(t) \hat{g}(Z^\tau t).$$

Für R^r -wertige Funktionen definieren wir entsprechende Z -Faltungen koordinatenweise.

Unter den Voraussetzungen des Theorems existiert eine gemeinsame $\mu_1 \otimes \mu_2$ -Dichte von X und Y , $p^{X,Y}(x,y) = p^\varepsilon(y - Zx)p^X(x)$, mit p^ε der μ_2 -Dichte, p^X der μ_1 -Dichte von X . Daher schreibt sich der bedingte Erwartungswert $\mathbb{E}[X|Y=y]$ $\mu_2(dy)$ -fast sicher als

$$\mathbb{E}[X|Y=y] = \frac{\int xp^\varepsilon(y - Zx)p^X(x)\mu_1(dx)}{\int p^\varepsilon(y - Zx)p^X(x)\mu_1(dx)}$$

und somit ist die vorausgesetzte Linearität nichts anderes als

$$\int (x - My)p^\varepsilon(y - Zx)p^X(x)\mu_1(dx) \stackrel{!}{=} 0 \quad [\mu_2(dy)] \quad (1.33)$$

mit M aus (28). Dies läßt sich auch $\mu_2(dy)$ -fast sicher schreiben als

$$M \int (y - Zx)p^\varepsilon(y - Zx)p^X(x)\mu_1(dx) \stackrel{!}{=} (\mathbb{I}_p - MZ) \int xp^\varepsilon(y - Zx)p^X(x)\mu_1(dx) \quad (1.34)$$

Unter den gemachten Voraussetzungen ist dabei stets (vgl. Lemma 7.1.1 (c)) $\mathbb{I}_p - MZ$ invertierbar, und es gilt mit $q^X(x) = xp^X(x)$, $q^\varepsilon(y) = yp^\varepsilon(y)$

$$(\hat{p}^X)'(s) = i\hat{q}^X(s), \quad s \in \mathbb{R}^p \quad (\hat{p}^\varepsilon)'(t) = i\hat{q}^\varepsilon(t), \quad t \in \mathbb{R}^q \quad (1.35)$$

Schreiben wir nun (34) mit Z -Faltungen, so erhalten wir

$$Mq^\varepsilon *_Z p^X \stackrel{!}{=} (\mathbb{I}_p - MZ)p^\varepsilon *_Z q^X \quad (1.36)$$

$$\iff (\mathbb{I}_p - MZ)^{-1}M\hat{q}^\varepsilon \hat{p}^X(Z^\tau \cdot) \stackrel{!}{=} \hat{p}^\varepsilon \hat{q}^X(Z^\tau \cdot)$$

$$\iff (\mathbb{I}_p - MZ)^{-1}M(\hat{p}^\varepsilon)' \hat{p}^X(Z^\tau \cdot) \stackrel{!}{=} \hat{p}^\varepsilon (\hat{p}^X)'(Z^\tau \cdot)$$

$$\iff (\mathbb{I}_p - MZ)^{-1}M(\log \hat{p}^\varepsilon)' \hat{p}^X(Z^\tau \cdot) \stackrel{!}{=} (\hat{p}^X)'(Z^\tau \cdot) \quad (1.37)$$

Hierbei dürfen wir wegen $\hat{p}^\varepsilon \neq 0$ in der vorletzten Äquivalenzumformung durch \hat{p}^ε teilen. Halten wir nun eine beliebige Richtung t_0 , $Z^\tau t_0 \neq 0$ fest, so handelt es sich um die gewöhnliche, homogene lineare Differentialgleichung in s mit Anfangswert $g(0) = \hat{p}^X(0) = 1$,

$$\{t_0^\tau Z(\mathbb{I}_p - MZ)^{-1}M(\log \hat{p}^\varepsilon)'(t_0 s)\} g(s) \stackrel{!}{=} g'(s) \quad \forall s.$$

Diese hat die eindeutige Lösung

$$\hat{p}_{t_0}^X(Z^\tau t_0 s) = g(s) = \exp \left\{ \int_0^s t_0^\tau Z(\mathbb{I}_p - MZ)^{-1} M(\log \hat{p}^\varepsilon)'(t_0 r) dr \right\}, \quad (1.38)$$

so daß auf $(\ker Z)^\perp$ maximal eine Lösung $p_{t_0}^X(Z^\tau t_0 s)$ existieren kann. Also existiert nach dem Eindeutigkeitssatz für Fouriertransformierte höchstens eine Verteilung von ZX , die zu einem linearen, bedingten Erwartungswert führt.

Hängt die Lösung $\hat{p}_{t_0}^X(Z^\tau t_0 s)$ nur über das Argument $Z^\tau t_0 s$ von t_0 ab, so existiert auch eine Lösung $\hat{p}_0^X(Z^\tau \cdot)$ auf ganz \mathbb{R}^q . Schreibt man hierzu (38) mithilfe der Kettenregel um, so erhält man Bedingung (30).

(c): Für $p = q = 1$ ist $Z \neq 0$, $(\mathbb{I}_p - MZ)^{-1} M = Z\Sigma/V$, und wir erhalten als notwendige Gestalt der charakteristischen Funktion von P^X

$$\hat{p}^X(s) = g(s/Z) = (\hat{p}^\varepsilon(s))^{(Z^2\Sigma/V)}.$$

Wegen der Stetigkeit der charakteristischen Funktion und des Stetigkeitssatzes dürfen wir ohne Einschränkung davon ausgehen, daß $d := (Z^2\Sigma/V) \in \mathbb{Q}$, also $d = d_1/d_2$, $d_1, d_2 \in \mathbb{N}$, $\text{ggT}(d_1, d_2) = 1$. Daß aber $\hat{p}^X(s)$ für beliebiges d , also beliebiges d_2 , die charakteristische Funktion eines Wahrscheinlichkeitsmaßes auf \mathbb{B} ist, ist nichts anderes als die Eigenschaft von P^ε , eine unendlich teilbare Verteilung zu sein; diese besitzt nach Loève (1960) die Normalform (31), wobei wir das dort verwendete Maß von endlicher Variation K als cG geschrieben haben.

(d): Für normalverteiltes ε wissen wir bereits, daß die charakteristische Funktion zu $\mathcal{N}_p(0, \Sigma)$, \hat{p}_0^X diese Differentialgleichung für jedes t_0 löst. ///

1.4 Nichtrobustheit des klassischen Verfahrens⁵

Der klassische Kalman–Filter und sein Glättungs–Pendant basieren auf zweiten Momenten der zugrundegelegten Verteilung. Dies stellt ein Robustheitsproblem dar, denn kleine Abweichungen von den Modellannahmen — klein im Sinn der schwachen Topologie der Verteilungskonvergenz — können die Qualität der Filter /Glätter /Vorhersagen stark beeinträchtigen.

Im Unterschied zur “u.i.v.” Situation können Ausreißer sehr unterschiedliche Effekte haben; in der Terminologie von Fox (1972) unterscheidet man daher AO– und IO–Ausreißer.

1.4.1 AO–Ausreißer

Es handelt sich um Ausreißer in den Beobachtungsfehlern ε_t , die also für den Zustandsprozeß keine Auswirkungen haben, so daß der Beobachtungsprozeß nur kurzfristig ausgelenkt wird. Ziel einer Behandlung von solchen Ausreißern muß es also sein, ihren Einfluß zu dämpfen.

Als Beispiel denke man sich die Steuerung eines Satelliten. Der Zustand des

⁵Teile dieses Abschnitts sowie die Graphiken in diesem und den nächsten drei Kapiteln sind mit finanzieller Unterstützung der Deutschen Forschungsgemeinschaft im Rahmen eines Forschungsaufenthaltes auf Einladung von Prof. Dr. Härdle im Oktober 1999 am SFB 373 (“Quantifikation und Simulation Ökonomischer Prozesse”) an der Humboldt Universität zu Berlin entstanden und bereits in Form eines Tutorials in Ruckdeschel (2000b) publiziert.

Systems seien die 3 Ortskoordinaten des Satelliten im Weltraum; die Kontrollstation am Boden empfangt Messungen über die aktuelle Position des Satelliten in Form eines möglicherweise verrauschten Signals. Ein AO-Ausreißer könnte nun durch eine wetterbedingte, kurzzeitige Überlagerung des Signals mit systemfremden Rauschen entstehen.

Ein guter Schätzer sollte in diesem Fall den Einfluß großer Beobachtungen auf die Zustandsschätzung dämpfen. Zu diesem Zweck werden wir in Kapitel 2 und Kapitel 3 mit dem rLS- und dem rIC-Filter robuste Alternativen zum Kalman-Filter einführen.

1.4.2 IO-Ausreißer

Ganz anders verhalten sich IO-Ausreißer. Hier wird der Zustandsprozeß selbst ausgelenkt, indem die Innovationen v_t kontaminiert werden. Es handelt sich also um Strukturbrüche, denen so schnell wie möglich gefolgt werden muß.

Um bei unserem Satelliten-Steuerungsbeispiel zu bleiben, stelle man sich vor, ein kleiner Asteroid treffe unseren Satelliten und lenke ihn damit vom Soll-Kurs ab. In diesem Fall will man so schnell wie möglich auf den neuen Kurs reagieren und versuchen, ihn zu kontrollieren.

Offenbar erschwert ein gedämpftes Signal ein schnelles Reagieren, so daß die simultane Behandlung von IO's und AO's im allgemeinen weniger effizient sein wird als jede der beiden für sich genommen.

Außerdem wird man im allgemeinen auch anhand einer Beobachtung plus Zustands-Schätzung basierend auf den bisherigen Beobachtungen schlecht entscheiden können, ob es sich bei einem großen Beobachtungssignal um einen IO oder einen AO handelt⁶. Daher empfiehlt es sich, für diese Aufgabe auf die strikte Rekursivität zu verzichten in dem Sinn, daß man sich bei der Zustandsschätzung auf die letzten $m + 1$ Beobachtungen y_t, \dots, y_{t-m+1} und die geglätteten/gefilterten/vorhergesagten Zustände $\beta_{t-i|t}$, $i = -1, \dots, m$. stützt und nicht nur auf y_t und $\beta_{t|t-1}$. In diesem Sinn wird in Kapitel 4 der mIC-Filter und Glätter definiert.

1.4.3 andere Ausreißer-Typen

Die Unterscheidung zwischen AO und IO ist bei weitem nicht die einzig mögliche. In der Literatur sind auch schon andere Typen betrachtet worden, wie in Mustern auftretende "patchy outliers" (PO)⁷ und substitutive Ausreißer (SO), auf die wir in Kapitel 9 noch näher eingehen werden. Bei den Robustifizierungen der Glätter in Kapitel 4 werden wir schließlich noch Level-Ausreißer (LO) betrachten.

Die wichtigsten Robustheitsprobleme werden aber bereits mit den Typen IO und AO voll abgedeckt.

1.4.4 Beispiele für AO's und IO's

Um einen Eindruck zu vermitteln, auf welche Weise AO's und IO's Daten im Zustandsraummodell beeinträchtigen können, haben wir Daten aus einem 10%-AO/IO-kontaminierten, normalverteilten Zustandsraum simuliert.

⁶vgl. auch West et al. (1985) p.79, erste Spalte

⁷vgl. Martin und Yohai (1986)

Beispiel 1.4.1 (Steady-State Modell) Wir spezifizieren das ideale Modell auf

$$\begin{aligned} p, q &= 1 \\ a_0 &= 0, \quad Q_0 = 0, \quad Q_t = Q = 1, \quad V_t = V = 1 \\ F_t &= F = 1, \quad Z_t = Z = 1 \end{aligned} \quad (1.39)$$

und realisieren AO-kontaminierte Daten, indem wir ε_t aus einer Konvex-Kontaminations-Umgebung mit Radius $r = 0.1$ um $\mathcal{N}(0, 1)$ simulieren, genauer

$$\varepsilon_t \sim 0.9\mathcal{N}(0, 1) + 0.1\mathcal{N}(10, 0.1),$$

und für die IO-kontaminierten Daten simulieren wir entsprechend

$$v_t \sim 0.9\mathcal{N}(0, 1) + 0.1\mathcal{N}(10, 0.1).$$

Eine graphische Darstellung von Pfaden der β - und y -Prozesse unter den verschiedenen Kontaminationen findet sich in Abbildung 1.1.

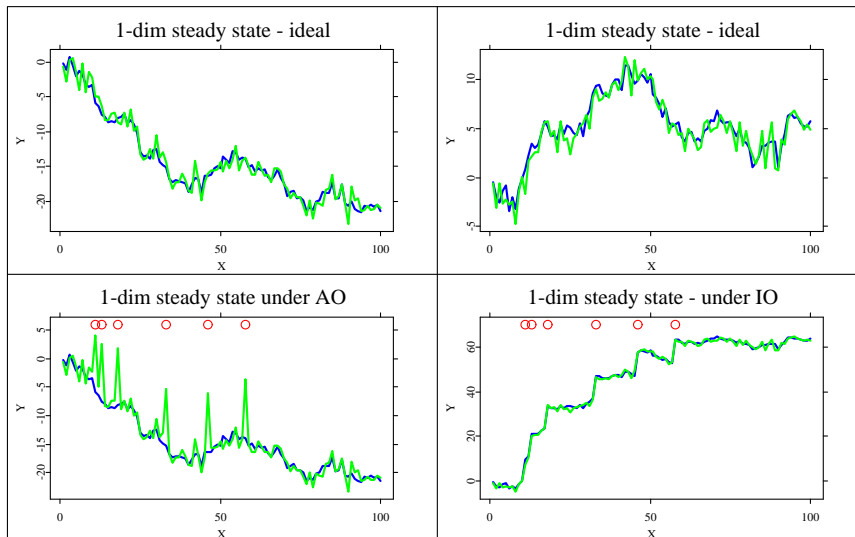


Abbildung 1.1: AO's und IO's im Steady State Modell:

Die AO- und die IO-kontaminierte Situation in Beispiel 1.4.1 sind jeweils unterhalb der entsprechenden idealen Situation dargestellt. Dabei sind die Zeitpunkte, an denen eine Kontamination stattfindet, durch Kreise markiert; AO's verursachen singuläre Peaks, IO's ziehen längeranhaltende Level-Changes nach sich. Grau gestrichelt sind dabei die Beobachtungen y_t , schwarz durchgezogen der Zustand β_t .

1.5 Ansätze zur Robustifizierung des Kalman-Filters

Sind die Verteilungsannahmen (V^*) nur annähernd erfüllt, oder liegen Ausreißer in den Daten vor, so ist damit zu rechnen, daß der klassische Kalman-Filter

sich zu stark auslenken läßt (bei AO's) oder zu träge auf einen Level-Change (IO) reagiert.

Dementsprechend kann man folgende Ansätze verfolgen, dieses Problem zu beheben⁸.

1.5.1 Robustheit in der Kontrolltheorie

Innerhalb der Kontrolltheorie ist in den letzten Jahren auch eine Robustheitstheorie entstanden; entsprechende Arbeiten hierzu finden sich etwa in Başar und Bernhard (1991) und Rotea und Khargonekar (1995). Allerdings wird hier Robustheit nicht als Stabilität gegen Abweichungen in den Verteilungen definiert — so diese überhaupt spezifiziert werden —, sondern man läßt nur beschränkte Steuerungen (L_∞ , bzw. \mathcal{H}_∞) zu und erhofft sich so eine Stabilität bei einer unvollkommen spezifizierten Übergangsfunktion ϕ .

1.5.2 Robustheit durch “Hard Rejection”

Masreliez und Martin (1977) berichten von der gängigen Praxis in der Flugzeugnavigation, sich eine Schätzung von $\Delta_t^{KK} = \text{Cov}[\Delta y_t^{KK}]$ zu verschaffen und dann anhand einer $3\sqrt{\Delta_t^{KK}}$ -Regel ausreißer-verdächtige Beobachtungen zu verwerfen. Schick und Mitter (1994) nennen in diesem Zusammenhang auch die Arbeit von Meyr und Spies (1984). Allerdings nutzen diese Verfahren nicht die Information aus, in welcher Richtung der Ausreißer zustande gekommen ist, so daß man in diesem Zusammenhang sicherlich effizientere Verfahren entwickeln kann, die ansonsten die gleichen Robustheitseigenschaften besitzen.

1.5.3 Robustheit durch v_t, ε_t mit stärkeren Flanken als in (V3) oder (V4)

Künstler (1995), p. 17, spricht von einem *robusten Modell*, sofern man “die Normalverteilung durch eine Verteilung mit breiteren Enden (‘heavy tails’) [ersetzt]”, also in unserem Kontext das ideale (Zentral-)Modell (V*) durch ein anderes ersetzt; das Ziel, ein Verfahren zu finden, das gleichmäßig über einer geeigneten Familie / Umgebung von Verteilungen um das ideale Modell herum “gut” ist, wird nicht in der Konstruktion des Verfahrens verwendet, sondern man hofft durch die Wahl eines entsprechend ungünstigen Zentralmodells bereits Robustheit über allen “günstigeren” Verteilungen zu erreichen. Dem ist auch so, allerdings um den Preis eines unnötig hohen Effizienzverlustes im idealen Modell — vgl. Kapitel 5. Dieser Ansatz wird in verschiedenen Ausprägungen verfolgt:

⁸Die Literatur zu diesem Thema ist sehr umfangreich — allein im Übersichtsartikel von Kassam und Poor (1985) sind bereits 209 Referenzen aufgeführt! Wir beschränken uns daher darauf, die verschiedenen Ansätze exemplarisch mit entsprechenden Arbeiten zu belegen. Für weitere Referenzen seien genannt: Ershov und Lipster (1978), Kassam und Poor (1985), Stockinger und Dutter (1987), Martin und Raftery (1987), Schick und Mitter (1994), Künsch (2001).

elliptisch–symmetrische Verteilungen

Szablowski (1990) ersetzt die Verteilungsannahmen (V^*) durch elliptisch–symmetrische Verteilungen mit fettschwänziger Verteilung des Betrages; interessanterweise ist bei deren Verwendung der bedingte Erwartungswert linear (c.f. Abschnitt 5 in Szablowski (1990)), so daß wir auch hier auf den klassischen Kalman–Filter stoßen, wenn wir die entsprechenden Varianzterme im Kalman–Filter durch Skalenterme ersetzen.

Bayes–scher Ansatz

Hierbei wird versucht, den Posterior–Erwartungswert $\beta_{t|t}$ längs nicht–normaler Verteilungen $\mathcal{L}(\beta_t, y_{1:t})$ zu ermitteln.

West (1981) schlägt hierzu die Verwendung von Verteilungen $\mathcal{L}(\varepsilon_t)$ mit stärkeren Flanken vor und spezifiziert diese in West (1984) auf Skalenmischungen von Normalverteilungen, was in West (1985) laut West et al. (1985) auf nicht–normale, ideale Modelle verallgemeinert wird.

Meinhold und Singpurwalla (1989) verwenden zur Robustifizierung multivariate Verallgemeinerungen von t –Verteilungen anstelle von Normalverteilungen, Ershov (1978)⁹, Ershov und Lipster (1978)⁹, Shirazi et al. (1988) und Peña und Guttman (1988) Konvexkombinationen von Normalverteilungen für die Verteilung der ε_t , von denen eine eine “größere” Kovarianz als im idealen, unkontaminierten Modell besitzt. Üblicherweise realisiert man solche Konvexkombinationen, indem man einen stochastisch unabhängigen Auswahlmechanismus U_t entscheiden läßt, ob Kontamination vorliegt oder nicht.

Ershov und Lipster (1978)⁹ und Ershov (1978)⁹ unterscheiden sich darin von Peña und Guttman (1988) — und auch von Shirazi et al. (1988), daß erstere eine Schätzung für U_t durchführen und anhand dieser dann entscheiden, aus welcher der beiden Komponenten der Konvexkombination die Daten stammen, in diesem Sinn also adaptiv filtern, während Peña und Guttman (1988) und Shirazi et al. (1988) tatsächlich $\mathcal{L}((1 - U_t)\varepsilon_t + U_t\tilde{\varepsilon}_t)$ verwenden.

Die Vorgehensweise, Konvexkombinationen zweier Normalverteilungen zu betrachten, ist ein Spezialfall des *Gaussian–Sum–Filters*¹⁰, bei dem man Konvexkombinationen endlich vieler Normalverteilungen betrachtet¹¹, um dadurch beliebige, absolutstetige, nicht–Gaußsche Verteilungen in Totalvariation zu approximieren. Die Konvexkombinationen können dabei entsprechend mit Multinomialverteilungen $U_t \sim \text{mult}(1, (p_1, \dots, p_{m-1}, 1 - \sum p_i)^\tau)$ statt Bin(1, p)–Verteilungen als Auswahlmechanismus realisiert werden. Tanaka und Katayama (1987) gehen in einem solchen Setup analog zu Ershov (1978) und Shirazi et al. (1988) vor, indem sie U_t — mithilfe eines Posterior–Modus–Schätzers schätzen.

Numerische Verfahren zur Berechnung dieser nichtlinearen, bedingten Erwartungswerte finden sich in Tsai und Kurz (1983), Kitagawa (1987) und Chow (1994). Kitagawa (1987) verwendet stückweise lineare Funktionen zur Approximation einer speziellen nicht–normalen Fehlerverteilung und entscheidet sich für diese Fehlerverteilung aufgrund des AIC–Kriteriums, dessen Verwendung aber in diesem Kontext nicht theoretisch abgesichert ist. Tsai und Kurz (1983) hingegen stützen sich auf stückweise polynomiale Approximation. Mit dem Ziel

⁹laut Schick und Mitter (1994)

¹⁰vgl. Anderson und Moore (1979), pp 211–221; weitere Referenzen hierzu siehe dort;

¹¹wiederum als diskrete Variante der Skalenmischungen von West (1984) interpretierbar

einer IO–Robustifizierung gibt Chow (1994) in seinem AGSF¹² eine effizient berechenbare Näherung an den Gaussian–Sum–Filter an.

Posterior–Modus

Posterior–Modus–Schätzer¹³ werden für den unbekanntem Zustand bei nicht-normalverteilten Setups seit langem verwendet — vgl. Sage (1970), Maltz (1972).

Diesen Ansatz greifen Fahrmeir und Kaufmann (1991), Künstler (1995), Fahrmeir und Künstler (1999) auf, um durch Verwendung entsprechend fettschwänzigerer Fehler– resp. Innovationsverteilungen AO– resp. IO–robuste Glätter und Filter zu erhalten. Als numerische Verfahren zur Lösung des Maximierungsproblems wird hierbei ein speziell an die Situation angepaßter Gauß–Newton Algorithmus mit einer Block–Cholesky–Zerlegung der Fisher–Information verwendet — vgl. Abschnitt 4 in Fahrmeir und Kaufmann (1991).

Çelebi und Kurz (1996) schlagen unabhängig von den letztgenannten Arbeiten verschiedene lokale Näherungen zur Berechnung des Posterior–Modus vor, wobei sie zusätzlich auch rein rekursiv definierte Verfahren betrachten.

1.5.4 Robustheit durch Verwenden parallel laufender Filter

ohne Samplingtechniken

Bei den adaptiven Ansätzen, wie denen von Ershov (1978) und Ershov und Lipster (1978), Tanaka und Katayama (1987) müssen immer mehrere Filter parallel betrachtet werden. Ähnlich geht auch Pupeikis (1998) vor, indem er in jedem Schritt t einen optimalen Filter aus einem solchen Bündel paralleler Filter wählt. Auch beim Auswerten des Filters von Schick (1989) laufen immer n Filter parallel, jedoch ist die Motivation hierfür eine andere — siehe weiter unten. Dasselbe gilt für die Lösung von Birmiwal und Shen (1993) für das Filterproblem: Indem man die Lösungen für alle Kontaminationsvergangenheiten bestimmen muß, sind in jedem Schritt t auch t verschiedene Filter zu betrachten. Vor allem der Filter von Birmiwal und Shen (1993) und der exakte Filter von Schick (1989) sind wegen der wachsenden Zahl parallel zu betrachtender Filter keinesfalls “on–line” anwendbar.

mit Samplingtechniken

In Carlin et al. (1992), Carter und Kohn (1994), Künstler (1995), Kapitel 5, Carter und Kohn (1996), Hürzeler (1998), Künsch (2001) werden Samplingtechniken beschrieben, mit denen man den bedingten Erwartungswert in einer nicht–Gaußschen Situation näherungsweise ermitteln kann. Dazu werden mit Simulationstechniken Daten aus der Verteilung $\mathcal{L}(\beta_t|Y_{1:t[-1]})$ gezogen.

Carlin et al. (1992), Carter und Kohn (1994) und Künstler (1995) verwenden hierzu *Markov Chain Monte Carlo (MCMC)*–Verfahren, genauer den *Gibbs Sampler* und setzen dieses Verfahren in kontaminierten Situationen ein. Allerdings ist es nicht rekursiv, denn man braucht hierzu¹⁴ eine Stichprobe aus

¹²Approximate Gaussian Sum Filter

¹³oft auch MAP für *Maximum à Posteriori*

¹⁴zumindest für $1 < t < T$, T der Beobachtungshorizont

$\mathcal{L}(\beta_t | \beta_{t+1}, \beta_{t-1}, Y_{1:t[-1]})$, und ist daher für das “on–line”–Filtern ungeeignet. In Hürzeler (1998) und Künsch (2001) finden sich für diesen Zweck rekursive Verfahren basierend auf dem *Particle Filter*. Diese verwendet Hürzeler (1998) in seinem Beispiel 8.2 für eine robuste Zustandsschätzung im AO–kontaminierten Zustandsraum. Trotzdem handelt es sich auch hierbei um sehr rechenaufwendige Verfahren, und es ist nicht klar, ob sie “on–line”–tauglich sind.

1.5.5 Robustheit mit robusten Verfahren aus anderen Kontexten

Eine ganze Reihe von Arbeiten überträgt Robustheitskonzepte aus einem anderen Kontext, wie dem Regressions– oder dem Lokationsmodell in den Kontext des Zustandsraummodells, ohne aber im Modell selbst die entsprechenden (optimalen) Robustheitseigenschaften nachzuweisen. Tatsächlich kann man anhand von Simulationsstudien in vielen Fällen diese Robustheit belegen, die Frage der Optimalität, also der Effizienz unter allen gleich robusten Verfahren bleibt aber offen¹⁵.

M–Schätzer im Regressionsmodell

In einer Serie von Arbeiten, Boncelet und Dickinson (1983; 1987), Boncelet (1985), verwenden die Autoren die formale Darstellung des Zustandsraummodells als Regressionsmodell mit stochastischem Parameter — vgl. auch Kapitel 3 —, um darauf aufbauend dann Techniken aus der robusten Regression zum Filtern / Glätten / Vorhersagen zu verwenden. Unabhängig zu diesen Arbeiten gehen Cipra und Romera (1991) analog vor, gelangen aber zu einem in einer unterschiedlichen Norm gestutzten M–Schätzer — vgl. Bemerkung 3.2.1.

L–Schätzer

Die Verwendung von Ordnungsstatistiken beim Glätten und Filtern geht auf Tukey und dessen 3R–Glätter, ein *running median*, zurück — vgl. Tukey (1977). Derartige Filter erfreuen sich vor allem in der Spracherkennung, bei der Bildrekonstruktion und bei der Mustererkennung großer Beliebtheit, vgl. die Übersichtsarbeit von Gabbouj et al. (1992), sowie Arce et al. (1998) und Barner und Arce (1998).

Im Zustandsraummodell haben Kirilin und Moghaddamjoo (1986) eine Modifikation des Kalman–Filters vorgeschlagen, bei dem sie als “adaptive Biaskorrektur” die Mediane der geschätzten Beobachtungsfehler und Innovationen $\varepsilon_{t|t}$ und $v_{t|t}$ im Korrekturschritt mitverarbeiten. Über die Robustheit dieses Verfahrens werden aber keine theoretischen Resultate nachgewiesen.

1.5.6 Minimax–Robustheit

Eines der wichtigsten Konzepte der robusten Statistik besteht darin, Verfahren zu entwerfen, deren ungünstigstes Verhalten längs einer Umgebung des idealen Modells optimal ist.

¹⁵Auch wir werden in Kapiteln 2–4 die Verfahren rLS, rIC, mIC zunächst mit Analogieschlüssen einführen, dann aber in Kapitel 8 noch einmal gesondert der Frage der Optimalität nachgehen.

Robustheit über reduzierten Umgebungen

Ein erster Ansatz, die Stabilität eines Verfahrens unter Abweichungen von den Modellannahmen zu analysieren, ist es sicher, das Verhalten des Verfahrens unter reduzierten Umgebungen wie parametrischen Umgebungen oder Umgebungen mit einer Varianzschranke zu untersuchen und gegebenenfalls dort das Verhalten im "worst case" zu optimieren. Bei einer Schranke an die Varianz erhält Morris (1976) wieder den Kalman–Filter als minimax–optimal längs dieser Teilumgebung; daß diese offensichtlich nicht den Anforderungen der Praxis genügt, belegt Abbildung 1.1 in Beispiel 1.4.1.

Parametrische Abweichungen, also Störungen der Hyperparameter F_t , Z_t , Q_t , V_t werden u.a. von Chen und Patton (1996) und Bolzern et al. (1996) betrachtet, bleiben aber bei linearen Verfahren und sind daher in unserem Sinn nicht robust.

Robustheit über Umgebungen im Frequenzbereich

Bei stationären Zustandsräumen¹⁶ können wir auch Umgebungen um das ideale Spektralmaß legen, und Minimax–Lösungen auf diesen Umgebungen gewinnen. Kassam und Poor (1985) führen Kassam und Lim (1977) als erste Referenz für diesen Ansatz an, Existenzfragen von Sattelpunkten sind in Franke (1985), Franke und Poor (1984), Abschnitte 2, 3 und Verdú und Poor (1984) bearbeitet. Für weitere Referenzen, sowie eine ganze Reihe von Lösungen zu speziellen Umgebungen verweisen wir auf Kassam und Poor (1985).

Der Nachteil dieses Ansatzes besteht zum einen in der Annahme einer Schranke an die Varianz, die sich hier als gemeinsame Schranke für die Totalmasse der Spektralmaße in der Umgebung äußert. Nach der Theorie von Wiener und Kolmogoroff müssen die auf diese Weise gewonnenen Minimax–Verfahren linear sein, sind also auch nicht robust in unserem Sinn. Zum anderen sind von vornherein nur stationäre Modelle zugelassen, ein Umstand an der Theorie von Wiener und Kolmogoroff, der zur Entwicklung des Kalman–Filters geführt hat.

Robustheit über Umgebungen beliebiger Verteilungen

Die erste Arbeit mit einem Minimax–Resultat im Zeitbereich über beliebigen Verteilungen ist Masreliez und Martin (1977). Der dort hergeleitete Filter wird in Martin (1979) als *ACM–Filter*¹⁷ bezeichnet und zum ACM–type–Filter verallgemeinert.

Franke und Poor (1984) behandeln in ihrem Abschnitt 4 die Frage der Existenz eines Sattelpunktes im Filter– und Vorhersageproblem, untersuchen aber keine der in dieser Arbeit verwendeten Umgebungen.

Schick (1989) und Schick und Mitter (1994) liefern für AO–Kontamination Argumente für eine Nähe von $\mathcal{L}(\beta_t|y_{1:t-1})$ zu einer Normalverteilung, und Birmiwal und Shen (1993) und Birmiwal und Papantoni-Kazakos (1994) verwenden — ohne diese so zu nennen — SO–Umgebungen¹⁸ und geben für diese explizite Formeln für den Sattelpunkt an.

Alle diese letztgenannten Ansätze werden wir noch detaillierter in Abschnitt 8.1.4 besprechen.

¹⁶zeitinvariante Modelle mit $\rho(F) < 1$

¹⁷für *approximate conditional mean*

¹⁸vgl. Abschnitt 8.1.3

Kapitel 2

der rLS–Filter

2.1 Motivation

Beim ersten Ansatz¹ zur Robustifizierung setzen wir an der Herleitung des klassischen Kalman–Filters als bestem linearen Verfahren in Abschnitt 1.2.3 an. Wir nehmen daher stets zumindest (V2) an.

Ziel ist es dabei, ein strikt rekursives Verfahren zu konstruieren, das also die gesamte Information aus der Vergangenheit im Wert einer Zufallsgröße $\beta_{t|t-1}$ zusammenfaßt. Dabei beschränken wir uns auf die Behandlung von AO–Ausreißern, da es sich, wie wir in Abschnitt 1.4.2 festgehalten haben, bei IO’s empfiehlt, auf diese strikte Rekursivität zu verzichten.

Offenbar ist der Einfluß eines AO–Ausreißers in den Beobachtungen im Korrekturschritt des klassischen Kalman–Filters in der linearen, orthogonalen Projektion

$$\text{op}(\Delta\beta_t^{\text{KK}} | \Delta y_t^{\text{KK}}) = M_t^{\text{KK}} \Delta y_t^{\text{KK}}$$

unbeschränkt, so daß man hier mit einer Robustifizierung anzusetzen hätte; beim Prädiktions–Schritt hingegen gehen keine potentiellen AO–Ausreißer ein, da sich hier die Beobachtungsbasis nicht ändert und nur der Zustand mithilfe einer (unkontaminierten) Innovation fortgeschrieben wird. Dies legt folgende Robustifizierung nahe:

2.2 Robustifizierung

Als Robustifizierung von $M_t^{\text{KK}} \Delta y_t^{\text{KK}}$ bietet es sich an, die aus anderen Zusammenhängen bekannte Form

$$H_b(M\Delta y) := M\Delta y \min\left\{1, \frac{b}{|M\Delta y|}\right\} \quad (2.1)$$

zu verwenden und eventuell die euklidische Norm im \mathbb{R}^p als Verlustfunktion zu ersetzen durch

$$\rho_K(x) := \begin{cases} |x|^2, & |x| \leq K \\ 2K|x| - K^2 & |x| > K \end{cases} \quad (2.2)$$

¹Die Ansätze rLS, rIC, L5 und mIC sind in enger Zusammenarbeit mit H. Rieder und V. Azmjakov im Rahmen des DFG–Projektes DFG Ri 332/8–1 im Zeitraum 1997–99 entstanden.

Man beachte, daß sich für $K = \infty$ wieder die euklidische Norm ergibt.

2.2.1 robustes Optimierungsproblem

Als robustes Optimierungsproblem erhalten wir nun

$$F(M) := \mathbb{E} \rho_K(\Delta\beta - H_b(M\Delta y)) = \min_{M \in \mathbb{R}^{p \times q}} ! \quad (2.3)$$

Die Lösung bezeichnen wir mit M^{rLS} .

Bemerkung 2.2.1 Für normalverteilte ε_t , v_t und $b = \infty$ hat die Einführung von ρ_K keinen Einfluß auf M^{rLS} , genauer gilt

$$M^{\text{rLS}}|_{K=0} = M^{\text{rLS}}|_{K=\infty} = M^{\text{KK}} = M^{\text{rLS}}|_{0 < K < \infty}.$$

Hierbei können wir wegen der Linearität von $M\Delta y$ koordinatenweise argumentieren, und uns daher auf den Fall $p = 1$ beschränken. Dann ist $\Delta\beta_{t|t}(M) := \Delta\beta - M\Delta y \sim \mathcal{N}_1(0, \Sigma_{t|t})$ mit

$$\Sigma_{t|t} = \Sigma_{t|t}(M) = \Sigma_{t|t-1} - M Z_t \Sigma_{t|t-1} - \Sigma_{t|t-1} Z_t^T M^T + M(Z_t \Sigma_{t|t-1} Z_t^T + V_t) M^T.$$

Nun ist $\mathbb{E} \rho_K(\Delta\beta_{t|t})$ für jedes K eine monoton steigende Funktion in $\Sigma_{t|t}$, so daß das Minimum am minimalen Wert von $\Sigma_{t|t}$, also dem Kalman-Filter, angenommen wird.

2.2.2 der rLS-Filter

Insgesamt ergibt sich, nachdem für jedes t die Matrix M_t^{rLS} gefunden worden ist, folgender, rekursiv zu berechnender Filter, den wir *rLS-Filter* für *robust Least Squares* nennen wollen:

$$\begin{aligned} \beta_{0|0}^{\text{rLS}} &= \mathbb{E}[\beta_0] = a_0 \\ \text{und für } t > 0 \quad \beta_{t|t-1}^{\text{rLS}} &= F_t \beta_{t-1|t-1}^{\text{rLS}} \\ \beta_{t|t}^{\text{rLS}} &= \beta_{t|t-1}^{\text{rLS}} + H_b(M_t^{\text{rLS}} \Delta y_t^{\text{rLS}}), \end{aligned} \quad (2.4)$$

wobei $\Delta\beta^{\text{rLS}}$, Δy^{rLS} analog zu den Größen $\Delta\beta^{\text{KK}}$, Δy^{KK} definiert sind, nur daß sich der $\beta_{t|t-i}^{\text{rLS}}$ -Prozess wie in (4) definiert aufbaut. Dieses Vorgehen hätte folgende Eigenschaften:

2.2.3 algorithmische Eigenschaften des rLS-Filters

- eine einfach verständliche Struktur, die in einen Initialisierungs-Schritt, einen Vorhersage-Schritt und einen Korrektur-Schritt zerfällt
- der Korrektur-Schritt läßt sich als (fast) lineare Funktion schnell auswerten
- die gesamte Information aus der Vergangenheit, die für eine Schätzung eines zukünftigen Zustandes genutzt wird, ist im Wert von $\beta_{t|t-1}^{\text{rLS}}$ zusammengefaßt
- der Einfluß von y_t im Korrektur-Schritt ist beschränkt, AO's haben also nur begrenzte Auslenkungsmöglichkeiten

2.2.4 Wahl von b und K

Wie bei den allgemeinen Influenzkurven vom Hampel–Krasker–Typ stehen wir vor dem Problem ein Kriterium für die Tuning–Parameter b und K anzugeben. Hierzu schlagen wir vor, die durchschnittliche “Größe”, genauer die Quadrate der L_2 -Normen zu betrachten:

Wir geben also eine relative Abweichung $\delta > 0$ für die Spuren der Kovarianzmatrizen vor:

$$\mathbb{E}[|\Delta\beta - H_b(M^{\text{rLS}} \Delta y)|^2] \stackrel{!}{=} (1 + \delta) \mathbb{E}[|\Delta\beta - M^{\text{KK}} \Delta y|^2] \quad (2.5)$$

Diese Gleichung kann man auch als Anscombe²-Versicherungskriterium lesen: *Wieviel an relativer Effizienz-minderung gegenüber dem klassisch optimalen Verfahren — dem Kalman-Filter — bin ich bereit, im idealen Modell als “Versicherungsprämie gegen Ausreißer” zu bezahlen, damit sich das Verfahren auch unter Abweichungen vom idealen Modell “vernünftig” verhält?*

Dieser Effizienzverlust im idealen Modell bemißt sich in einer relativen Verschlechterung des erreichbaren MSE.

Was K anlangt, wählen wir den Effekt, der durch das Herunterwichten mit ρ_K entsteht, so daß

$$(1 + \delta) \mathbb{E}[\rho_K(X - M^{\text{KK}} Y)] = \mathbb{E}[|X - M^{\text{KK}} Y|^2] \stackrel{!}{=} \Sigma_{t|t}. \quad (2.6)$$

2.3 Berechnung der Tuningkonstanten b und K

2.3.1 vereinfachte Bestimmung von b

Weil die Berechnung des optimalen M^{rLS} nur sehr mühsam zu bewerkstelligen ist, schlagen wir unter Verteilungsannahmen (V3) bzw. (V4) vor, (5) zu ersetzen durch

$$\mathbb{E}[|\Delta\beta - H_b(M^{\text{KK}} \Delta y)|^2] \stackrel{!}{=} (1 + \delta) \mathbb{E}[|\Delta\beta - M^{\text{KK}} \Delta y|^2] = (1 + \delta) \Sigma_{t|t}^{\text{KK}} \quad (2.7)$$

bzw. im eindimensionalen

$$\mathbb{E}[|\Delta\beta - H_b(\tilde{M}^{\text{KK}} \Delta y)|^2] \stackrel{!}{=} (1 + \delta) \mathbb{E}[|\Delta\beta - \tilde{M}^{\text{KK}} \Delta y|^2] = (1 + \delta) \tilde{\Sigma}_{t|t}^{\text{KK}}. \quad (2.8)$$

Dabei sind M^{KK} und $\Sigma_{t|t}^{\text{KK}}$ die Größen aus dem klassischen Kalman-Filter (12), und die $\tilde{\cdot}$ -Größen stammen aus dem Kalman-Filter, der berücksichtigt, daß wir mit der Robustifizierung jeweils im Korrekturschritt im idealen Modell eine höhere Varianz des Filterfehlers in Kauf nehmen. Dies schlägt sich in einer Aufdatierung von $\tilde{\Sigma}_{t|t}^{\text{KK}}$ als

$$\tilde{\Sigma}_{t|t}^{\text{KK}} = (1 + \delta) (\tilde{\Sigma}_{t|t-1}^{\text{KK}} - \tilde{\Sigma}_{t|t-1}^{\text{KK}} Z_t^T (Z_t \tilde{\Sigma}_{t|t-1}^{\text{KK}} Z_t^T + V_t)^{-1} Z_t \tilde{\Sigma}_{t|t-1}^{\text{KK}}) \quad (2.9)$$

nieder.

²vgl. Anscombe (1960)

2.3.2 numerische Bestimmung der Tuningkonstanten

Verwendet man Näherung (7) bzw. (8), so kann man b und K ohne Kenntnis von M^{rLS} ermitteln. Für die auftretenden Integrale $\delta E[\rho_K(X - M^{\text{KK}} Y)]$ und $E[|\Delta\beta - H_b(M^{\text{KK}} \Delta y)|^2]$ schlagen wir je nach Dimension numerische oder Monte-Carlo-(MC-) Integration mit hinreichend vielen Stützstellen vor, wobei wir unter (V3) bzw. (V4) — ebenfalls vereinfachend — so tun, als sei

$$\Delta\beta_t^{\text{rLS}} \sim \mathcal{N}_p(0, \Sigma_{t|t-1}^{\text{KK}}[\tilde{\Sigma}_{t|t-1}^{\text{KK}}]).$$

In den Beispielen in `xploRe` haben wir für Dimension $p = 1$ im Quantlet `r1sbnorm1` ein Romberg-Verfahren mit 100 Stützstellen und für $p > 1$ in `r1sbnorm` MC-Integration mit 10000 Stützstellen verwendet.

Die angegebenen Kriterien sind jeweils monoton in b respektive K , so daß sich Bisektionsverfahren zu deren Ermittlung anbieten.

2.3.3 Resultat für $t = 1$

Für den Fall, daß $(\Delta\beta, \Delta y)$ linear ist — vgl. auch Theorem 1.3.1 —, also zum Beispiel unter (V3) bzw. (V4) für den rLS-Filter und den später noch eingeführten rIC-Filter im Zeitpunkt $t = 1$, läßt sich noch mehr aussagen; dann nämlich gilt folgende Proposition:

Proposition 2.3.1 Falls $E[\Delta\beta|\Delta y] = M\Delta y$, so löst M auch folgendes Problem

$$E|\Delta\beta - M\Delta y \min\{1, \frac{b}{|M\Delta y|}\}|^2 = \min_{M \in \mathbb{R}^{p \times q}} ! \quad (2.10)$$

BEWEIS: Wegen der Projektoreigenschaft des bedingten Erwartungswertes gilt für jedes Δy -meßbare f

$$\begin{aligned} E|\Delta\beta - f|^2 &= E|\Delta\beta - E[\Delta\beta|\Delta y]|^2 + E|E[\Delta\beta|\Delta y] - f|^2 \\ &= E|\Delta\beta - M\Delta y|^2 + E|M\Delta y - f|^2. \end{aligned}$$

Nach Lemma C.2.2 wird das Problem

$$E|\Delta\beta - f(\Delta y)|^2 = \min_f ! \quad \text{NB } |f| \leq b \quad (2.11)$$

daher durch $\hat{f} = M\Delta y \min\{1, \frac{b}{|M\Delta y|}\}$ gelöst. Umgekehrt ist auch jedes $f_A = A\Delta y \min\{1, \frac{b}{|A\Delta y|}\}$ zulässig für (11), so daß \hat{f} tatsächlich auch (10) löst. $////$

2.4 Varianten

Ausgehend vom rLS-Filter in (4) bieten sich folgende Varianten an, die sich von (4) in der Verwendung anderer Matrizen M und anderer Normen zur Messung einer “großen” Beobachtungsinnovation Δy unterscheiden.

2.4.1 Hybridfilter

Weil die Berechnung von M^{rLS} sehr aufwendig ist, bietet es sich an, stattdessen gleich M^{KK} zu verwenden, insbesondere, weil die Resultate im folgenden Abschnitt suggerieren werden, daß bei Verteilungsannahmen (V3) bzw. (V4) der “Verlust”, gemessen in einem größeren MSE, gering ist, wenn man statt M^{rLS} das klassische M^{KK} verwendet. Der so entstehende Filter heißt rLS(M^{KK})-Filter. In den folgenden Kapiteln werden wir stets mit dieser rLS-Variante arbeiten.

2.4.2 der Filter von Masreliez/Martin

In Masreliez und Martin (1977) schlagen die Autoren für Verteilungsannahmen (V3) bzw. (V4) im idealen Modell in Theorem 3 folgenden Filter vor:

$$\begin{aligned} \text{und für } t > 0 \quad \beta_{0|0}^{\text{MM}} &= \text{E}[\beta_0] = a_0 \\ \beta_{t|t-1}^{\text{MM}} &= F_t \beta_{t-1|t-1}^{\text{MM}} \\ \beta_{t|t}^{\text{MM}} &= \beta_{t|t-1}^{\text{MM}} + \sum_{k|t-1}^{\text{MM}} Z_k^T T_k^T \Psi(T_k \Delta y_t^{\text{MM}}) \\ \Psi_i(x) &= x_i \min\{1, \frac{b}{|x_i|}\} \end{aligned} \quad (2.12)$$

Dabei gewährleistet T_t , daß

(a) $\mathcal{L}(T_t \Delta y_t^{\text{MM}})$ invariant ist unter der Untergruppe der Gruppe der orthogonalen Matrizen $\mathcal{O}(q)$, in $\mathbb{R}^{q \times q}$, die durch die Matrizen

$$P_i := \text{diag}(1, \dots, 1, -1, 1, \dots, 1)$$

mit -1 an der Stelle i, i erzeugt wird, und

(b) daß sich jede eindimensionale Randverteilung $\mathcal{L}((T_t \Delta y_t^{\text{MM}})_i)$ auffassen läßt als Konvexkontamination zum Radius r von $\mathcal{N}(0, 1)$.

Halten wir t fest, so würde unter der von Masreliez und Martin (1977) gemachten, aber unzutreffenden Annahme, daß $\Delta \beta^{\text{MM}} \sim \mathcal{N}_p(0, \Sigma^{\text{MM}})$, gelten

$$\text{E}_{\text{id}}[\Delta \beta^{\text{MM}} | \Delta y^{\text{MM}}] = \tilde{M}^{\text{MM}} \Delta y^{\text{MM}}$$

mit

$$\begin{aligned} \tilde{M}^{\text{MM}} &= \text{E}[\Delta \beta (\Delta y^{\text{MM}})^\tau] (\text{E}[\Delta y^{\text{MM}} (\Delta y^{\text{MM}})^\tau])^{-1} = \\ &= \Sigma^{\text{MM}} Z^\tau (\Delta^{\text{MM}})^{-1}, \end{aligned} \quad (2.13)$$

denn nach Proposition 1.2.2 ist der bedingte Erwartungswert linear, und der beste lineare Filter ergibt sich wie in der Herleitung der optimalen Matrix M^{KK} in Abschnitt 1.2.3. Für $b = \infty$ ist damit notwendig $T = U(\Delta^{\text{MM}})^{-\frac{1}{2}}$ mit einer gewissen orthogonalen Matrix U . Diese Wahl ist offenbar auch zulässig, d.h. T erfüllt (a) und (b). Wir halten außerdem fest, daß (13) unabhängig von der Normalität von ε gilt.

Betrachten wir im Huberschen Minimax-Ansatz³ für die univariate Lokation die ungünstigsten Verteilung innerhalb der Konvex-Kontaminationsumgebung um $\mathcal{N}(0, 1)$ mit einem Radius r_0 ,

$$\hat{P}(dy) = (1 - r_0) \lambda(dy) \begin{cases} \varphi(y) & |y| \leq k \\ \varphi(k) \exp(-k(|y| - k)) & |y| > k \end{cases},$$

³c.f. Huber (1964)

k so daß

$$2\Phi(k) - 1 + \frac{2\varphi(k)}{k} \stackrel{!}{=} \frac{1}{1-r_0}.$$

Wenn wir in Abhängigkeit von b den Radius r_0 so wählen, daß wir auf $k(r_0) = b$ stoßen, können wir $\tilde{P}(dy)$ durch die kontaminierende Verteilung

$$\tilde{P}(dy) = \frac{(1-r_0(b))}{r_0(b)} [\varphi(b) \exp(-b(|y| - b)) - \varphi(y)] \mathbb{I}_{\{|y| > b\}} \lambda(dy)$$

realisieren. Kontaminieren wir nun $\mathcal{L}((\Delta^{\text{MM}})^{-\frac{1}{2}} \Delta y^{\text{MM}}) = \mathcal{N}(0, \mathbb{I}_q)$ zu diesem Radius r_0 mit $\mathcal{L}(\omega|s)$, wobei⁴ $\omega \sim \text{ufo}(\omega_{q-1})$ und unabhängig dazu $s \sim \tilde{P}$, so erfüllt ein \hat{Y}^0 mit $Y^0 \sim (1-r_0)\mathcal{N}(0, \mathbb{I}_q) + r_0\mathcal{L}(\omega|s)$ Voraussetzungen (a) und (b), und mit $\Delta \hat{y} = (\Delta^{\text{MM}})^{\frac{1}{2}} \hat{Y}^0$ erhalten wir ein kontaminiertes $\Delta \hat{y}$, das im Lokationsfall auf das von Masreliez und Martin angegebene Ψ führen würde. An dieser Stelle können wir aber ebensowenig wie diese Autoren nachweisen, daß

$$\mathcal{L}(\Delta \hat{Y}^{\text{MM}}) = \mathcal{L}(\Delta \beta^{\text{MM}}) * ((1-r_0)\mathcal{N}(0, V) + r_0\mathcal{L}((\Delta^{\text{MM}})^{\frac{1}{2}} \varepsilon^0))$$

für ein irgendwie verteiltes ε^0 . Allerdings kommt man so zu folgendem Filter

$$\begin{aligned} \text{und für } t > 0 \quad \beta_{0|0}^{\text{MM}} &= \mathbb{E}[\beta_0] = a_0 \\ \beta_{t|t-1}^{\text{MM}} &= F_t \beta_{t-1|t-1}^{\text{MM}} \\ \beta_{t|t}^{\text{MM}} &= \beta_{t|t-1}^{\text{MM}} + M_t^{\text{MM}} \Delta y_t^{\text{MM}} w_t^{\text{MM}} \\ w_t^{\text{MM}} &= \min\left\{1, \frac{b}{|(\Delta_t^{\text{MM}})^{-\frac{1}{2}} \Delta y_t^{\text{MM}}|}\right\} \end{aligned} \quad (2.14)$$

Diesen Filter wollen wir *MM-Filter* nennen. Wie bei der Betrachtung des rLS-Filters schlagen wir auch hier vor, anstelle der Matrix \tilde{M}^{MM} die Matrix M^{KK} zu verwenden. Den entsprechenden Filter nennen wir *MM(M^{KK})-Filter*.

2.4.3 optimierter MM-Filter

Aus der Herleitung der optimalen Matrix M^{KK} in Abschnitt 1.2.3 wird unmittelbar ersichtlich, daß man den MM-Filter in (14) noch verbessern kann, indem man anstelle von M^{MM} den “Fourier-Koeffizienten”,

$$M^{\text{MM}} = \mathbb{E}[\Delta \beta (\Delta y^{\text{MM}})^\tau w^{\text{MM}}] (\mathbb{E}[\Delta y^{\text{MM}} (\Delta y^{\text{MM}})^\tau (w^{\text{MM}})^2])^{-1} \quad (2.15)$$

verwendet. Der so entstehende Filter heiße *optimierter MM-Filter* oder kurz *MM(M^{MM})-Filter*.

Bemerkung 2.4.1 Es handelt sich weder beim MM- noch beim optimierten MM-Filter um den ACM-(type-)Filter von Martin (1979); bis auf den eindimensionalen Fall unterscheiden sich die beiden MM-Filter vom ACM-Filter bei der Verwendung der entsprechenden Matrix $M^{[1]}$. Obige Argumentation macht aber deutlich, daß bei letzterem die Behandlung des mehrdimensionalen Falls noch Fragen aufwirft.

⁴ $\omega_{q-1} \triangleq$ die Oberfläche $q-1$ -dimensionalen Einheitssphäre

2.5 Ist M^{rLS} wirklich besser als M^{KK} ?

2.5.1 einige analytische Resultate

Leider können wir im Gegensatz zu Problem (1.24) im allgemeinen nicht hoffen, daß die Beziehungen (1.25) bis (1.26) global gelten, sondern bestenfalls lokal. Genauer muß, wenn F differenzierbar in M ist, notwendig $\frac{\partial}{\partial M} F(M^{\text{rLS}}) = 0$ sein, was dann die mit (1.26) vergleichbare Darstellung liefert.

Wenn wir im folgenden $\Delta\beta$ mit X und Δy mit Y abkürzen, sehen wir leicht, daß der Integrand $f(M, X, Y) := \rho_K(X - H_b(MY))$, den wir stückweise auf den vier Gebieten

$$\begin{aligned} A_1 &:= \{|MY| < b, |X - MY| < K\} \\ A_2 &:= \{|MY| < b, |X - MY| > K\} \\ A_3 &:= \{|MY| > b, |X - b\frac{MY}{|MY|}| < K\} \\ A_4 &:= \{|MY| > b, |X - b\frac{MY}{|MY|}| > K\} \end{aligned} \quad (2.16)$$

definieren, stetig ist. Was die Ableitung anlangt, so gilt folgendes Theorem, das wir hier allerdings nicht beweisen:

Proposition 2.5.1 *Sei F definiert wie in (3). Weiter mögen X, Y Lebesgue-Dichten besitzen. Dann ist F überall in M differenzierbar und*

$$\frac{\partial}{\partial M} F(M) = \sum_{i=1}^4 \mathbb{E}[I_{A_i}(M, X, Y) \frac{\partial}{\partial M} f(M, X, Y)]. \quad (2.17)$$

Verwenden wir die Abkürzung $U_M := MY$ und setzen $\bar{V} := V/|V|$ für einen Vektor V , so gilt mit H definiert als

$$\begin{aligned} H^{(1)}(M, X, Y) &= (U_M - X) \\ H^{(2)}(M, X, Y) &= K \bar{U}_M - \bar{X} \\ H^{(3)}(M, X, Y) &= \frac{b}{|\bar{U}_M|} (\mathbb{I}_p - \bar{U}_M \bar{U}_M^\tau) (b \bar{U}_M - X) \\ H^{(4)}(M, X, Y) &= \frac{Kb}{|\bar{U}_M|} (\mathbb{I}_p - \bar{U}_M \bar{U}_M^\tau) b \bar{U}_M - \bar{X}, \end{aligned} \quad (2.18)$$

daß

$$(I_{A_i} \frac{\partial}{\partial M_{j,k}} f)(M, X, Y) = 2(I_{A_i} H_j^{(i)})(M, X, Y) Y_k \quad (2.19)$$

und das lokale Analogon zu (26) als notwendige Bedingung für ein optimales M^{rLS} :

$$\sum_{i=1}^4 \mathbb{E}[I_{A_i} H^{(i)} Y^\tau] \Big|_{M^{\text{rLS}}} = 0 \quad (2.20)$$

Bemerkung 2.5.2 Beim Nachweis des Theorems geht essentiell die Stetigkeit von f ein, da dann die Abhängigkeit der Gebiete von M keine Auswirkung auf den Wert des Integrals hat. In Dimension $q = 1$, und unter Normalität von ε , gilt für $K = \infty$

$$\frac{d^2}{dM^2} F \neq \sum_{i=1}^4 \mathbb{E}[I_{A_i} Y \frac{d}{dM} H^{(i)}],$$

wie man in folgendem Theorem zeigen kann, das wir hier ohne Beweis präsentieren.

Proposition 2.5.3 *Seien die Voraussetzungen von Proposition 2.5.1 in Kraft; außerdem sei $K = \infty$, $q = 1$, und für $Y = ZX + \varepsilon$, sei $\varepsilon \sim \mathcal{N}(0, V)$. Dann ist F unendlich oft differenzierbar.*

Sei $d_{1,2} := (\pm \frac{b}{|M|} - ZX)/\sqrt{V}$ und möge Φ die Verteilungsfunktion von $\mathcal{N}(0, 1)$ und φ deren Dichte bezeichnen. Wenn wir dann mit Φ_i $i = 1, 2$ und $d\Phi$ die Ausdrücke $\Phi(d_i)$ resp., $\Phi(d_2) - \Phi(d_1)$ abkürzen und analog für φ , so gilt:

$$\begin{aligned} \frac{\partial^2}{\partial M \partial M^\tau} F(\bar{M}) &= 2 \mathbb{E}[\{(ZX)^2 + V\}d\Phi - \frac{b\sqrt{V}}{|M|}(\varphi_1 + \varphi_2) - ZX\sqrt{V}d\varphi] \mathbb{I}_p + \\ &+ \frac{2b}{|M|^2} \mathbb{E}[(1 - \Phi_1 - \Phi_2)\{(X^\tau \bar{M})\mathbb{I}_p + X\bar{M}^\tau + \bar{M}X^\tau - 3(X^\tau \bar{M})\bar{M}\bar{M}^\tau\}] + \\ &+ \frac{2b^2}{|M|^3\sqrt{V}} \mathbb{E}[(X^\tau \bar{M})d\varphi - b(\varphi_1 + \varphi_2)]\bar{M}\bar{M}^\tau \end{aligned} \quad (2.21)$$

Insbesondere ist $\frac{\partial^2}{\partial M \partial M^\tau} F(\bar{M}) \neq \sum_{i=1}^4 \mathbb{E}[I_{A_i} \frac{\partial^2}{\partial M \partial M^\tau} f(\bar{M}; \cdot)]$, denn

$$\begin{aligned} \frac{\partial^2}{\partial M \partial M^\tau} F(\bar{M}) &- \sum_{i=1}^4 \mathbb{E}[I_{A_i} \frac{\partial^2}{\partial M \partial M^\tau} f(\bar{M}; \cdot)] = \\ &= \frac{2b^2}{|M|^3\sqrt{V}} \mathbb{E}[(X^\tau \bar{M})d\varphi - b(\varphi_1 + \varphi_2)]\bar{M}\bar{M}^\tau \neq 0. \end{aligned} \quad (2.22)$$

2.5.2 Bestimmung der optimalen Matrix M^{rLS}

Um festzustellen, inwiefern unter (V3) und (V4) Näherungen (7) und (8) gerechtfertigt sind, beziehungsweise inwiefern dann der $\text{MM}(M^{\text{MM}})$ - besser als der $\text{MM}(M^{\text{KK}})$ -Filter ist, haben wir für verschiedene Modelle die optimalen Matrizen M^{rLS} bzw. M^{MM} numerisch bestimmt.

Weil dazu die Berechnung von Ableitungen $\partial/\partial M(\mathbb{E}\rho_K(\Delta\beta - H_b(M\Delta y)))$ numerisch schwierig ist und für $p, q > 1$ hohe Dimensionen involviert — vgl. auch Abschnitt 2.5.1 —, haben wir zur Bestimmung von M^{rLS} Algorithmus C.2.1 mit Zielfunktion $F(M)$ aus (3) verwendet. Dabei haben wir den Erwartungswert definitionsgemäß aus den Prozeßvergangenheiten $(\Delta\beta_{1:t-1}^{\text{LS}}, \Delta y_{1:t-1}^{\text{LS}})$ heraus per Monte-Carlo-Integration mit Stichprobengröße 7000 ermittelt. Die Größen der M -Gitter im Algorithmus haben wir dabei auf $n_1 = 200$, $n_2 = 10$ und $n_3 = 200$ gesetzt, wobei wir die Gitterpunkte in Phase 1 als $M_i = M^{\text{KK}} + Z^{(i)}$,

$$Z^{(i)} = (Z_{k,l}^{(i)})_{\substack{k=1,\dots,q, \\ l=1,\dots,p}} \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}(0, |M^{\text{KK}}|) \quad (2.23)$$

gewählt haben. Die Varianz in Phase 3 schließlich schrumpft mit j^{-1} , wobei wir als Störungen $Z^{(j)}/\sqrt{j}$, $Z^{(j)}$ wie in (23), verwendet haben.

Im Steady-State Modell aus Beispiel 1.4.1 haben wir dabei die Resultate erhalten, wie sie in Tabelle 2.1 dargestellt sind. Wir haben dabei Pfade bis zu $t = 50$ erzeugt und an dieser Stelle dann den MSE als $F(M) = \text{MSE}(M)$ ausgewertet. Die Zeilen entsprechen jeweils verschiedenen Filtern. In der ersten Spalte ist die verwendete Filter-Vergangenheit abgetragen, in der zweiten das verwendete Clipping- b an der Stelle $t = 50$; letzteres wurde jeweils auf $\delta = 15\%$ geeicht. In den Abbildungen 2.1 und 2.2 ist die Situation jeweils in unterschiedlichen Ausschnitten dargestellt.

Das in Abbildung 2.2 sichtbare Auseinanderfallen von M^{MM} und M^{rLS} ist möglicherweise auch auf numerische Effekte zurückzuführen, denn die Funktion $F(M)$ verläuft bei $\text{rLS}(M^{\text{KK}})$ und $\text{rLS}(M^{\text{rLS}})$ um das Minimum herum sehr flach.

Für die in Tabelle 2.2 aufgeführten, weiteren Zustandsraummodelle haben wir die Untersuchung ebenfalls durchgeführt und sind — koordinatenweise — zu

ganz ähnlichen Verhalten und Graphiken gekommen, so daß wir an dieser Stelle darauf verzichten, diese im einzelnen darzustellen.

Tabelle 2.1: $M^{\text{rLS}}, M^{\text{MM}}$ vs M^{KK} im Steady-State Modell

$\begin{pmatrix} \Delta\beta_{1:t-1} \\ \Delta y_{1:t-1} \end{pmatrix}$	b	$\bar{M} =$ argmin(F)	$F(\bar{M})$ (=: \bar{F})	$F(M^{\text{KK}})$ (=: F^{KK})	rel.Abw. $\bar{F} : F^{\text{KK}}$	rel.Abw. $\bar{M} : M^{\text{KK}}$
KK	∞	0.61515	0.62484	0.62484	0.00000 %	0.00000 %
KK	b_{50}^{rLS}	0.62760	0.71933	0.71943	0.01401 %	2.02464 %
rLS(M^{KK})	b_{50}^{rLS}	0.64458	0.78991	0.79066	0.09443 %	4.78512 %
rLS(M^{rLS})	b_{50}^{rLS}	0.64458	0.78924	0.78999	0.09569 %	4.78512 %
MM(M^{KK})	b_{50}^{MM}	0.80190	0.73567	0.79227	7.69384 %	30.35918 %
MM(M^{MM})	b_{50}^{MM}	0.78481	0.71600	0.76253	6.49800 %	27.58184 %

Tabelle 2.2: weitere untersuchte Modelle

p	q	F	Z	Q	V	a_0
1	1	1	1	9	9	10
2	2	$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0.3 & 1 \\ -0.3 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 0 & 9 \end{pmatrix}$	$\begin{pmatrix} 9 & 0 \\ 0 & 9 \end{pmatrix}$	$\begin{pmatrix} 20 \\ 0 \end{pmatrix}$
2	1	$\begin{pmatrix} 0.5 & 1 \\ -0.3 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$	4	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$
2	2	$\begin{pmatrix} 0.5 & 0.3 \\ 0.6 & 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.46 & 0.25 \\ 0.55 & 0.46 \end{pmatrix}$	$\begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$	$\begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$
3	2	$\begin{pmatrix} 0.5 & 0.3 & 0 \\ 0.6 & 0.5 & 0 \\ 0 & 0 & 0.8 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 3 & 2 & 0 \\ 2 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 2 & -0.2 \\ -0.2 & 0.5 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

2.6 Beispiele

Wir unterstellen in allen Beispielen jeweils (V3).

Beispiel 2.6.1 Als erstes Beispiel haben wir die Daten von (Franěk (2000)) aus dem Datensatz `kalman.dat` zusammen mit den dort angegebenen Systemspezifikationen für die Hyper-Parameter verwendet, also

$$a_0 = 10, \Sigma_0 = 0, Z_t = Z = 1, F_t = F = 1, V_t = V = 9, Q_t = Q = 9. \quad (2.24)$$

Die Beobachtungen $t = 50, 60, 90$ wurden anschließend manipuliert, indem an diesen Stellen artifizielle Ausreißer eingefügt wurden.

Der rLS-Filter ist mit $\delta = 5\%$ kalibriert. Die entsprechende Graphik ist in Abbildung 2.3 dargestellt: Der rLS-Filter entdeckt alle tatsächlichen Ausreißer und stützt die entsprechenden Δy_t^{rLS} . Der Kalman-Filter dagegen reagiert abrupt auf die Ausreißer.

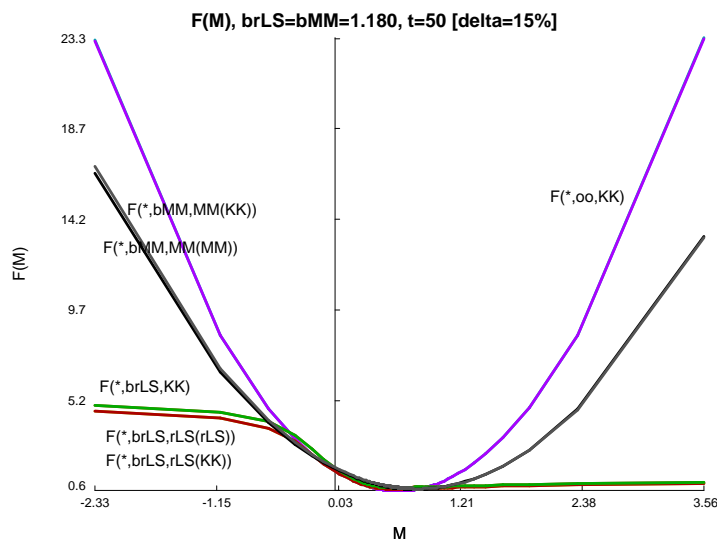


Abbildung 2.1: M^{MM} , M^{rLS} vs M^{KK} im Steady-State Modell:

Darstellung von $F(M, b, \{\text{Filtervergangenheit}\}) = \text{MSE}(\Delta\beta_{t|t}^{[-]}(M, b))$ im gesamten Bereich; das Risiko beim klassischen Kalman-Filter ist unbeschränkt, genauer eine Parabel in M , bei den rLS-Varianten stellt sich eine waagrechte Asymptote ein, und bei den MM-Varianten ergeben sich wieder Parabeln.

Beispiel 2.6.2 Auch das zweite Beispiel übernimmt ein Zustandsraummodell, das bereits von Franěk (2000) als Beispiel verwendet worden ist, nämlich

$$\begin{aligned} a_0 &= \begin{pmatrix} 20 \\ 0 \end{pmatrix}, & \Sigma_0 &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ Z_t = Z &= \begin{pmatrix} 0.3 & 1 \\ -0.3 & 1 \end{pmatrix}, & F_t = F &= \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \\ Q_t = Q &= \begin{pmatrix} 0 & 0 \\ 0 & 9 \end{pmatrix}, & V_t = V &= \begin{pmatrix} 9 & 0 \\ 0 & 9 \end{pmatrix}. \end{aligned} \quad (2.25)$$

Die Verteilung der ε_t unterliegt nun einer 10%-AO-Kontamination. Die kontaminierende Verteilung \tilde{P}^ε ist dabei

$$\tilde{P}^\varepsilon = \mathcal{N}_2\left(\begin{pmatrix} 25 \\ 30 \end{pmatrix}, \begin{pmatrix} 0.9 & 0 \\ 0 & 0.9 \end{pmatrix}\right). \quad (2.26)$$

Gemäß diesem Modell haben wir 50 Zustände und Beobachtungen sowohl im idealen Modell als auch unter Kontamination simuliert. Der rLS-Filter ist auf $\delta = 10\%$ kalibriert. Die graphische Darstellung der Filter findet sich in Abbildung 2.4: In beiden Koordinaten reagiert der Kalman-Filter heftig auf additive Ausreißer, während sich der rLS-Filter auch bei Ausreißern akzeptabel nahe des zu schätzenden Zustandes bleibt.

Beispiel 2.6.3 Um zu zeigen, daß die Robustifizierung nicht “gratis” ist, haben wir den Datensatz `kalman2.dat` unverändert zusammen mit den Hyperparameter-Spezifikationen von Franěk (2000) übernommen. Die Daten stammen daher aus einer idealen Situation, und der klassische Kalman-Filter sollte

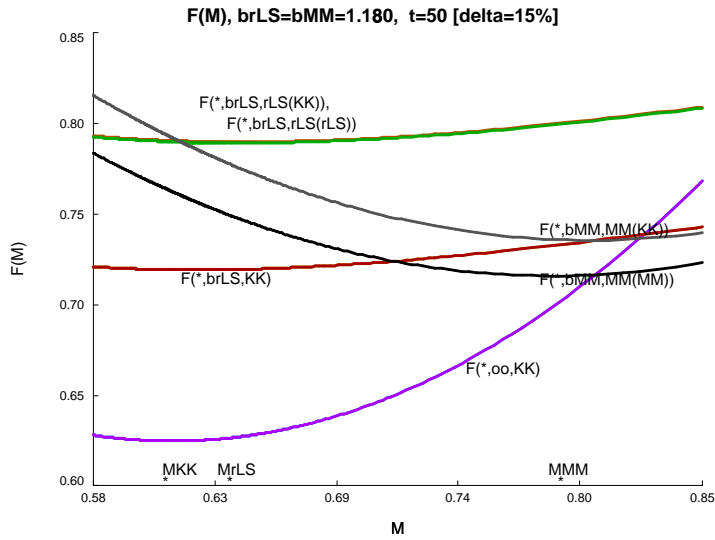


Abbildung 2.2: M^{MM} , M^{rLS} vs M^{KK} im Steady-State Modell:
 Ausschnitt von $F(M, b, \{\text{Filtervergangenheit}\}) = \text{MSE}(\Delta\beta_{t|t}(M, b))$ im
 interessanten Bereich; die Minimalstellen sind markiert;

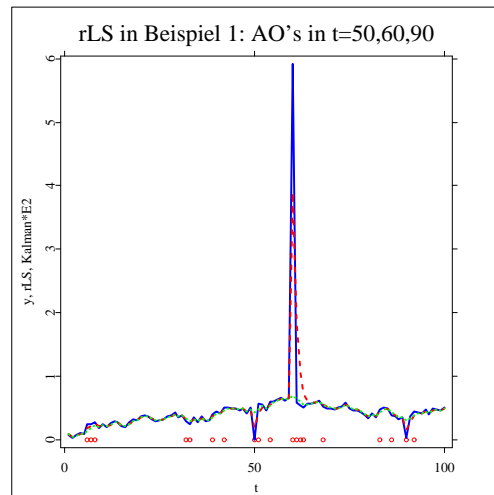


Abbildung 2.3: Die tatsächlichen Zustände sind für den Filter nicht verfügbar, da-
 her stellen wir die tatsächlichen Beobachtungen in einer durchgezogenen, schwarzen Linie dar; der klassische Kalman-Filter ist in einer
 gestrichelten Linie, der rLS-Filter durch eine gepunktete helle Linie
 geplottet; Zeitpunkte, an denen der rLS-Filter die Δy_t^{rLS} stützt,
 sind durch Kringle unterhalb der Graphik markiert.

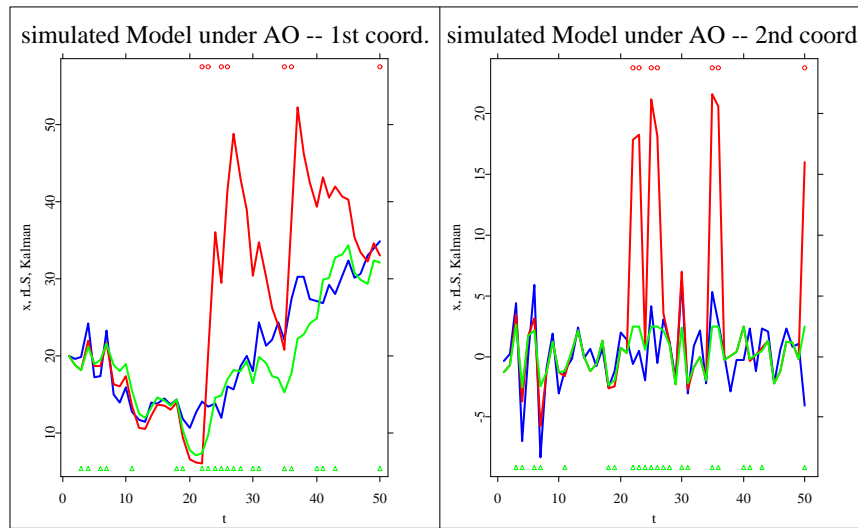


Abbildung 2.4: Die tatsächlichen Zustände sind in einer schwarz durchgezogenen, der Kalman-Filter in einer gestrichelten, dunklen, und der rLS-Filter in einer gepunkteten, hellen Linie geplottet; Zeitpunkte, an denen der rLS-Filter stützt, sind durch Krinkel unterhalb der Graphik, Zeitpunkte mit AO's durch Krinkel unterhalb der Graphik markiert.

in diesem Fall bessere Resultate liefern als der rLS-Filter. Die Hyper-Parameter lauten

$$\begin{aligned}
 a_0 &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}, & \Sigma_0 &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\
 Z_t = Z &= \begin{pmatrix} 1 & 0 \end{pmatrix}, & F_t = F &= \begin{pmatrix} 0.5 & 1 \\ -0.3 & 0 \end{pmatrix} \\
 Q_t = Q &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, & V_t = V &= 4.
 \end{aligned} \tag{2.27}$$

Der rLS-Filter ist auf einen Effizienzverlust von $\delta = 5\%$ geeicht.

Weil wir in diesem Beispiel keine Information über die tatsächlichen Zustände besitzen, untersuchen wir die gefilterten Beobachtungen $y_{t|t}^{[i]} = Z\beta_{t|t}^{[i]}$, die durch Kalman-Filter und rLS-Filter erzeugt werden, mit den tatsächlichen Beobachtungen y_t . Diese gefilterten Beobachtungen sind dann wie üblich in Abbildung 2.5 abgetragen: Offenbar ist der Kalman-Filter nicht so träge wie der rLS-Filter und folgt daher den Beobachtungen enger.

2.7 Bewertung

Mit dem rLS-Filter haben wir ein rekursives Verfahren, das sehr gut mit AO-Ausreißern zurecht kommt. Folgende Eigenschaften sind dabei festzuhalten:

Kalibrierung “off-line”, Auswertung “on-line”

Beim Aufbau des Filters ist zu bemerken, daß in unserem Kontext, also bei be-

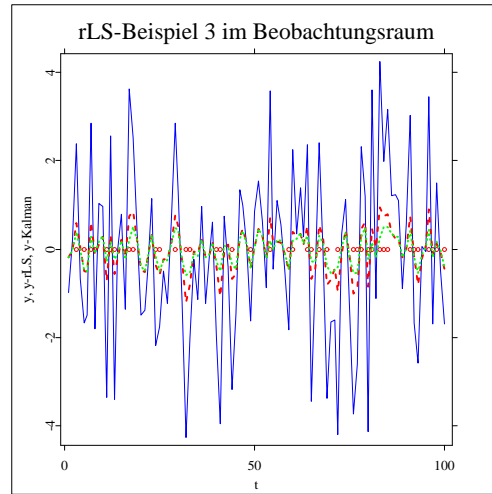


Abbildung 2.5: Die tatsächlichen Beobachtungen sind in einer schwarz durchgezogenen, der Kalman-Filter in einer gestrichelten dunklen, und der rLS-Filter in einer gepunkteten hellen Linie abgetragen; Zeitpunkte, an denen der rLS-Filter clippt, sind durch Kringle markiert.

kannten Systemmatrizen, die rechentchnisch aufwendige Kalibrierung des Filters, d.h. das Finden von b , K und \tilde{M} , vorneweg erledigt werden kann und dann beim eigentlichen Auswerten des Filters an den Beobachtungen y_t die abgespeicherten Werte für b^{rLS} , K , und M^{rLS} verwendet werden können. Das Auswerten geht dann extrem schnell, da nur der Ausdruck $\beta_{t|t-1}^{\text{rLS}} + H_{b^{\text{rLS}}} (M^{\text{rLS}} \Delta y_t^{\text{rLS}})$ zu berechnen ist.

Berechnung von b_t (und M_t^{rLS}) für wenige t

Unserer Erfahrung nach kann man sich im Fall konvergenter $\Sigma_{t|t-1}^{\text{KK}}$ bei der Berechnung von \tilde{M}_t auf die Zeitpunkte t beschränken, bis $\Sigma_{t|t-1}^{\text{KK}}$ sich numerisch nicht mehr ändert.

kaum Unterschied $\text{rLS}(M^{\text{KK}})$: $\text{rLS}(M^{\text{rLS}})$, $K < \infty$: $K = \infty$

Ebenso macht es nach unserer Erfahrung numerisch keinen Unterschied, ob man den $\text{rLS}(M^{\text{KK}})$ - oder den $\text{rLS}(M^{\text{rLS}})$ -Filter verwendet, und ob man mit $K = \infty$ arbeitet oder mit der modifizierten Verlustfunktion. Es bleibt daher nur die Aufgabe, ein weniger aufwendige Bisektionsverfahren zur Ermittlung der Clipping-Schwelle b für jedes t — bis zu dessen Stabilisierung in t — ablaufen zu lassen.

2.8 Verfügbarkeit

In der XploRe-Quantlib `kalman.lib` wurde der rLS-Filter in Form der Quantlets `rLSfil` — der $\text{rLS}(M^{\text{KK}})$ -Filter zu gegebenem b^{rLS} — und `calibrLS` — zur Kalibrierung des rLS-Filters auf ein vorgegebenes δ — umgesetzt; vgl. auch Ruckdeschel (2000b) für die entsprechenden Eingabe-/Ausgabeparameter.

Kapitel 3

der rIC–Filter

3.1 das Filterproblem als Regressionsproblem

Ein weiterer Ansatz, den Kalman-Filter zu robustifizieren, stammt aus dem Regressions-Kontext, wo eine umfangreiche Robustheitstheorie zur Verfügung steht und wir auf den Begriff der “Influenzkurve” oder “Influenzfunktion” zurückgreifen können. Den in diesem Abschnitt hergeleiteten Schätzer wollen wir daher als *rIC* für *recursive Influence Curve* bezeichnen.

3.1.1 der klassische Kalman–Filter als KQ–Regressions–Schätzer

Um die Verbindung zur Regressionstheorie herzustellen, verweisen wir auf ein Lemma, das sich bei Duncan und Horn (1972) und Cipra und Romera (1991) findet:

Lemma 3.1.1 *Der klassische Kalman–Filter (12) läßt sich als ein gewichteter Kleinste–Quadrate–Regressions–Schätzer auffassen.*

Weil wir Teile davon später noch brauchen, geben wir hier einen Beweis dieser Aussage.

BEWEIS: Indem wir die Minimalstelle $\hat{\beta}$ von

$$Q(\beta) = (\beta_{t|t-1}^{\text{KK}} - \beta)^\tau \Sigma_{t|t-1}^{\text{KK}^{-1}} (\beta_{t|t-1}^{\text{KK}} - \beta) + (y_t - Z_t \beta)^\tau V_t^{-1} (y_t - Z_t \beta) \quad (3.1)$$

als Nullstelle der Gleichung

$$\Sigma_{t|t-1}^{\text{KK}^{-1}} (\beta_{t|t-1}^{\text{KK}} - \beta) + Z_t^\tau V_t^{-1} (y_t - Z_t \beta) = 0 \quad (3.2)$$

berechnen, erhalten wir, daß

$$(\Sigma_{t|t-1}^{\text{KK}^{-1}} + Z_t^\tau V_t^{-1} Z_t) \hat{\beta} = Z_t^\tau V_t^{-1} y_t + \Sigma_{t|t-1}^{\text{KK}^{-1}} \beta_{t|t-1}^{\text{KK}}.$$

Nun ist wegen $M_t^{\text{KK}} = \Sigma_{t|t-1}^{\text{KK}} Z_t^\tau (Z_t \Sigma_{t|t-1}^{\text{KK}} Z_t^\tau + V_t)^{-1}$

$$M_t^{\text{KK}} ((Z_t \Sigma_{t|t-1}^{\text{KK}} Z_t^\tau) V_t^{-1} + \mathbb{I}_q) = \Sigma_{t|t-1}^{\text{KK}} Z_t^\tau V_t^{-1}. \quad (3.3)$$

Verwenden wir das Matrixinversionslemma, Anderson und Moore (1979), Problem 3.7, so erhalten wir, daß

$$(\Sigma_{t|t-1}^{\text{KK}^{-1}} + Z_t^T V_t^{-1} Z_t) = (\Sigma_{t|t-1}^{\text{KK}} - M_t^{\text{KK}} Z_t \Sigma_{t|t-1}^{\text{KK}})^{-1},$$

und sehen somit, daß

$$\begin{aligned} \hat{\beta} &= (\Sigma_{t|t-1}^{\text{KK}} - M_t^{\text{KK}} Z_t \Sigma_{t|t-1}^{\text{KK}}) (Z_t^T V_t^{-1} y_t + \Sigma_{t|t-1}^{\text{KK}^{-1}} \beta_{t|t-1}^{\text{KK}}) \\ &= \beta_{t|t-1}^{\text{KK}} + M_t^{\text{KK}} (y_t - Z_t \beta_{t|t-1}^{\text{KK}}) = \beta_{t|t}^{\text{KK}}. \end{aligned}$$

////

3.1.2 das Zustandsraummodell als Regressionsmodell

Auf dieses gewichtete KQ-Problem stößt man im folgenden "Regressionsmodell"

$$\begin{pmatrix} \beta_{t|t-1}^{\text{KK}} \\ y_t \end{pmatrix} = \begin{pmatrix} \mathbb{I}_p \\ Z_t \end{pmatrix} \beta_t + \begin{pmatrix} \beta_{t|t-1}^{\text{KK}} - \beta_t \\ \varepsilon_t \end{pmatrix}. \quad (3.4)$$

Dabei schreiben wir Regressionsmodell in Anführungszeichen, weil der Parameter dieses Modells nicht deterministisch ist, sondern stochastisch. Pagan (1980) nennt in diesem Zusammenhang das Buch von Swamy (1971) und die Artikel von Rosenberg (1973) und Cooley und Prescott (1976) als Referenzen. Modelle dieser Art werden im engeren Sinn nicht durch die klassische asymptotische Theorie¹ erfaßt, und so können wir vorerst einige Schlüsse nur per Analogie ziehen. In Abschnitt 8.8 aber werden wir Argumente liefern, die unsere Vorgehensweise innerhalb des Modells mit stochastischem Parameter absichern. Einstweilen verwenden wir die Theorie für ein deterministisches β_t .

Schreiben wir D_t für die Design-Matrix und fassen wir Beobachtung und Fehler jeweils zusammen zu $z_t = (\beta_{t|t-1}^{\text{KK} \tau}, y_t^T)^T$ und $\zeta_t = ((\beta_{t|t-1}^{\text{KK}} - \beta_t)^T, \varepsilon_t^T)^T$, so wird aus (4) unter (V3) bzw. (V4)

$$z_t = D_t \beta_t + \zeta_t, \quad \zeta_t \sim \mathcal{N}_{p+q}(0, C_t), \quad C_t = \begin{pmatrix} \Sigma_{t|t-1}^{\text{KK}} & 0 \\ 0 & V_t \end{pmatrix}. \quad (3.5)$$

Geht man nun vor wie im klassischen, multivariaten Regressionsmodell

$$Y = X\theta + \varepsilon$$

mit unabhängigen, \mathbb{R}^l -wertigen Beobachtungen y_i und Designpunkten $x_i \in \mathbb{R}^{k \times l}$, so hätte man für den ML-Schätzer die Nullstelle $\hat{\theta}$

$$\sum \Lambda_\theta(y_i) \stackrel{!}{=} 0$$

zu suchen, wobei wegen der Normalität der Fehlerterme $\Lambda_\theta(y_i) = x_i(y_i - x_i^T \theta)$ wäre. Mit offensichtlichen Identifikationen erhalten wir für unseren Fall (mit einer Beobachtung)

$$\Lambda_\theta(z_t) = \Sigma_{t|t-1}^{\text{KK}^{-1}} (\beta_{t|t-1}^{\text{KK}} - \beta_t) + Z_t^T V_t^{-1} (y_t - Z_t \beta_t), \quad (3.6)$$

also das Zusammenfallen mit (2), so daß wir unter (V3) bzw. (V4) den Kalman-Filter als Maximum-Likelihood Schätzer auffassen können.

Wir notieren an dieser Stelle, daß unter (V3) bzw. (V4)

$$\Lambda_\theta(z_t) \sim \mathcal{N}_p(0, \Sigma_{t|t}^{\text{KK}^{-1}}). \quad (3.7)$$

¹vgl. Rieder (1994), 2.4

3.1.3 Influenzkurven in diesem Regressionsmodell

Ausgehend von diesem formalen Regressionsmodell schlagen wir, wie schon Boncelet und Dickinson (1983), Boncelet (1985), und Cipra und Romera (1991), eine Robustifizierung des Kalman-Filters mit Konzepten aus der robusten Regression vor.

Im Unterschied zu diesen Arbeiten setzen wir nicht gleich Regressions-M-Schätzer an, sondern verwenden den Begriff der Influenzkurve im Regressionsmodell aus Rieder (1994), Kap. 7.

Dazu nehmen wir auch den Regressor D_t als stochastisch und unabhängig von ε an und erhalten als Modell

$$\mathcal{P} = \{P_\theta(dx, dy) = P^\varepsilon(dy - x^\tau \theta)P^X(dx) \mid \theta \in \mathbb{R}^p\}, \quad (3.8)$$

wobei wir für P^X endliches zweites Moment und für P^ε endliche Lokations-Fisher-Information im Sinn von Definition B.2.4 verlangen.

Hierfür können wir im Prinzip² auf die in Rieder (1994), Thm. 2.4.2 und Thm. 2.4.6 gezeigte L_2 -Differenzierbarkeit dieses Modells mit L_2 -Ableitung und Fisher-Information

$$\Lambda_\theta(x, y) = x\Lambda_\varepsilon(y - x^\tau \theta), \quad \mathcal{I}_\theta = \mathbb{E}[x\mathcal{I}_{P^\varepsilon}x^\tau] = \text{block-tr } \mathbb{E}[x^\tau x]\mathcal{I}_{P^\varepsilon} \quad (3.9)$$

mit Λ_ε , $\mathcal{I}_{P^\varepsilon}$ der Lokations- L_2 -Ableitung³ bzw. Lokations-Fisher-Information von P^ε zurückgreifen und daher als Regressions-Influenzkurven wie dort in Definition 4.2.12 bzw. 7.4.1 definieren

Definition 3.1.2 Sei \mathcal{P} wie in (8) und Λ_ε die Lokations- L_2 -Ableitung von P^ε . Dann heißt $\psi_\theta \in L_2^p(P_\theta)$ Influenzkurve falls

$$\mathbb{E}_\theta[\psi_\theta | X] = 0 \quad (3.10)$$

$$\mathbb{E}_\theta[\psi_\theta \Lambda_\theta^\tau] = \mathbb{I}_p \quad (3.11)$$

Wie generell bei der linearen Regression haben wir Äquivarianz im folgenden Sinn, — c.f. Rieder (1994), (16), p. 265 —

$$\Lambda_\theta \begin{pmatrix} \beta_{t|t-1}^{\text{KK}} \\ y_t \end{pmatrix} = \Lambda_0 \begin{pmatrix} \beta_{t|t-1}^{\text{KK}} - \theta \\ y_t - Z_t \theta \end{pmatrix}, \quad (3.12)$$

so daß wir im folgenden nur noch den Fall $\theta = 0$ betrachten und den Index für θ weglassen.

Als klassisch optimale Influenzkurve $\hat{\psi}$ ergibt sich $\hat{\psi} = \mathcal{I}^{-1}\Lambda$ mit \mathcal{I} der Fisher-Information des Modells. In unserem formalen Regressionsmodell ergibt sich unter (V3) bzw. (V4) gerade $\mathcal{I} = \Sigma_{t|t}^{\text{KK}}{}^{-1}$, also

$$\hat{\psi} = \Sigma_{t|t}^{\text{KK}} (\Sigma_{t|t-1}^{\text{KK}}{}^{-1} \beta_{t|t-1}^{\text{KK}} + Z_t^\tau V_t^{-1} y_t).$$

²Dort ist nur der Fall univariater Beobachtungen ausgeführt, der Fall multivariater Beobachtungen folgt aber mit dem Begriff der multivariaten Lokations-Fisher-Information aus Definition B.2.6 analog.

³Mit p^ε , der Lebesgue-Dichte von P^ε , ergibt sich $\Lambda_\varepsilon = -\nabla p^\varepsilon / p^\varepsilon$.

3.1.4 der Kalman-Filter als One-Step-Schätzer

Basierend auf dieser Funktion kann man nun zu einer Influenzkurve ψ und einem \sqrt{n} -konsistenten Startschätzer θ_0 einen One-Step-Schätzer $\hat{\theta}$ definieren (Rieder (1994), 6.4):

$$\hat{\theta}^{\text{os}} = \theta_0 + \frac{1}{n} \sum_i \psi_{\theta_0}(x_i, y_i) \quad (3.13)$$

Für diesen One-Step-Schätzer hätte man, sofern beliebig viele Beobachtungen zur Schätzung zur Verfügung stünden, daß für jedes $c > 0$ für alle $h \in \mathbb{R}^p$, $|h| \leq c$ gleichmäßig

$$\sqrt{n}(\hat{\theta} - (\theta + h/\sqrt{n})) \circ P_{\theta+h/\sqrt{n}} \xrightarrow{w} \mathcal{N}(0, E_{\theta}[\psi_{\theta}\psi_{\theta}^{\tau}]).$$

Per formaler Analogie verwenden wir $\beta_{t|t-1}^{\text{KK}}$ als Startschätzer und definieren damit in unserem Kontext zu einer Regressions-Influenzkurve ψ den One-Step-Schätzer

$$\tilde{\beta}_{t|t} = \tilde{\beta}_{t|t-1} + \psi \left(\begin{array}{c} \beta_{t|t-1}^{\text{KK}} - \tilde{\beta}_{t|t-1} \\ y_t - Z_t \tilde{\beta}_{t|t-1} \end{array} \right) \quad (3.14)$$

und dazu wie üblich die Einsrittvorhersage $\tilde{\beta}_{t+1|t} = F_t \tilde{\beta}_{t|t}$.

Lemma 3.1.3 *Unter (V3) bzw (V4) fallen bei Verwendung von $\hat{\psi}$ der One-Step-Schätzer (14) und der Kalman-Filter zusammen.*

BEWEIS: Wir schreiben die Influenzfunktion aus als

$$\begin{aligned} \hat{\psi} &= \Sigma_{t|t}^{\text{KK}} (\Sigma_{t|t-1}^{\text{KK}-1} (\beta_{t|t-1}^{\text{KK}} - \tilde{\beta}_{t|t-1}) + \\ &\quad + Z_t^{\tau} V_t^{-1} (y_t - Z_t \tilde{\beta}_{t|t-1})) + \tilde{\beta}_{t|t-1} \end{aligned}$$

wobei $\tilde{\beta}_{0|0} = \beta_{0|0}^{\text{KK}}$ und $\tilde{\beta}_{1|0} = \beta_{1|0}^{\text{KK}}$. Sei $\tilde{\beta}_{t|t-1} = \beta_{t|t-1}^{\text{KK}}$, dann ist

$$\tilde{\beta}_{t|t} = \beta_{t|t-1}^{\text{KK}} + \Sigma_{t|t}^{\text{KK}} Z_t^{\tau} V_t^{-1} (y_t - Z_t \beta_{t|t-1}^{\text{KK}}) = \beta_{t|t}^{\text{KK}},$$

weil

$$(\Sigma_{t|t-1}^{\text{KK}-1} + Z_t^{\tau} V_t^{-1} Z_t)^{-1} = \Sigma_{t|t-1}^{\text{KK}} Z_t^{\tau} (Z_t \Sigma_{t|t-1}^{\text{KK}} Z_t^{\tau} + V_t)^{-1}.$$

Weiterhin gilt $\tilde{\beta}_{t+1|t} = \beta_{t+1|t}^{\text{KK}}$, und damit ist die Behauptung per Induktion gezeigt. ////

3.2 der rIC-Filter

Die Idee beim rIC-Filter besteht nun darin, die Influenzkurve $\hat{\psi}$ der klassischen Scoresfunktion durch allgemeinere, beschränkte Influenzfunktionen zu ersetzen. Dabei setzen wir von nun an stets (V3) bzw. (V4) voraus.

3.2.1 Definition

Für die lineare Regression konnte man zeigen (c.f. Rieder (1994), chap. 7), daß Influenzkurven von Hampel-Krasker-Gestalt⁴, also

$$\psi_{\theta}^{\text{rIC}} = A(\Lambda_{\theta} - a) \min\left\{1, \frac{b}{|A\Lambda_{\theta} - a|}\right\}, \quad (3.15)$$

⁴vgl. auch Hampel et al. (1986), chap. 6, graphisch dargestellt dort auch in Figure 1, p.322

die Spur der as. Kovarianz unter einer Schranke b an den as. Bias in einer Konvexkontaminationsumgebung minimieren. Dabei ist unter (V3) bzw (V4) wegen der dann geltenden Symmetrie von $\mathcal{L}(\Lambda)$ stets $a = 0$ wählbar.

Wegen der Äquivarianz der Regression (12) ergibt sich

$$\psi_{\theta}^{\text{rIC}}(z_t) = \psi_0^{\text{rIC}} \left(\begin{array}{c} \beta_{t|t-1}^{\text{KK}} - \beta_{t|t-1}^{\text{rIC}} \\ y_t - Z_t \beta_{t|t-1}^{\text{rIC}} \end{array} \right) =: \psi \left(\begin{array}{c} \beta_{t|t-1}^{\text{KK}} - \beta_{t|t-1}^{\text{rIC}} \\ y_t - Z_t \beta_{t|t-1}^{\text{rIC}} \end{array} \right). \quad (3.16)$$

Für die Bestimmung der Matrix A ergibt sich aus (11)

$$A^{-1} \stackrel{!}{=} \text{E}[\Lambda \Lambda^\tau \min\{1, \frac{b}{|\Lambda \Lambda|}\}]. \quad (3.17)$$

Ist zu gegebenem b einmal die Matrix $A = A(b)$ bestimmt, kann man im Korrekturschritt eine robuste Schätzung erzielen, indem man folgenden One-Step-Schätzer, eben den *rIC-Filter*,

$$\beta_{t|t}^{\text{rIC}} = \beta_{t|t-1}^{\text{rIC}} + \psi^{\text{rIC}} \left(\begin{array}{c} \beta_{t|t-1}^{\text{KK}} - \beta_{t|t-1}^{\text{rIC}} \\ y_t - Z_t \beta_{t|t-1}^{\text{rIC}} \end{array} \right) \quad (3.18)$$

und dazu wie üblich die Einschrittvorhersage $\beta_{t+1|t}^{\text{rIC}} = F_t \beta_{t|t}^{\text{rIC}}$ verwendet.

Bemerkung 3.2.1 Boncelet und Dickinson (1983), Boncelet (1985) und Cipra und Romera (1991) schlagen als Robustifizierung des formalen Regressionsmodells (5) jeweils M-Schätzer vor, Boncelet (1985) gibt auch die Möglichkeit eines One-Steps an. Im Unterschied zu unserer Hampel-Krasker-Influenzkurve halten sich Cipra und Romera (1991) an die Vorgehensweise von Huber (1981), Kapitel 7, indem der Regressor nicht mitgestutzt wird. Die Robustifizierung erfolgt koordinatenweise in z_t aus (5).

Boncelet (1985) hingegen verwendet die “dekorrelierten Residuen”

$$C_t^{-\frac{1}{2}}(z_t - D_t \beta_t)$$

als Argument einer ψ -Funktion, so daß auch hier ein “großes” $C_t^{-\frac{1}{2}}$ Hebelpunkte erzeugt.

3.2.2 Kriterium für b

Wie im klassischen Regressionskontext erhalten wir mit (15) eine ganze Familie von Influenzkurven, die durch die Stutzhöhen b parametrisiert wird. Für diese Stutzhöhe b hätte man gerne ein Kriterium, das man unabhängig von den Modelldaten aufstellen kann. Dazu schlagen wir vor, b so zu wählen, daß

$$\text{E}_{\text{id}} |\psi^{\text{rIC}}|^2 \stackrel{!}{=} (1 + \delta) \text{E}_{\text{id}} |\hat{\psi}|^2 = (1 + \delta) \text{tr}(\Sigma_{t|t}^{\text{KK}}), \quad (3.19)$$

wobei E_{id} für die Integration unter (V3) bzw (V4) steht. (19) und (17) resp. das noch zu definierende (24) definieren dann ein Gleichungssystem, das simultan numerisch zu lösen ist.

Bemerkung 3.2.2 (a) Gleichung (19) kalibriert die L_2 -Norm von ψ so, daß sie um $100\delta\%$ größer ist als die L_2 -Norm der klassisch-optimalen Influenzkurve. Wenn uns eine Asymptotik wie im Regressionskontext zur Verfügung

stünde, würde $E_{id} |\psi^{\text{rIC}}|^2$ mit der Spur der asymptotischen Varianz des Schätzers zusammenfallen, so daß $1/(1 + \delta)$ als *asymptotische relative Effizienz (ARE)* aufgefaßt werden könnte.

(b) Die Bestimmung von A und b gemäß (19) und (17) resp. (24) wirft numerische Probleme auf und wird etwas ausführlicher in Anhang D erläutert.

(c) Anders als beim rLS-Filter müssen wir uns hier keine Gedanken machen, daß wir durch die Verwendung des robusten Verfahrens von der idealen Normalität von $\Delta\beta$ abweichen: Dadurch, daß wir immer den klassischen Kalman-Filter mitführen, dürfen wir auch in jedem Schritt t von der Normalität der entsprechenden Scores im idealen Modell ausgehen.

3.2.3 eine Variante dieser Robustifizierung

Wir haben gesehen, daß die Scoresfunktion Λ in unserem Kontext darstellbar ist als $\Lambda = \Lambda_{\text{IO}} + \Lambda_{\text{AO}}$ mit

$$\Lambda_{\text{IO}} = \Sigma_{t|t-1}^{\text{KK}-1} (\beta_{t|t-1}^{\text{KK}} - \beta_t) \quad (3.20)$$

$$\Lambda_{\text{AO}} = Z_t' V_t^{-1} (y_t - Z_t \beta_t), \quad (3.21)$$

und

$$\begin{pmatrix} \Lambda_{\text{IO}} \\ \Lambda_{\text{AO}} \end{pmatrix} \sim \mathcal{N}_{2p}(0, \begin{pmatrix} C_{\text{IO}} & 0 \\ 0 & C_{\text{AO}} \end{pmatrix}), \quad (3.22)$$

$\mathcal{I} = C_{\text{IO}} + C_{\text{AO}} + C_{1,2} + C_{1,2}^T$, wobei $C_{\text{IO}} = \Sigma_{t|t-1}^{-1}$, $C_{\text{AO}} = Z_t' V_t^{-1} Z_t$.

Will man sich nun spezifisch gegen additive Ausreißer schützen⁵, so ist diese Darstellung besser geeignet, denn sie erlaubt es, Innovations- und additive Ausreißer separat zu behandeln, da Innovationsausreißer sich nur im Term Λ_{IO} auswirken und additive entsprechend nur in Λ_{AO} . Als separat gestutzte Influenzkurve ergibt sich

$$\psi^{\text{rICs}} := \tilde{A}(\Lambda_{\text{IO}} \min\{1, \frac{b_{\text{IO}}}{|\tilde{A}\Lambda_{\text{IO}}|}\} + \Lambda_{\text{AO}} \min\{1, \frac{b_{\text{AO}}}{|\tilde{A}\Lambda_{\text{AO}}|}\}). \quad (3.23)$$

Da wir es nur auf AO's abgesehen haben, ist es sinnvoll, b_{IO} grundsätzlich auf ∞ zu setzen, also nur Term Λ_{AO} bei Bedarf zu stutzen. Für die Bestimmung der Matrix \tilde{A} ergibt sich damit aus (11) wegen der Unabhängigkeit der beiden Summanden Λ_i im idealen Modell die Bestimmungsgleichung

$$\tilde{A}^{-1} \stackrel{!}{=} E[\Lambda_{\text{AO}} \Lambda_{\text{AO}}^T \min\{1, \frac{b}{|\tilde{A}\Lambda_{\text{AO}}|}\}] + \Sigma_{t|t-1}^{\text{KK}-1}. \quad (3.24)$$

Zur Kalibrierung von b schlagen wir eine entsprechende Variante von (19) vor.

3.3 Äquivarianz beim rIC und andere Normen

3.3.1 Fall $p = 1$

In Bemerkung 3.2.2 (b) haben wir auf das numerische Problem hingewiesen, die Matrix A in (17) bzw. (24) zu bestimmen. Im Fall $p = 1$ sind jedoch einige

⁵Denkbar wäre auch ein Schutz gegen Innovationsausreißer; dieser stellt sich aber vor allem im Vergleich zum später eingeführten mIC-Glätter und -Filter als nicht befriedigend heraus.

Vereinfachungen möglich. Genauer kann man zeigen, daß man zu einem gewissen δ die Größen A und b zur Bestimmung einer Influenzkurve für beliebige Fisher-Information \mathcal{I} ausgehend von den Größen A_0 und b_0 einer Referenz-Influenzkurve zum selben δ für die Fisher-Information \mathcal{I}_0 gewinnen kann durch

$$A = A_0 \mathcal{I}_0 / \mathcal{I} \quad b = b_0 \sqrt{\mathcal{I}_0 / \mathcal{I}}. \quad (3.25)$$

Der Beweis dieser Tatsache findet sich in Rieder und Ruckdeschel (2000b).

Bemerkung 3.3.1 Weil sich für $p = 1$ alle Normen nur durch einen skalaren Faktor unterscheiden, macht es keinen Sinn, wie in den Abschnitten 3.3.2 und 3.3.3 andere Normen zu verwenden; die hier für ein festes δ gefundenen A und b sind unabhängig von der Wahl der Norm.

3.3.2 informations-standardisierte IC

Anstelle der euklidischen Norm kann man bei der Definition der Hampel-Krasker-Influenzkurve in (15), zur Bestimmung von A in (17) resp. (24), sowie anstelle der Spur $\text{tr } E \psi \psi^\tau = E |\psi|^2$ zur Bestimmung von b in (19) auch andere Normen verwenden. Eine mögliche Wahl ist dabei die Norm $|\cdot|_{\mathcal{I}}$, die für ein $x \in \mathbb{R}^p$ definiert ist als $\sqrt{x^\tau \mathcal{I} x}$. Die mit dieser Norm definierten Influenzkurven besitzen gewisse Äquivarianzeigenschaften (vgl. Rieder (1994), Abschnitt 5.5.4 und Rieder und Ruckdeschel (2000b)). Insbesondere aber gilt in unserem Kontext, d.h. $\Lambda \sim \mathcal{N}_p(0, \mathcal{I})$: Man kann A als $\alpha \mathcal{I}^{-1}$ mit $\alpha \in \mathbb{R}$ wählen und α, b sind unabhängig von \mathcal{I} und hängen nur von der Dimension des Problems ab. Daher ist die Dimension des Problems, die Influenzkurve zu gegebenem δ zu bestimmen, von vornherein zwei — (α, b) —, so daß sich zu deren Bestimmung ein Newtonverfahren anbietet. Genauer hat man für $\tilde{\Lambda} \sim \mathcal{N}(0, \mathbb{I}_p)$ mit den Funktionen

$$\mathcal{F}_u(t) := E[\mathbb{I}_{\{|\tilde{\Lambda}| \leq t\}}], \quad \mathcal{E}_u(t) := E[|\tilde{\Lambda}| \mathbb{I}_{\{|\tilde{\Lambda}| \leq t\}}], \quad \mathcal{V}_u(t) := E[|\tilde{\Lambda}|^2 \mathbb{I}_{\{|\tilde{\Lambda}| \leq t\}}], \quad (3.26)$$

die ausführlicher in Anhang D motiviert werden, die Gleichung $F(b, \alpha) \stackrel{!}{=} 0$ zu lösen:

Proposition 3.3.2 Sei $c := b/\alpha$; dann gilt mit

$$F(b, \alpha) := \begin{pmatrix} F_1(b, \alpha) \\ F_2(b, \alpha) \end{pmatrix} = \begin{pmatrix} \mathcal{V}_u(c) + c(Eu - \mathcal{E}_u(c)) - p/\alpha \\ \mathcal{V}_u(c) + c^2(1 - \mathcal{F}_u(c)) - \delta p/\alpha^2 \end{pmatrix} \quad (3.27)$$

$$dF(b, \alpha) = \begin{pmatrix} \frac{\partial F_1}{\partial b} & \frac{\partial F_1}{\partial \alpha} \\ \frac{\partial F_2}{\partial b} & \frac{\partial F_2}{\partial \alpha} \end{pmatrix} (b, \alpha) = \begin{pmatrix} \frac{Eu - \mathcal{E}_u(c)}{\alpha} & \frac{p - b(Eu - \mathcal{E}_u(c))}{\alpha^2} \\ \frac{2c(1 - \mathcal{F}_u(c))}{\alpha} & \frac{2(\delta p - b^2(1 - \mathcal{F}_u(c)))}{\alpha^3} \end{pmatrix}. \quad (3.28)$$

Der entsprechende Newton-Algorithmus für (α, b) als Nullstelle von F konvergiert lokal superquadratisch.

Der Beweis der hier gemachten Aussagen und der Proposition findet sich in Ruckdeschel (2000a).

3.3.3 selbststandardisierte IC

Ähnliche Eigenschaften wie informations–standardisierte Influenzkurven besitzen auch sogenannte selbststandardisierte Influenzkurven (vgl. Rieder (1994), Abschnitt 5.5.4 und Rieder und Ruckdeschel (2000b)). Anstelle der euklidischen Norm verwendet man dann als Norm $|\cdot|_{\mathcal{C}(\eta)}$, die für ein $x \in \mathbb{R}^p$ definiert ist als $\sqrt{x^\tau \mathcal{C}(\eta)^{-1} x}$ mit $\mathcal{C}(\eta) = E \psi \psi^\tau$. Auch hier gilt in unserem Kontext, daß man A als $\alpha \mathcal{I}^{-1}$ mit $\alpha \in \mathbb{R}$ wählen kann und (α, b) wieder unabhängig von \mathcal{I} ist, so daß sich auch hier zur Bestimmung von (α, b) ein Newtonverfahren anbietet. Es gilt:

Proposition 3.3.3 *Schreiben wir wieder kurz $c := b/\alpha$ und definieren die Funktion $\tilde{r}(c) := r(b, A, Y)$, so gilt mit*

$$G(b, \gamma) := \begin{pmatrix} G_1(b, \gamma) \\ G_2(b, \gamma) \end{pmatrix} = \begin{pmatrix} \mathcal{V}_u(c) + c^2(1 - \mathcal{F}_u(c)) - \delta p / \gamma^2 \\ \gamma^2 - \delta p / \tilde{r}(c) \end{pmatrix} \quad (3.29)$$

$$\begin{aligned} dG(b, \gamma) &= \begin{pmatrix} \frac{\partial G_1}{\partial b} & \frac{\partial G_1}{\partial \gamma} \\ \frac{\partial G_2}{\partial b} & \frac{\partial G_2}{\partial \gamma} \end{pmatrix} (b, \gamma) \\ &= \begin{pmatrix} \frac{2c(1 - \mathcal{F}_u(c))}{\gamma} & \frac{2(\delta p - b^2(1 - \mathcal{F}_u(c)))}{\gamma^3} \\ \frac{\delta p(E u - \mathcal{E}_u(c))}{\gamma \tilde{r}(c)^2} & 2\gamma - \frac{c \delta p(E u - \mathcal{E}_u(c))}{\gamma \tilde{r}(c)^2} \end{pmatrix}, \end{aligned} \quad (3.30)$$

daß ein entsprechender Newton-Algorithmus für (γ, b) als Nullstelle von G lokal superquadratisch konvergiert. α kann dann explizit als $\frac{p}{\tilde{r}(c)}$ gewonnen werden.

Der Beweis dieser Aussagen wie auch der Proposition findet sich wieder in Ruckdeschel (2000a).

3.4 separat gestutzte Varianten des rIC mit anderen Normen

Auch für die in Abschnitt 3.2.3 eingeführte Variante des rIC, bei der man den AO– und den IO–Teil der Scoresfunktion separat stützt, hätte man gerne die Invarianzeigenschaften der standardisierten Varianten des letzten Abschnitts. Dies liegt hier nicht auf der Hand, da man die entsprechenden Normen erst konstruieren muß. Möglich wird dies, wenn man jeweils nur einen Teil der Scoresfunktion robustifizieren möchte. Man erhält:

3.4.1 informations–standardisierte IC

Ohne Einschränkung sei die erste Komponente Λ_{AO} zu robustifizieren. Verwenden wir als standardisierende Norm $|\cdot|_{C_{AO}^{-1}}$ und lösen damit das Fixpunktproblem $B^{-1} \stackrel{!}{=} E[\Lambda_{AO} \Lambda_{AO}^\tau w_{AO}]$ mit

$$w_{AO} = \min\left\{1, \frac{b_{AO}}{|B\Lambda|_{C_{AO}^{-1}}}\right\},$$

so kann man wegen der Normalität von Λ_{AO} im idealen Modell (V3) bzw. (V4) B als $B = \gamma C_{\text{AO}}^{-1}$ mit $0 < \gamma \in \mathbb{R}$ wählen. Aus der Unabhängigkeit von Λ_{AO} und Λ_{IO} im idealen Modell folgt dann mit $A := (\frac{1}{\gamma}C_{\text{AO}} + C_{\text{IO}})^{-1}$

$$\begin{aligned} A^{-1} &= \frac{1}{\gamma}C_{\text{AO}} + C_{\text{IO}} = \text{E}[\Lambda_{\text{AO}}w_{\text{AO}}\Lambda_{\text{AO}}^\tau] + \text{E}[\Lambda_{\text{IO}}\Lambda_{\text{IO}}^\tau] = \\ &= \text{E}[(\Lambda_{\text{AO}}w_{\text{AO}} + \Lambda_{\text{IO}})(\Lambda_{\text{AO}} + \Lambda_{\text{IO}})^\tau] \end{aligned}$$

und somit ist die Korrelationsbedingung erfüllt. Insgesamt ergibt sich als Stutznorm für ψ damit $|\cdot|_G$ mit

$$G = A(BC_{\text{AO}}B)^{-1}A = (\gamma C_{\text{IO}} + C_{\text{AO}})^{-1}C_{\text{AO}}(\gamma C_{\text{IO}} + C_{\text{AO}})^{-1},$$

mit der sich dann ψ schreiben läßt als

$$\psi = A(\Lambda_{\text{AO}} \min\{1, \frac{b}{|A\Lambda_{\text{AO}}|_G}\} + \Lambda_{\text{IO}}).$$

Für die Bestimmung der Stutzhöhe b macht es allerdings dann keinen Sinn, die “Größe” von ψ in der nur durch Λ_{AO} induzierten Norm zu betrachten. Stattdessen verlangen wir

$$\text{E}[|\psi|^2] = \text{E}[|A(\Lambda_{\text{AO}}w_{\text{AO}} + \Lambda_{\text{IO}})|^2] \stackrel{!}{=} (1 + \delta) \text{tr} \mathcal{I}^{-1}.$$

In Proposition 3.3.2 ergeben sich — wie in Ruckdeschel (2000a) nachzulesen — hieraus folgende Modifikationen: Mit $c = b/\gamma$, $A = (C_{\text{IO}} + \frac{1}{\gamma}C_{\text{AO}})^{-1}$ und $\rho := \mathcal{V}_u(c) + c^2(1 - \mathcal{F}_u(c))$, sowie

$$\rho_{\text{AO}} := \gamma \text{tr}[AC_{\text{AO}}A(C_{\text{IO}} + C_{\text{AO}}\rho)A] - b^2(1 - \mathcal{F}_u(c)) \text{tr}(AC_{\text{AO}}A)$$

haben wir

$$F(b, \gamma) := \begin{pmatrix} \mathcal{V}_u(c) + c(\text{E}u - \mathcal{E}_u(c)) - k/\gamma \\ \text{tr}[A(C_{\text{IO}} + C_{\text{AO}}\rho)A] - \delta \text{tr} \mathcal{I}^{-1} \end{pmatrix} \quad (3.31)$$

$$dF(b, \gamma) = \begin{pmatrix} \frac{\text{E}u - \mathcal{E}_u(c)}{\gamma} & \frac{k - b(\text{E}u - \mathcal{E}_u(c))}{\gamma^2} \\ \frac{2c(1 - \mathcal{F}_u(c)) \text{tr}(AC_{\text{AO}}A)}{\gamma} & \frac{2\rho_{\text{AO}}}{\gamma^3} \end{pmatrix}. \quad (3.32)$$

3.4.2 selbststandardisierte IC

Mit den gleichen Techniken wie im vorausgegangenen Abschnitt können wir auch eine selbststandardisierte, separat gestutzte Variante des rIC definieren:

Wir lösen das Fixpunktproblem $B^{-1} \stackrel{!}{=} \text{E}[\Lambda_{\text{AO}}\Lambda_{\text{AO}}^\tau \tilde{w}_{\text{AO}}^2]$, dieses Mal mit

$$\tilde{w}_{\text{AO}} = \min\{1, \frac{b_{\text{AO}}}{|\Lambda|_{B^{-1}}}\}.$$

Wir erhalten wieder, daß B wählbar ist als $\gamma C_{\text{AO}}^{-1}$. Setzen wir dann $A^{-1} := C_{\text{IO}} + \text{E}[\Lambda_{\text{AO}}\Lambda_{\text{AO}}^\tau \tilde{w}_{\text{AO}}]$, so erfüllen wir die Korrelationsbedingung. Wir merken hier noch an, daß auch $\text{E}[\Lambda_{\text{AO}}\Lambda_{\text{AO}}^\tau \tilde{w}_{\text{AO}}] = \alpha C_{\text{AO}}$ für ein $0 < \alpha \in \mathbb{R}$.

Insgesamt ergibt sich als Stutznorm für ψ damit $|\cdot|_G$ mit

$$G = AB^{-1}A = \frac{1}{\gamma}(C_{\text{IO}} + \alpha C_{\text{AO}})^{-1}C_{\text{AO}}(C_{\text{IO}} + \alpha C_{\text{AO}})^{-1},$$

mit der sich dann ψ schreiben läßt als

$$\psi = A(\Lambda_{AO} \min\{1, \frac{b}{|A\Lambda_{AO}|_G}\} + \Lambda_{IO}).$$

Für die Bestimmung von b gehen wir vor wie in Abschnitt 3.4.1; modifiziert man G und dG entsprechend, so kann man analog zum informations-standardisierten Fall wieder ein Newtonverfahren zur Bestimmung von α, b verwenden.

3.5 Beispiele

Der Einfachheit halber stellen wir nur die simultan AO/IO-gestutzte Variante des rIC dar. Um eine bessere Vergleichbarkeit zu den Beispielen des rLS-Filters zu gewährleisten, sind die Beispiele für den rIC-Filter unter den gleichen Bedingungen erstellt, wie die des rLS-Filters.

Beispiel 3.5.1 Wir übernehmen den Setup von Beispiel 2.6.1 inklusive der Kalibrierung auf einen Effizienzverlust von $\delta = 5\%$. Das Resultat ist in Abbildung 3.1 dargestellt: Wie beim rLS-Filter entdeckt der rIC-Filter alle tatsächlichen Ausreißer und stützt die entsprechenden Δy_t^{rIC} . Der Kalman-Filter hingegen reagiert abrupt auf die Ausreißer.

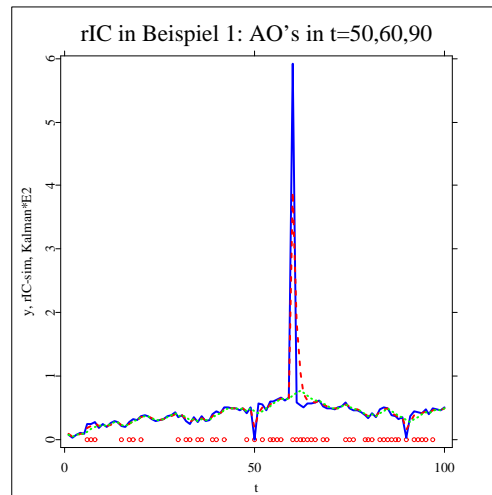


Abbildung 3.1: Die tatsächlichen Zustände sind für den Filter nicht verfügbar, daher stellen wir die tatsächlichen Beobachtungen in einer durchgezogenen, schwarzen Linie dar; der klassische Kalman-Filter ist in einer dunklen, gestrichelten Linie, der rIC-Filter durch eine gepunktete, helle Linie geplottet; Zeitpunkte, an denen der rIC-Filter die Δy_t^{rIC} stützt, sind durch Kringle unterhalb der Graphik markiert.

Beispiel 3.5.2 Setup und Kalibrierungs- δ sind genau wie in Beispiel 2.6.2. Das Resultat ist in Abbildung 3.2 dargestellt: Im Gegensatz zum rLS-Filter erscheinen dem rIC-Filter Ausreißer mit Mittelwert 25 nicht als Ausreißer-verdächtig; das Clipping in der ersten Koordinate hat kaum einen Effekt. In

der zweiten Koordinate reagiert der Kalman-Filter dagegen heftig auf additive Ausreißer, während der rIC-Filter auch bei den Ausreißern nahe am Zustand bleibt.

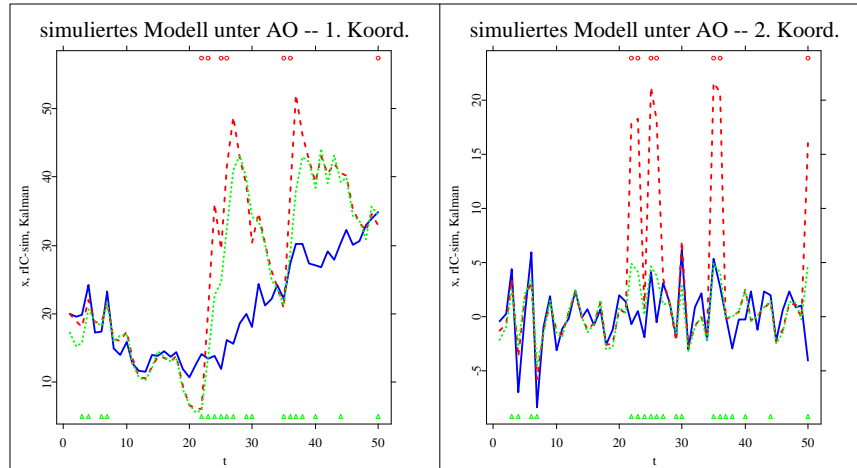


Abbildung 3.2: Die tatsächlichen Zustände sind in einer schwarz durchgezogenen, der Kalman-Filter in einer gestrichelten, dunklen, und der rIC-Filter in einer gepunkteten, hellen Linie geplottet; Zeitpunkte, an denen der rIC-Filter stutzt, sind durch Krinkel unterhalb der Graphik, die Zeitpunkte mit AO's durch Krinkel oberhalb der Graphik markiert.

Beispiel 3.5.3 Wie in Beispiel 2.6.3 wollen wir anhand dieses Beispiels zeigen, daß wir bei Verwendung der rIC-Filters im idealen Modell an Effizienz gegenüber dem klassischen Kalman-Filter verlieren. Setup und Kalibrierungs- δ sind gegenüber Beispiel 2.6.3 unverändert. Das Resultat ist in Abbildung 3.3 dargestellt: Wie beim rLS-Filter folgt der Kalman-Filter den Beobachtungen enger als der robuste Filter.

3.6 Bewertung

Wie der rLS-Filter ist der rIC-Filter ein (strikt) rekursives Verfahren, das im allgemeinen sehr gut mit AO-Ausreißern zurecht kommt.

Kalibrierung “off-line”, Auswertung “on-line”

Wie beim rLS-Filter kann auch hier bei bekannten Systemmatrizen die rechen-technisch aufwendige Bestimmung der Influenzkurven ψ_t^{rIC} , das heißt das Finden von b und A noch vor der Datenerhebung “off-line” geschehen. Die eigentliche Auswertung des Filters an den Beobachtungen y_t ist dann ähnlich unkompliziert wie beim rLS-Filter.

Berechnung von b_t und A_t für wenige t

Im zeitinvarianten Fall gilt für konvergentes $\Sigma_{t|t[-1]}^{\text{KK}}$ automatisch die Mitkonvergenz der Lagrange-Multiplikatoren A_t und der Clipping-Schwellen b_t^{rIC} . Weil

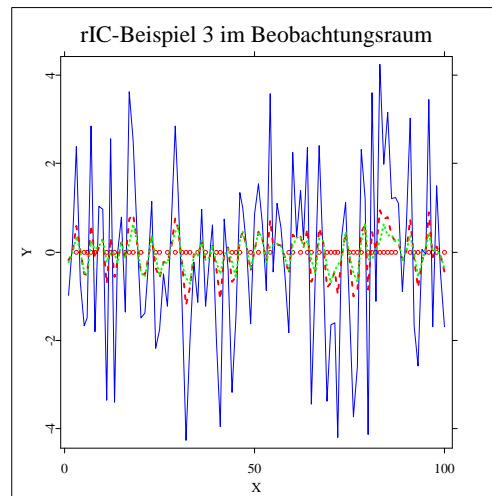


Abbildung 3.3: Die tatsächlichen Beobachtungen sind in einer schwarz durchgezogenen, der Kalman-Filter in einer gestrichelten dunklen, und der rLS-Filter in einer gepunkteten hellen Linie abgetragen; Zeitpunkte, an denen der rIC-Filter clippt, sind durch Kringel markiert.

diese Konvergenz geometrisch ist⁶, stabilisiert sich $\Sigma_{t|t[-1]}^{\text{KK}}$ sehr schnell — im Fall des Steady-State Modells aus Beispiel 1.4.1 ist etwa

$$|\Sigma_{t|t-1}^{\text{KK}} - \lim_t \Sigma_{t|t-1}^{\text{KK}}| < 10^{-5} \quad \text{für } t > 6.$$

Verwendet man zusätzlich eine der standardisierenden Normen aus Abschnitt 3.3, so ist auch die Bestimmung der Matrizen A im Prinzip “online” möglich.

Vergleich rIC : rLS

Die Frage, ob der rLS-Filter dem rIC-Filter vorzuziehen ist, hängt wiederum von der Kontamination ab und läßt sich nicht allgemein beantworten:

Während im Beispiel 2 von P. Franěk der rLS-Filter eindeutig die besseren Ergebnisse liefert, erzielt der rIC-Filter in der Mini-Studie in Kapitel 5 in den betrachteten Kontaminationen bei Eichung auf denselben Effizienzverlust bessere MSE-Werte.

3.7 Verfügbarkeit

Wie der rLS-Filter, so steht auch der rIC-Filter im Rahmen des Programmpaketes `XploRe` in der `XploRe-Quantlib` `kalman.lib` zur Verfügung. Der rIC-Filter wurde dabei in Form der Quantlets `ricfil` und `calibrIC` realisiert, `ricfil` für den rIC- bzw. rICs-Filter — je nach Wahl der Eingabeparameter — zu gegebenem $(A[\tilde{A}], b)_t^{\text{rIC}*}$, und `calibrIC` zur Kalibrierung des rIC[s]-Filters auf ein vorgegebenes δ ; vgl. auch Ruckdeschel (2000b).

⁶Weil $\Sigma_{t|t[-1]}$ konvergiert, ist (F, Z) ermittelbar im Sinn von Definition 7.2.13; daher gibt es ein K , so daß $\rho(F - KZ) < 1$, und dieses $\rho(F - KZ)$ ist — vgl. Anderson und Moore (1979), p. 81 zugleich im wesentlichen die Konvergenzrate.

Kapitel 4

m -Schritt-Robustifizierung: der mIC-Glätter und -Filter

4.1 Problemstellung

4.1.1 Innovationsausreißer

Wenn wir Innovationsausreißer für möglich erachten und auch dagegen “robust” sein wollen, ist es, wie bereits erwähnt, nicht sinnvoll, auf der rekursiven Struktur des Kalman-Filters zu bestehen, weil wir nur dann einen Strukturbruch im Prozeß erkennen können, wenn eine gewisse Vorgeschichte des Prozesses bekannt ist. Nur so ist es möglich, dem tatsächlichen Zustand bei Niveausprüngen schneller zu folgen als der klassische Kalman-Filter.

4.1.2 ein großes Regressionsproblem

Wir betrachten nun die Aufgabe, simultan Schätzungen für β_T (Filterproblem) und $\beta_{T-m:T-1}$ (Glättungsproblem), basierend auf $Y_{T-m:T}$ und $\beta_{T-m|T-m-1}^{\text{KK}}$ zu finden.

Wenn wir hierfür das Regressionsmodell mit unbekanntem “Parameter” $\beta_{T-m:T}$ zugrundelegen, wie es von Duncan und Horn (1972) vorgeschlagen wird, und zusätzlich für die ideale Situation (V3) bzw. (V4) annehmen, können wir schreiben

$$\begin{aligned} Y_t &= Z_t \beta_t + \epsilon_t & t &= T-m, \dots, T \\ 0 &= \beta_t - F \beta_{t-1} - v_t & t &= T-m+1, \dots, T \\ \beta_{T-m|T-m-1}^{\text{KK}} &= \beta_{T-m} + \zeta_{T-m} \end{aligned} \quad (4.1)$$

so daß im idealen Modell (V3) bzw. (V4) $\zeta_{T-m} \sim \mathcal{N}_p(0, \Sigma_{T-m|T-m-1}^{\text{KK}})$ und stochastisch unabhängig von allen v_t, ϵ_t ist, die in (1) auftauchen.

Bemerkung 4.1.1 Wie im rIC-Fall schreiben wir “Parameter” in Anführungszeichen, weil im Gegensatz zur klassischen Regressionstheorie der Parameter selbst stochastisch ist. Im Unterschied zum rIC-Filter und ebenfalls anders als in der klassischen Regressionstheorie sind sogar einige “Beobachtungskordinaten” deterministisch.

In Matrixschreibweise schreibt sich (1) als

$$\begin{pmatrix} \beta_{T|T-m-1}^{\text{KK}} \\ y_{T-m} \\ 0 \\ y_{T-m+1} \\ \vdots \\ 0 \\ y_T \end{pmatrix} = \begin{bmatrix} \mathbb{I}_p & 0 & & & & & \\ Z_{T-m} & 0 & & & & & \\ -F_{T-m} & \mathbb{I}_p & & 0 & & & \\ 0 & Z_{T-m+1} & & & & & \\ & \ddots & & & & & \\ & & & & \ddots & & \\ 0 & & & -F_{T-1} & \mathbb{I}_p & & \\ & & & 0 & Z_T & & \end{bmatrix} \begin{pmatrix} \beta_{T-m} \\ \vdots \\ \beta_T \end{pmatrix} + \begin{pmatrix} \zeta_{T-m} \\ \varepsilon_{T-m} \\ -v_{T+1-m} \\ \varepsilon_{T+1-m} \\ \vdots \\ -v_T \\ \varepsilon_T \end{pmatrix} \quad (4.2)$$

4.2 klassische Lösung

4.2.1 die Kleinste-Quadrate-Lösung

Man erhält die Kleinste-Quadrate-Lösung für $\beta^{\text{LS}} = \beta_{T-m:T}^{\text{LS}}$, indem man

$$\begin{aligned} 2G(\beta) &:= \sum_{t=T-m}^T (y_t - Z_t \beta_t)^\tau V_t^{-1} (y_t - Z_t \beta_t) + \\ &\quad + \sum_{t=T-m+1}^T (\beta_t - F_t \beta_{t-1})^\tau Q_t^{-1} (\beta_t - F_t \beta_{t-1}) + \\ &\quad + (\beta_{T-m} - \beta_{T-m|T-m-1}^{\text{KK}})^\tau \Sigma_{T-m|T-m-1}^{\text{KK}-1} (\beta_{T-m} - \beta_{T-m|T-m-1}^{\text{KK}}) \\ &= \min_{\beta_{T-m}, \dots, \beta_T}! \end{aligned} \quad (4.3)$$

löst.

Bemerkung 4.2.1 Im idealen Modell (V3) bzw. (V4) fällt dieses Minimum sowohl mit dem bedingten Erwartungswert als auch mit dem Posterior-Modus zusammen (vgl. Proposition 4.2.3), sowie wegen der Linearität des bedingten Erwartungswertes zugleich mit $\beta_{T-m:T|T}^{\text{KK}}$.

Diese Bemerkung führt uns zu einer Robustifizierung analog zum rIC-Kontext:

4.2.2 die $(m+1)p$ -dimensionale Scores-Funktion

Gehen wir nun wie bei der robusten Regression (vgl. Rieder (1994), chap. 7) vor, so ermitteln wir unter (V3) und (V4) die L_2 -Ableitung Λ , indem wir G nach β ableiten und erhalten eine Scoresfunktion, die in $m+1$ p -dimensionale Blöcke $\Lambda_{(t)}$, $t = T-m, \dots, T$, zerfällt:

$$\begin{aligned} \frac{\partial}{\partial \beta_t} G &=: \Lambda_{(t)}(y, \beta) = \Lambda_{(t)}^{\text{AO}}(y, \beta) + \Lambda_{(t)}^{\text{IO}}(y, \beta) \\ &\quad \text{für } t = T-m, \dots, T \\ \text{mit} \quad \Lambda_{(t)}^{\text{AO}} &= Z_t^\tau V_t^{-1} (y_t - Z_t \beta_t) \\ &\quad \text{für } t = T-m, \dots, T \\ \Lambda_{(t)}^{\text{IO}} &= -Q_t^{-1} (\beta_t - F_t \beta_{t-1}) + F_{t+1}^\tau Q_{t+1}^{-1} (\beta_{t+1} - F_{t+1} \beta_t) \\ &\quad \text{für } t = T-m+1, \dots, T-1 \\ &= -Q_T^{-1} (\beta_T - F_T \beta_{T-1}) \\ &\quad \text{für } t = T \\ &= -\Sigma_{T-m|T-m-1}^{\text{KK}-1} (\beta_{T-m} - \beta_{T-m|T-m-1}) + \\ &\quad + F_{T-m+1}^\tau Q_{T-m+1}^{-1} (\beta_{T-m+1} - F_{T-m+1} \beta_{T-m}) \\ &\quad \text{für } t = T-m. \end{aligned} \quad (4.4)$$

Ihre Fisher-Information $\mathcal{I} := E[\Lambda\Lambda^\tau]$ besitzt Block-Tridiagonal-Gestalt,

$$\begin{aligned} \mathcal{I} := & \text{diag}[(Z_t^\tau V_t^{-1} Z_t) + \Sigma_{T-m|T-m-1}^{\text{KK}^{-1}} + F_{T-m+1}^\tau Q_{T-m+1}^{-1} F_{T-m+1}, \\ & (Q_t^{-1} + F_{t+1}^\tau Q_{t+1}^{-1} F_{t+1})_{T-m < t < T}, Q_T^{-1}] + \\ & - \text{ndiag}_1[(F_t Q_t^{-1})_{T-m < t \leq T}] - \\ & - \text{ndiag}_1[(F_t Q_t^{-1})_{T-m < t \leq T}]^\tau, \end{aligned} \quad (4.5)$$

wobei “diag[.]” für die $(m+1) \times (m+1)$ Diagonalmatrix mit entsprechenden $p \times p$ Matrizen als Elementen steht. Analog bezeichnet “ndiag₁[.]” die $(m+1) \times (m+1)$ Matrix, bei der die erste obere Nebendiagonale mit den $p \times p$ Matrizen aus dem Argument besetzt ist.

4.2.3 Influenzkurven in diesem Regressionsmodell

Wie schon im rIC-Fall nehmen wir den Regressor in (2) als stochastisch und unabhängig von den Fehlern an und definieren

Definition 4.2.2 Sei die Scoresfunktion Λ wie in (4) definiert. Dann heißt $\psi_\theta \in L_2^{(m+1)p}(P_\theta)$ Influenzkurve, falls

$$E_\theta[\psi_\theta | \text{Regressor}] = 0 \quad (4.6)$$

$$E_\theta[\psi_\theta \Lambda_\theta^\tau] = \mathbb{I}_{(m+1)p}. \quad (4.7)$$

Wie im rIC-Fall dürfen wir eine analog zu (12) definierte Äquivarianz verwenden, so daß wir im folgenden wieder nur noch den Fall $\theta = 0$ betrachten und den Index für θ weglassen. Als klassisch optimale Influenzkurve $\hat{\psi}$ ergibt sich $\hat{\psi} = \mathcal{I}^{-1} \Lambda$.

4.2.4 verschiedene Konstruktionen im idealen Modell

Proposition 4.2.3 Im idealen Modell (V3) bzw. (V4) fallen One-Step mit beliebigem Startschätzer β_0 , bedingter Erwartungswert, orthogonale Projektion $oP(\beta_{T-m:T} | Y_{1:T})$ und Posterior-Modus zusammen.

BEWEIS: Daß im Fall normalverteilter Übergangsdichten bedingter Erwartungswert und Posterior-Modus zusammenfallen ist eine bekannte Tatsache (vgl. Künstler (1995), Fahrmeir und Künstler (1999)), und weil im Gaußschen Fall der bedingte Erwartungswert linear ist, fällt er auch mit der orthogonalen Projektion auf \mathcal{L}_t zusammen.

Schreibt man nun in den Bezeichnungen von Künstler (1995), p. 66 im Fall fester Fensterbreite¹

$$\begin{aligned} \bar{y} & := (\beta_{T-m|T-m-1}, y_{T-m}, \dots, y_T) \\ \mathbf{Z} & := \text{diag}[\mathbb{I}_p, Z_{T-m}, \dots, Z_T], \\ \mathbf{W} & := \text{diag}[\Sigma_{T-m|T-m-1}^{-1}, V_{T-m}^{-1}, \dots, V_T^{-1}], \end{aligned}$$

respektive in variabler

$$\begin{aligned} \bar{y} & := (a_0, y_1, \dots, y_T) \\ \mathbf{Z} & := \text{diag}[\mathbb{I}_p, Z_{T-m+1}, \dots, Z_T] \\ \mathbf{W} & := \text{diag}[Q_0^{-1}, V_{T-m+1}^{-1}, \dots, V_T^{-1}], \end{aligned}$$

¹vgl. auch Lemma 3.1.3

so schreibt sich die Scoresfunktion als

$$\Lambda(\bar{y}, \beta) = \mathbf{Z}^\tau \mathbf{W} \bar{y} - \mathcal{I}\beta.$$

Wie eben bemerkt, fallen im idealen Modell (V3) bzw. (V4) Kleinste-Quadrate-Lösung und Posterior-Modus zusammen. Daher löst der klassische Kalman Glätter und Filter $\beta_{0|T-m}^{\text{KK}}:T|T$ auch das Nullstellenproblem $\Lambda_\beta(\bar{y}) \stackrel{!}{=} 0$ in β . Wegen der Linearität von Λ in β und der Invertierbarkeit von \mathcal{I} ist dies äquivalent mit $\beta_{0|T-m}^{\text{KK}}:T|T = \mathcal{I}^{-1} \mathbf{Z}^\tau \mathbf{W} \bar{y}$. Damit aber haben wir für beliebige Startschätzer β^0

$$\begin{aligned} \beta^{\text{os}} &:= \beta^0 + \psi_{\beta^0}(\bar{y}) = \beta^0 + \mathcal{I}^{-1} \Lambda(\beta^0, \bar{y}) = \\ &= \beta^0 + \mathcal{I}^{-1} (\mathbf{Z}^\tau \mathbf{W} \bar{y} - \mathcal{I}\beta^0) = \mathcal{I}^{-1} \mathbf{Z}^\tau \mathbf{W} \bar{y} = \beta_{0|T-m}^{\text{KK}}:T|T \end{aligned}$$

////

4.3 Robustifizierung

4.3.1 Definition von ψ^{mIC}

Weil im idealen Model (V3) bzw. (V4) dieses hoch-dimensionale Λ normalverteilt ist, können wir im Prinzip die Methoden aus Abschnitt D verwenden, um wie in Abschnitt 3.2 eine robustifizierte Version ψ^{mIC} von $\mathcal{I}^{-1}\Lambda$ zu erstellen, wengleich die Bestimmung des Lagrange-Multiplikators A möglicherweise numerisch an der hohen Dimension scheitert. Davon abgesehen können wir aber für ein geeignet gewähltes $b > 0$ definieren:

$$\psi^{\text{mIC}} = \psi_\beta^{\text{mIC}}(Y_T \dots Y_{T-m}, \beta_{T-m|T-m-1}) = A\Lambda_\beta(\cdot) \min\left\{1, \frac{b}{|A\Lambda_\beta|}\right\}, \quad (4.8)$$

wobei $A \in \mathbb{R}^{(m+1)p \times (m+1)p}$ so zu wählen ist, daß

$$\text{E}_{\text{id}}[\psi^{\text{mIC}} \Lambda^\tau] = \mathbb{I}_{(m+1)p}. \quad (4.9)$$

4.3.2 feste und variable Fensterweite

Aufbauend auf diesem ψ^{mIC} wollen wir analog zum rIC-Fall ein One-Step-Verfahren zur Schätzung von β benutzen, wobei wir das Superskript mIC bei ψ^{mIC} weglassen. Dazu können wir zwei unterschiedliche Vorgehensweisen verwenden:

- $m = T - 1$ — variable Fensterweite:

Wir nutzen unabhängig von der Situation alle zur Verfügung stehende Information, was eine wachsende Dimension von ψ nach sich zieht, so daß streng gesprochen, eine Folge von ψ 's $\psi = \psi^{(T)}$ vorliegt und wir das folgende, aufsteigende Dreiecksschema für T definieren:

$$\begin{aligned} \beta_{0|0}^{\text{mIC}} &= \text{E}[\beta_0] = a_0 \\ \beta_{T|T-1}^{\text{mIC}} &= F_T \beta_{T-1|T-1}^{\text{mIC}} \\ \beta_{0:T|T}^{\text{mIC}} &= \beta_{0:T|T-1}^{\text{mIC}} + \psi_{\beta_{0:T|T-1}^{\text{mIC}}}(y_{1:T}) \end{aligned} \quad (4.10)$$

- $m < T - 1$, m fixiert — feste Fensterweite;
Wir fangen wie im ersten Fall an und wenden (10) für $T \leq m$ an. Von $T = m$ an allerdings verwenden wir die feste Funktion $\psi = \psi^{(m)}$ und setzen

$$\begin{aligned}\beta_{T|T-1}^{\text{niC}} &= F_T \beta_{T-1|T-1}^{\text{niC}} \\ \beta_{T-m:T|T}^{\text{niC}} &= \beta_{T-m:T|T-1}^{\text{niC}} + \psi_{\beta_{T-m:T|T-1}^{\text{niC}}} (y_{T-m:T}, \beta_{T-m|T-m-1}^{\text{KK}})\end{aligned}\quad (4.11)$$

Dies entspricht der Vorstellung, daß die Information vor $t = T - m$ unverfälscht vorliegt, das heißt, daß man diese Information, wie im idealen Modell korrekt, im bedingten Erwartungswert zusammenfassen kann, der dort wiederum gerade der klassische Kalman-Filter ist.

Bemerkung 4.3.1 Weil im Unterschied zum Fall mit fixierter Fensterweite für den frühesten Zeitpunkt $t = T - m = 0$ keine Beobachtung y_0 vorliegt, wenn man mit variabler Fensterweite arbeitet, fällt in der Scoresfunktion der Term $\Lambda_{(0)}^{\text{AO}}$ ersatzlos weg und führt auch zu keinem Diagonaleintrag $Z_0^T V_0^{-1} Z_0$ in \mathcal{I} . Hingegen lautet

$$\Lambda_{(0)}^{\text{IO}} = F_1^T Q_1^{-1} (\beta_1 - F_1 \beta_0) - Q_0^{-1} (\beta_0 - a_0),$$

und man muß den Term $\Sigma_{T-m|T-m-1}^{-1}$ in \mathcal{I} ersetzen durch Q_0^{-1} .

4.3.3 Posterior-Modus für andere Eichverteilungen als die Normalverteilung

Ein anderer Ansatz, Robustheit im Glättungs- und Filterproblem

$$E |\beta_{t-m:t} - f_{t-m:t}(y_{1:t})|^2 = \min_{f_{t-m:t}} ! \quad (4.12)$$

zu erzielen, wird in der Dissertation von Künstler (1995) verfolgt:

Man interpretiert die zu minimierende Quadratsumme im Kleinste-Quadrate-Problem als Realisationen der klassischen Scoresfunktion, d.h. der negativen, logarithmierten Dichte, so daß das Minimum in (3) zugleich die Posterior-Dichte maximiert. Ersetzt man nun die Gaußverteilung als ideale Verteilung in (V3) bzw. (V4) durch eine Verteilung mit fetteren Schwänzen, wie zum Beispiel der t -Verteilung mit geringer Zahl an Freiheitsgraden, was einer Änderung der Eichverteilung für den posterior Modus entspricht, so erhält man für $m = T$ folgendes Optimierungsproblem:

$$\sum_{t=0}^T \rho_{\text{AO}}(y_t - Z_t \beta_t) + \sum_{t=0}^T \rho_{\text{IO}}(\beta_t - F_t \beta_{t-1}) = \min_{\beta_0, \dots, \beta_T} ! \quad (4.13)$$

Im Endeffekt bedeutet dies den Ersatz der quadratischen Verlustfunktion durch eine andere Verlustfunktion, die langsamer im Betrag der Argumente wächst, ja im Fall der t -Verteilung sogar zu einer Ableitung führt, die gegen 0 geht, falls das Argument im Betrag unbeschränkt wächst, was einem “redescender” als Influenzkurve entspräche.

Zusätzlich werden bei dem in Künstler (1995) vorgeschlagenen Verfahren bei

Verwendung der t -Verteilung als Eichverteilung simultan mit einem EM-Algorithmus Hyper-Parameter — Q_t und V_t —, sowie die Zahl der Freiheitsgrade geschätzt. Ein direkter Vergleich zu dem hier vorgestellten Verfahren ist also nur bedingt möglich, solange keine Mitschätzung von Hyperparametern erfolgt.

Bemerkung 4.3.2 Bei der Modifikation in Problem (13) gegenüber dem klassischen Ansatz hat man die Veränderung in der Stammfunktion der Scoresfunktion durchgeführt, im Gegensatz zu dem hier verfolgten Ansatz, der die Scoresfunktion selbst durch Stutzen modifiziert. Daraus entstehen Unterschiede bei der Behandlung von IO-Ausreißern: Durch die Summation in (13) wird so gruppiert, daß die “ersten” Differenzen $\beta_t - F_t\beta_{t-1}$ jeweils durch einen ρ'_{IO} -Term modifiziert werden. Hingegen werden bei der weiter unten in Abschnitt 4.4 eingeführten blockweisen, separaten Stutzung die “zweiten” Differenzen

$$\Lambda_t^{\text{IO}} := -Q_t^{-1}(\beta_t - F_t\beta_{t-1}) + F_{t+1}^T Q_{t+1}^{-1}(\beta_{t+1} - F_{t+1}\beta_t)$$

jeweils durch einen Gewichtsterm $w(x) = \min\{1, b/x\}$ robustifiziert zu $\tilde{\psi}_t^{\text{IO}} := A\tilde{\Lambda}_t^{\text{IO}} w(|A\tilde{\Lambda}_t^{\text{IO}}|)$.

Interpretiert man diese “zweiten” Differenzen Λ_t^{IO} als Glattheitsstrafsterme, so erreicht man durch ihre direkte Behandlung, daß die Datentreue-Strafterme $\Lambda_t^{\text{AO}} := Z_t^T V_t^{-1}(y_t - Z_t\beta_t)$ stärker bei der Optimierung in (13) ins Gewicht fallen, so daß der entsprechende Schätzvektor $\beta_{0:T}^{\text{mIC}}$ den Beobachtungen schneller folgt als der klassische Schätzer. Dies entspricht genau der Umgewichtung des Optimierungsproblems, von der Künstler (1995) p.65 ab Formel (3.7) spricht.

4.3.4 separates Stutzen

Um Innovations- und additive Ausreißer separat behandeln zu können, ist es wie schon beim rIC-Filter nützlich, eine Partition der Scoresfunktion in einen AO- und einen IO-Teil vorzunehmen. Dies wird durch die Zerlegung in die Summanden Λ^{AO} und Λ^{IO} erreicht, wie sie bereits in (4) blockweise eingeführt worden sind. Dabei steht Λ^{AO} für AO's, Λ^{IO} für IO's. Im idealen Modell (V3) bzw. (V4) sind Λ^{AO} und Λ^{IO} unabhängige, $\mathbb{R}^{(m+1)p}$ -wertige Zufallsvariablen mit Mittelwert 0 und Kovarianzmatrix

$$\mathcal{I}^{(\text{AO})} := \text{diag}[(Z_t^T V_t^{-1} Z_t)]_{t=T-m, \dots, T} \quad (4.14)$$

bzw.

$$\begin{aligned} \mathcal{I}^{(\text{IO})} := & \text{diag} \left[\begin{array}{l} \Sigma_{T-m|T-m-1}^{\text{KK}^{-1}} + F_{T-m+1}^T Q_{T-m+1}^{-1} F_{T-m+1}, \\ (Q_t^{-1} + F_{t+1}^T Q_{t+1}^{-1} F_{t+1})_{T-m < t < T}, Q_T^{-1} \end{array} \right] + \\ & + \text{ndiag}_1 \left[\begin{array}{l} (-F_t Q_t^{-1})_{T-m < t \leq T} \end{array} \right] + \\ & + \text{ndiag}_1 \left[\begin{array}{l} (-F_t Q_t^{-1})_{T-m < t \leq T} \end{array} \right]^T. \end{aligned} \quad (4.15)$$

Genau wie beim rIC-Filter können wir nun ψ definieren als

$$\psi = \psi_\beta(\hat{y}) = A \left[\Lambda_\beta^{\text{AO}}(\cdot) \min\left\{1, \frac{b^{\text{AO}}}{|A\Lambda_\beta^{\text{AO}}|}\right\} + \Lambda_\beta^{\text{IO}}(\cdot) \min\left\{1, \frac{b^{\text{IO}}}{|A\Lambda_\beta^{\text{IO}}|}\right\} \right] \quad (4.16)$$

mit geeignet gewählten $b^{(i)} > 0$ und mit einer Matrix A , die für die Korrelationsbedingung (9) sorgt.

4.4 blockweises Stutzen

In einer weiteren Variante von (8) führt man das Stutzen blockweise in den $m+1$ Blöcken durch. Im Zusammenhang mit allgemeinen Influenzkurven entspricht es dieser Variante am ehesten, wenn man im Optimierungsproblem $O_{c;\infty}^{\text{tr}}(b)$ aus Rieder (1994) koordinatenweise Schranken an den asymptotischen Bias auferlegt (vgl. Rieder (1994), Rem. 5.5.4, und Lem. 5.3.2). Für unser Problem bedeutet dies eine beträchtliche Reduktion des Rechenaufwandes, ja es ist für größere Dimensionen m , p und q sogar die einzige, in vertretbarem Aufwand berechenbare Variante.

4.4.1 Vorarbeiten

Für das blockweise Stutzen führen wir folgende Schreibweise für blockweise indizierte Matrizen und Vektoren ein:

Schreibweise 4.4.1 Für eine Matrix $A \in \mathbb{R}^{(m+1)p \times (m+1)p}$ bezeichne $A_{(s),(t)}$ die $p \times p$ Untermatrix $(A_{i,j})_{\substack{i=(s-1)p \dots sp \\ j=(t-1)p \dots tp}}$, und für einen Vektor $b \in \mathbb{R}^{(m+1)p}$ bezeichne $b_{(s)}$ den p -Block $(b_i)_{i=(s-1)p \dots sp}$. Auf diese Weise läßt sich das Matrixprodukt $C = AB$ solcher Matrizen als das blockweise Produkt von $p \times p$ -Untermatrizen schreiben:

$$C = (C_{(r),(t)}) = \left(\sum_{s=1}^{m+1} A_{(r),(s)} B_{(s),(t)} \right)$$

Der Block $A^{(s),(t)}$ bezeichne für invertierbare $A \in \mathbb{R}^{(m+1)p \times (m+1)p}$ die entsprechende $p \times p$ -Untermatrix der inversen Matrix A^{-1} . $\tilde{\Lambda}_{(t)}(\cdot)$ schließlich sei die kanonische Einbettung von $\Lambda_{(t)}(\cdot)$ in $\mathbb{R}^{(m+1)p}$, das heißt mit $\mathbf{0}_p = 0 \in \mathbb{R}^p$

$$\tilde{\Lambda}_{(t)}(\cdot) = (\mathbf{0}_{(t-1)p}^\tau, (\Lambda_{(t)}(\cdot))^\tau, \mathbf{0}_{(m+1-t)p}^\tau)^\tau. \quad (4.17)$$

4.4.2 blockweise definierte Scores, IC's und Fisher-Information

Wir betrachten nun Funktionen $\psi \in L_2^{(m+1)p}$ von der Gestalt

$$\psi = \psi_b(\cdot) = A \sum_{t=T-m}^T \tilde{\Lambda}_{(t)}(\cdot) \min\left\{1, \frac{b_t}{|A\tilde{\Lambda}_{(t)}(\cdot)|}\right\} \quad (4.18)$$

für ein $b = (b_{T-m:T}) \in \mathbb{R}_{>0}^{m+1}$ und eine Matrix $A \in \mathbb{R}^{(m+1)p \times (m+1)p}$, die für (9) sorgt, d.h.

$$E_{\text{id}}[\psi_b \Lambda^\tau] = \mathbb{I}_{(m+1)p}. \quad (4.19)$$

Diese Variante bewahrt die Tridiagonalgestalt von A^{-1} :

Weil A im idealen Modell (V3) bzw. (V4) bestimmt wird, ist $(\Lambda_{(i)}^\tau, \Lambda_{(j)}^\tau)^\tau$ normalverteilt für alle i, j . Wir haben schon in (5) gesehen, daß $\text{Cov}[\Lambda_{(i)}, \Lambda_{(j)}] = 0$ für $|i - j| > 1$, also sind diese Blöcke dann auch stochastisch unabhängig.

Dies gilt auch für die Einbettungen, daher haben wir wegen des symmetrischen Clippings

$$(A^{-1})_{(i),(j)} = \sum_{i,j=T-m}^T \mathbb{E}[\tilde{\Lambda}_{(i)} \tilde{\Lambda}_{(j)}^\tau \min\{1, \frac{b_i}{|A\tilde{\Lambda}_{(i)}}\}] = \quad (4.20)$$

$$= 0 \quad \text{für } |i-j| > 1. \quad (4.21)$$

Also müssen wir, um A zu berechnen nur die $m+1$ Integrale auf der Diagonalen $\mathbb{E}[\Lambda_{(i)} \Lambda_{(i)}^\tau \min\{1, \frac{b_i}{|A\Lambda_{(i)}}\}]$ berechnen, weil wir wegen der gemeinsamen Normalität schreiben dürfen:

$$\begin{aligned} & \mathbb{E}[\Lambda_{(i+1)} \Lambda_{(i)}^\tau \min\{1, \frac{b_i}{|A\Lambda_{(i)}}\}] = \\ &= \mathbb{E} \left[\mathbb{E}[\Lambda_{(i+1)} | \Lambda_{(i)}] \Lambda_{(i)}^\tau \min\{1, \frac{b_i}{|A\Lambda_{(i)}}\} \right] = \\ &= FQ^{-1} \mathbb{E}[\Lambda_{(i)} \Lambda_{(i)}^\tau]^{-1} \mathbb{E} \left[\Lambda_{(i)} \Lambda_{(i)}^\tau \min\{1, \frac{b_i}{|A\Lambda_{(i)}}\} \right] \end{aligned} \quad (4.22)$$

Diese Integration kann nun mit Techniken, wie sie beim rIC-Fall besprochen worden sind, numerisch berechnet werden.

4.4.3 Kalibrierung

Eine wichtige Frage, die sich in diesem Zusammenhang stellt, ist die Wahl der Stutzhöhen b_t für $t = T \dots T-m$.

Anders als beim rIC-Filter können wir nicht einfach auf das ‘‘Effizienzkriterium’’, d.h. $\mathbb{E}_{\text{id}}[|\psi|^2] \stackrel{!}{=} (1+\delta) \text{tr } \mathcal{I}^{-1}$ für einen gewissen ‘‘Effizienzverlust’’ $\delta > 0$ zurückgreifen, weil dies nur einen Freiheitsgrad eliminieren würde.

Stattdessen könnte man versuchen, dieses Kriterium blockweise für jeden p -Block $\psi_{(t)}$, $t = T \dots T-m$ vorzuschreiben. Um zu zeigen, daß wir dies tun können, benutzen wir folgendes Lemma

Lemma 4.4.2 *Sei $\psi \in L_2^{(m+1)p}$, so daß*

$$\mathbb{E}_{\text{id}}[\psi \Lambda^\tau] = \mathbb{I}_{(m+1)p} \quad (4.23)$$

gilt. Dann haben wir Cramér-Rao-Schranken derart, daß

$$\mathbb{E}_{\text{id}}[\hat{\psi}_{(t)} \hat{\psi}_{(t)}^\tau] \preceq \mathbb{E}_{\text{id}}[\psi_{(t)} \psi_{(t)}^\tau] \quad \forall t = T-m \dots T. \quad (4.24)$$

Daraus folgt unmittelbar

Korollar 4.4.3

$$\text{tr } \mathcal{I}^{(t),(t)} = \mathbb{E}_{\text{id}}[|\hat{\psi}_{(t)}|^2] \leq \mathbb{E}_{\text{id}}[|\psi_{(t)}|^2] \quad \forall t = T-m \dots T \quad (4.25)$$

Inbesondere gibt es für jedes $\delta \in (0; \delta_{\min})$, und für ψ von der Gestalt (18) einen Vektor $b = (b_t)_{t=T-m \dots T} \in \mathbb{R}_{>0}^{m+1}$, so daß

$$\mathbb{E}_{\text{id}}[|\psi_{(t)}(\cdot, b)|^2] = (1+\delta) \text{tr } \mathcal{I}^{(t),(t)} \quad \forall t = T-m \dots T. \quad (4.26)$$

BEWEIS VON LEMMA 4.4.2: Wir zeigen

$$\mathbb{E}_{\text{id}}[\hat{\psi}_{(t)}\psi_{(t)}^\tau] = \mathbb{E}_{\text{id}}[\psi_{(t)}\hat{\psi}_{(t)}^\tau] = \mathbb{E}_{\text{id}}[\hat{\psi}_{(t)}\hat{\psi}_{(t)}^\tau]. \quad (4.27)$$

Dann gilt auch

$$\mathbb{E}_{\text{id}}[\psi_{(t)}\psi_{(t)}^\tau] - \mathbb{E}_{\text{id}}[\hat{\psi}_{(t)}\hat{\psi}_{(t)}^\tau] = \mathbb{E}_{\text{id}}[(\psi - \hat{\psi})_{(t)}(\psi - \hat{\psi})_{(t)}^\tau] \succeq 0.$$

Um (27) zu zeigen, bemerken wir, daß wegen (19) $\mathbb{E}[\psi_{(s)}\Lambda_{(t)}^\tau] = \delta_{s,t}\mathbb{I}_p$ mit $\delta_{s,t}$ dem Kronecker- δ . Nun ist

$$\begin{aligned} \mathbb{E}_{\text{id}}[\psi_{(t)}\hat{\psi}_{(t)}^\tau] &= \mathbb{E}_{\text{id}}[\psi_{(t)}(I^{-1}\Lambda)_{(t)}^\tau] = \\ &= \sum_{s=T-m}^T \mathbb{E}_{\text{id}}[\psi_{(t)}\Lambda_{(s)}^\tau]\mathcal{I}^{(t),(s),\tau} = \sum_{s=T-m}^T \delta_{s,t}\mathbb{I}_p\mathcal{I}^{(s),(t)} = \mathcal{I}^{(t),(t)} \end{aligned}$$

und

$$\begin{aligned} \mathbb{E}_{\text{id}}[\hat{\psi}_{(t)}\hat{\psi}_{(t)}^\tau] &= \mathbb{E}_{\text{id}}[(\mathcal{I}^{-1}\Lambda)_{(t)}(\mathcal{I}^{-1}\Lambda)_{(t)}^\tau] = \\ &= \sum_{r,s=T-m}^T \mathcal{I}^{(t),(r)}\mathbb{E}_{\text{id}}[\Lambda_{(r)}\Lambda_{(s)}^\tau]\mathcal{I}^{(s),(t)} = \sum_{r,s=T-m}^T \mathcal{I}^{(t),(r)}\mathcal{I}_{(r),(s)}\mathcal{I}^{(s),(t)} = \\ &= \sum_{s=T-m}^T \delta_{s,t}\mathbb{I}_p\mathcal{I}^{(s),(t)} = \mathcal{I}^{(t),(t)}. \end{aligned}$$

$\mathbb{E}_{\text{id}}[\hat{\psi}_{(t)}\psi_{(t)}^\tau] = \mathcal{I}^{(t),(t)}$ folgt aus Symmetriegründen. ////

4.4.4 Verwendung numerischer Verfahren für Blockmatrizen

Beim Bestimmen des Lagrange-Multiplikators A für die blockweise gestutzten Influenzkurven müssen wir in der Lage sein $A\Lambda$ schnell und numerisch stabil auswerten zu können, wozu wir entweder eine Darstellung von A selbst oder von A^{-1} zur Verfügung haben müssen. Wegen der hohen Dimensionen — $A \in \mathbb{R}^{(T+1)p \times (T+1)p}$ — bei Verwendung eines wachsenden Fensters, benötigen wir Verfahren, die die spezielle Struktur von A , genauer eine Block-Tridiagonalgestalt von A^{-1} , ausnutzen. Hierzu wollen wir die Verfahren aus Anhang E verwenden.

Nicht-Symmetrie von A

Leider können wir im Fall der blockweise gestutzten Influenzkurven die Symmetrie der Matrix A nicht garantieren. Um dies einzusehen, betrachten wir A^{-1} . Mit der Schreibweise 4.4.1 haben wir

$$A^{(i),(j)} = \begin{cases} \mathbb{E}[\Lambda_{(i)}\Lambda_{(i)}^\tau]w_i & \text{für } j = i \\ \mathbb{E}[\Lambda_{(i)}\Lambda_{(i+1)}^\tau]w_i & j = i + 1 \leq T \\ \mathbb{E}[\Lambda_{(i)}\Lambda_{(i-1)}^\tau]w_i & j = i - 1 \geq T - m \\ 0 & \text{sonst} \end{cases} \quad (4.28)$$

mit $w_i = \min\{1, \frac{b_i}{|A\Lambda_{(i)}|}\}$. Im idealen Modell (V3) bzw (V4) gilt

$$\begin{pmatrix} \Lambda_{(i)} \\ \Lambda_{(i+1)} \end{pmatrix} \sim \mathcal{N}_{2p}(0, \begin{pmatrix} C_i & V_{i,i+1} \\ V_{i,i+1}^\tau & C_{i+1} \end{pmatrix}). \quad (4.29)$$

Daher folgt mit Lemma A.2.1

$$\begin{aligned} \Lambda_{(i)|(i+1)} &:= \mathbb{E}[\Lambda_{(i)}|\Lambda_{(i+1)}] = V_{i,i+1}C_{i+1}^{-1}\Lambda_{(i+1)} \\ \Lambda_{(i+1)|(i)} &:= \mathbb{E}[\Lambda_{(i+1)}|\Lambda_{(i)}] = V_{i,i+1}^\tau C_i^{-1}\Lambda_{(i)} \end{aligned} \quad (4.30)$$

Indem wir iterierte Erwartungswerte verwenden, erhalten wir

$$\begin{aligned} A^{(i),(i+1)} &= \mathbb{E}[\Lambda_{(i)}\Lambda_{(i+1)}^\tau w_i] = \mathbb{E}[\Lambda_{(i)}\Lambda_{(i+1|i)}^\tau w_i] = A^{(i),(i)}C_i^{-1}V_{i,i+1} \\ (A^{(i+1),(i)})^\tau &= (\mathbb{E}[\Lambda_{(i+1)}\Lambda_{(i)}^\tau w_{i+1}])^\tau = V_{i,i+1}C_{i+1}^{-1}A^{(i+1),(i+1)} \end{aligned} \quad (4.31)$$

Daher muß für die Symmetrie von A gelten

$$A^{(i),(i)}C_i^{-1}V_{i,i+1} = V_{i,i+1}C_{i+1}^{-1}A^{(i+1),(i+1)}, \quad (4.32)$$

was keineswegs klar ist.

Bemerkung 4.4.4 Im klassischen Fall, also $b_i = \infty$, gilt $A^{(i),(i)} = \mathbb{E}[\Lambda_{(i)}\Lambda_{(i)}^\tau] = C_i$, so daß wir für $b_i = b_{i+1} = \infty$ wieder die Symmetrie der Fisher-Information \mathcal{I} erhalten.

Anwendbarkeit der LR-Zerlegung auf A^{-1} ohne Spaltenpivotierung

Sofern nur die Fisher-Information \mathcal{I} aus (5) vollen Rang besitzt, zeichnet sich A^{-1} aus (28) dadurch aus, daß in der blockweise durchgeführten LR-Zerlegung keine Spaltenvertauschungen nötig sind, weil die in Algorithmus E.2.2 auftretenden Matrizen $D_{(i),(i)}$ für jedes i invertierbar sind. Demnach ist die LDM-Zerlegung E.2.2 anwendbar. Um dies zu zeigen, führen wir zunächst eine Gram-Schmidt-Orthogonalisation der $\Lambda_{(i)}$ ein. Wir setzen hierzu²

$$x_{(i)} := \begin{cases} \Lambda_{(T-m)} & \text{für } i = T - m, \\ \Lambda_{(i)} - \text{oP}(\Lambda_{(i)}|x_{(i-1)}) & \text{für } i = T - m + 1 \dots T \end{cases} \quad (4.33)$$

Wie in Abschnitt 1.2.3 sehen wir, daß

$$\text{oP}(\Lambda_{(i)}|x_{(i-1)}) = \mathbb{E}[\Lambda_{(i)}x_{(i-1)}^\tau]^{-1} \mathbb{E}[x_{(i-1)}x_{(i-1)}^\tau]^{-1}x_{(i-1)}. \quad (4.34)$$

Nach diesem Vorgehen bilden die x_i offensichtlich eine orthogonale Familie. Im idealen Modell (V3) bzw. (V4) gilt

$$\begin{pmatrix} x_{(i-1)} \\ \Lambda_{(i)} \end{pmatrix} \sim \mathcal{N}_{2p}(0, \begin{pmatrix} X_{i-1} & G_{i-1,i} \\ G_{i-1,i}^\tau & C_i \end{pmatrix}) \quad \text{für } i = T - m + 1 \dots T \quad (4.35)$$

und wegen der Normalität ist $\text{oP}(\Lambda_{(i)}|x_{(i-1)}) = \mathbb{E}[\Lambda_{(i)}|x_{(i-1)}]$. Weiterhin gilt

²Wegen der Block-Tridiagonalgestalt ist $\text{oP}(\Lambda_{(i)}|x_{(0):(i-1)}) = \text{oP}(\Lambda_{(i)}|x_{(i-1)})$.

Lemma 4.4.5

$$G_{i-1,i} = V_{i-1,i} \quad \text{für } i = T - m + 1 \dots T \quad (4.36)$$

und falls $\det \mathcal{I} \neq 0$,

$$\text{rk } X_i = p \quad \text{für } i = T - m \dots T \quad (4.37)$$

$$X_i = C_i - V_{i-1,i}^\tau X_{i-1}^{-1} V_{i-1,i} \quad (4.38)$$

BEWEIS: Für $i = T - m + 1$ gilt (36) nach Definition. Für $i > T - m + 1$ ist

$$\begin{aligned} \mathbb{E}[x_{(i-1)} \Lambda_{(i)}^\tau] &= \mathbb{E}[(\Lambda_{(i-1)} - \mathbb{E}[\Lambda_{(i-1)} | x_{(i-2)}]) \Lambda_{(i)}^\tau] \\ &= V_{i-1,i} - \mathbb{E}[\mathbb{E}[\Lambda_{(i-1)} | x_{(i-2)}] \Lambda_{(i)}^\tau]. \end{aligned}$$

Nun ist $\mathbb{E}[\Lambda_{(i-1)} | x_{(i-2)}]$ meßbar in $x_{(i-2)}$, was wiederum $\Lambda_{(1):(i-2)}$ -meßbar ist. Weil aber $\Lambda_{(i)}$ von $\Lambda_{(1):(i-2)}$ stochastisch unabhängig ist, folgt (36) aus der Zentriertheit von $\Lambda_{(i)}$.

Was (37) anlangt, so sehen wir, daß $\Lambda_{(i)} = x_{(i)} + \sum_{j < i} K_{i,j} x_{(j)}$ für gewisse reellwertige $p \times p$ Matrizen $K_{i,j}$. In Matrixform erhalten wir $\Lambda = Kx$. Nach Konstruktion ist K eine untere Dreiecksmatrix mit 1 auf der Hauptdiagonale, so daß $\det K = 1$. Daher hat $X = \text{diag}[X_i]$ dieselbe Determinante wie \mathcal{I} , und alle X_i haben vollen Rang. (38) sieht man durch Einsetzen in (34). ////

Die Anwendbarkeit des LDM-Algorithmus' E.2.2 wird nun abgesichert durch

Proposition 4.4.6

$$\text{rk } A^{(i),(i)} = p \quad (4.39)$$

$$D_{(i),(i)} = A^{(i),(i)} C_i^{-1} X_i \quad (4.40)$$

Daher ist der LDM-Algorithmus E.2.2 wohldefiniert, denn

Korollar 4.4.7

$$\det D_{(i),(i)} = \det A^{(i),(i)} \det X_i / \det C_i \neq 0 \quad (4.41)$$

BEWEIS ZU PROPOSITION 4.4.6: Offenbar ist $A^{(i),(i)} = \mathbb{E}[(\Lambda_{(i)} \sqrt{w_i})(\Lambda_{(i)} \sqrt{w_i})^\tau]$ positiv semidefinit. Damit ist $A^{(i),(i)}$ genau dann singular, wenn es einen Einheitsvektor $e \in \mathbb{R}^p$ gibt, so daß fast sicher $(e^\tau \Lambda_{(i)} \sqrt{w_i}) = 0$. Dies aber impliziert $(e^\tau \Lambda_{(i)}) = 0$ fast sicher, denn $w_i > 0$ für endliches $\|A \tilde{\Lambda}_{(i)}\|$, was auch fast sicher der Fall ist, da $\|A^{-1}\| > 0$. Nun ist $\det C_i \neq 0$, so daß $(e^\tau \Lambda_{(i)}) \neq 0$ fast sicher. Dies zeigt (39).

Für (40) führen wir eine vollständige Induktion in i .

Für $i = T - m$ gilt $X_{T-m} = C_{T-m}$, also ist $D_{(T-m),(T-m)} = A^{(T-m),(T-m)} = A^{(T-m),(T-m)} C_{T-m}^{-1} X_{T-m}$ nach Definition. Nehmen wir an, die Behauptung gelte für $j < i$. Dann ist

$$\begin{aligned} D_{(i),(i)} &\stackrel{\text{Def.}}{=} A^{(i),(i)} - A^{(i),(i-1)} D_{(i-1),(i-1)}^{-1} A^{(i-1),(i)} = \\ &\stackrel{\text{Ind.}}{=} A^{(i),(i)} - A^{(i),(i-1)} (A^{(i-1),(i-1)} C_{i-1}^{-1} X_{i-1})^{-1} A^{(i-1),(i)} = \\ &\stackrel{(5.31)}{=} A^{(i),(i)} - A^{(i),(i)} C_i^{-1} V_{i,i-1}^\tau (A^{(i-1),(i-1)} C_{i-1}^{-1} X_{i-1})^{-1} A^{(i-1),(i-1)} C_{i-1}^{-1} V_{i,i-1} = \\ &= A^{(i),(i)} - A^{(i),(i)} C_i^{-1} V_{i,i-1}^\tau X_{i-1}^{-1} V_{i,i-1} = \\ &= A^{(i),(i)} C_i^{-1} (C_i - V_{i,i-1}^\tau X_{i-1}^{-1} V_{i,i-1}) \stackrel{(5.38)}{=} A^{(i),(i)} C_i^{-1} X_i \end{aligned}$$

////

4.5 weitere Varianten: Standardisierung in anderen Normen

Wie in Abschnitt 3.3.2 beim rIC-Filter ist es auch möglich, die Dimension des Problems, den Lagrange-Multiplikator A und die entsprechende Stutzhöhe b zu finden, im idealen Modell (V3) bzw. (V4) auf 2 zu reduzieren; die Argumente hierfür gelten wörtlich auch für die $(m+1)p$ -dimensionale normalverteilte Scoresfunktion Λ . Auch ein separates Stutzen wie in Abschnitt 3.3.3 in äquivarianten Normen ist möglich. Allerdings ist das so vorgenommene Stutzen natürlich längst nicht so differenziert wie bei den blockweise gestutzten Varianten; genauer ist wegen der hohen Dimension von $(m+1)p$ Korollar D.4.2 zu beachten. Für die blockweise gestutzten Varianten ließen sich dann allerdings keine adäquaten Formulierungen für äquivariante Normen finden.

Schreibweise 4.5.1 Um die verschiedenen Varianten des mIC-Glätters und -Filters zu unterscheiden, verwenden wir folgende Notation

mICuvw	mit		
u	=	b	$\hat{=}$ blockweises Stutzen
		g	$\hat{=}$ gemeinsames Stutzen (keine Blöcke)
v	=	t	$\hat{=}$ totales Stutzen (AO,IO simultan)
		s	$\hat{=}$ separates Stutzen (AO,IO separat)
w	=	u	$\hat{=}$ unstandardisiert; $ \cdot $ ist euklid. Norm
		i	$\hat{=}$ informations-standardisiert
		s	$\hat{=}$ selbststandardisiert

4.6 Beispiele

Anders als bei den rekursiven Filtern rLS und rIC sind die mIC-Varianten (noch) nicht in `xploRe` integriert. Die Beispiele zum mIC-Filter und -Glätter stammen daher aus der Arbeit mit H. Rieder im DFG-Projekt DFG Ri 332/8-1 im Zeitraum 1997–99. Wir arbeiten dabei stets unter (V4) und beschränken uns auf die blockweise unstandardisierten Varianten mICbtu und mICbsu.

Beispiel 4.6.1 Als erstes Beispiel verwenden wir als ideales Modell das Steady-State Modell aus Beispiel 1.4.1, also

$$a_0 = 0, \quad Q_0 = 0, \quad Q = 1, \quad V = 1, \quad F = 1, \quad Z = 1.$$

Wir verwenden ein festes Fenster der Länge $m = 60$ und kalibrieren den mIC auf $\delta = 10\%$. Als Kontamination betrachten wir folgende Kombination:

$$\begin{aligned} v_t &\stackrel{\text{u.i.v.}}{\sim} (1 - r_{\text{IO}})\mathcal{N}(0, 1) + r_{\text{IO}}\mathcal{N}(10, 100), \\ \varepsilon_t &\stackrel{\text{u.i.v.}}{\sim} (1 - r_{\text{AO}})\mathcal{N}(0, 1) + r_{\text{AO}}\mathcal{N}(10, 100), \end{aligned} \quad (4.42)$$

mit $r_{\text{IO}} = 5\%$, $r_{\text{AO}} = 5\%$. Zusätzlich unterliegen die Zustände noch einer weiteren Kontamination, die wir *LO-Kontamination* — LO für Level-Ausreißer

— nennen wollen: Anstelle des idealen Zustandes β wird im Zustandsraum ein Zustand $\hat{\beta}_t = \beta_t + \tilde{l}_t$ definiert, und \tilde{l}_t baut sich unabhängig vom ε - und vom v -Prozess auf als $\tilde{l}_0 = 0$, $\tilde{l}_t = \tilde{l}_{t-1} + w_t$,

$$w_t \stackrel{\text{u.i.v.}}{\sim} \begin{cases} l_{\text{LO}} & r_{\text{LO}}/2 \\ 0 & \text{mit Ws } 1 - r_{\text{LO}} \\ -l_{\text{LO}} & r_{\text{LO}}/2 \end{cases}$$

In unserem Beispiel ist $r_{\text{LO}} = 10\%$ und $l_{\text{LO}} = 5$. Unter einer solchen Kontamination entstand Abbildung 4.1: Die Zustände unter Kontamination lösen sich auf Dauer vom Verlauf unter dem idealen Modell ab.

Im idealen Modell liegen die mIC-Filter, d.h. $\beta_{t|t}^{\text{mIC}}$ in den Varianten mICbtu

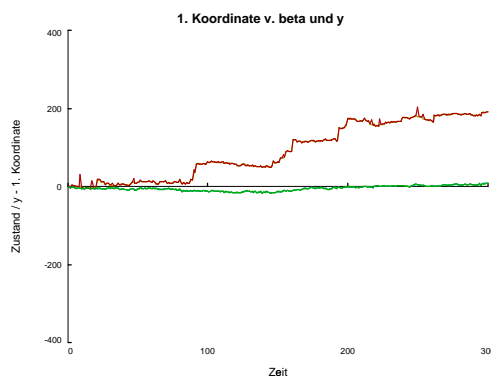


Abbildung 4.1: Zustände im Steady-State Modell bei LO's

Der Verlauf im idealen Modell, etwas heller dargestellt, pendelt im wesentlichen um 0, während bei LO-Kontamination — etwas heller — sich das Grundniveau stufenweise nach oben bewegt.

und mICbsu, der tatsächliche Zustand β_t und der klassische Kalman-Filter $\beta_{t|t}^{\text{KK}}$ in etwa im gleichen Bereich.

Innerhalb der vorliegenden Kontamination sind offenbar die IO- und LO-Komponenten die gravierenderen; der unspezialisierte mICbtu folgt den Niveausprüngen des Zustandes viel zu langsam. In dem hier dargestellten Beispiel folgt — wohl wegen der sehr günstigen Signal-to-Noise-Ratio $Q/V = 1$ — der klassische Kalman-Filter dem gestörten Signal sogar noch schneller als der robuste mICbsu, der aber auch ansprechende Werte liefert, wie wir in Abbildung 4.2 rechts feststellen. Wie in Abbildung 4.3 deutlich wird, ist das Bild bei den “extremen” Glätttern, $\beta_{t-m|t}^{\text{KK}}$ und $\beta_{t-m|t}^{\text{mIC}}$, ganz ähnlich: Im idealen Modell fallen alle Methoden zusammen; unter Kontamination fallen klassischer Kalman-Glätter und mICbsu zusammen, mICbtu reagiert etwas träger; dadurch aber, daß zur Schätzung von β_t auch noch die Beobachtungen $y_{t:t+59}$ zur Verfügung stehen, werden insgesamt bessere Ergebnisse erzielt als beim Filtern in Abbildung 4.2.

In Abbildung 4.4 stellen wir zu einem festem Endzeitpunkt $T = 140$, den Verlauf von $\beta_{t-s|t}^{\text{KK}}$ und $\beta_{t-s|t}^{\text{mIC}}$, für $s = 0, \dots, m = 60$ dar: Im idealen Modell fallen alle Methoden zusammen; unter Kontamination sind sie fast ununterscheidbar.

Beispiel 4.6.2 Als zweites Beispiel verwenden wir als ideales Modell das zeit-invariante Modell mit $p = q = 1$

$$a_0 = 0, \quad Q_0 = 0, \quad Q = 0.2, \quad V = 5, \quad F = 1, \quad Z = 1, \quad (4.43)$$

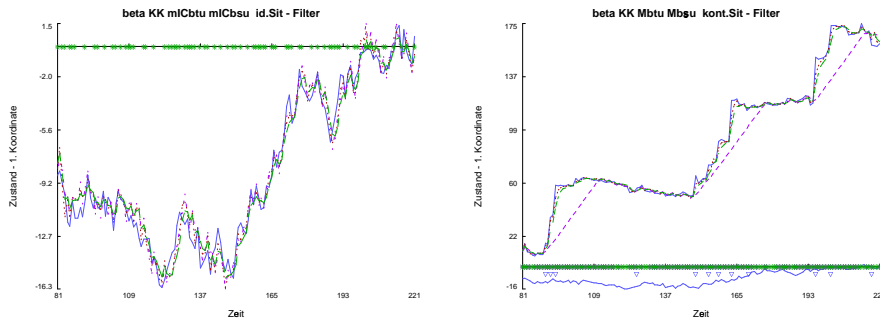


Abbildung 4.2: mIC- und KK-Filter im idealen Modell (links) und unter Kontamination (rechts);

durchgezogen β_t , gepunktet $\beta_{t|t}^{KK}$, kurz-gestrichelt $\beta_{t|t}^{mICbtu}$, lang-gestrichelt $\beta_{t|t}^{mICbsu}$;

man beachte, daß die Ordinatenskala links um Faktor 10 kleiner ist als rechts! Zeitpunkte, an denen der mIC-Filter stutzt, sind durch Sternchen auf der Linie $\beta = 0$, Zeitpunkte an denen Kontamination auftritt mit Dreiecken darunter markiert; ganz unten am Rand der rechten Graphik: der ideale Zustand;

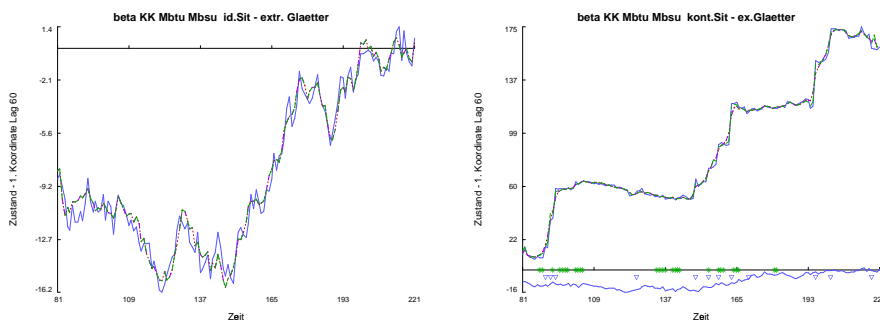


Abbildung 4.3: $\beta_{t-m|t}^{KK}$ und $\beta_{t-m|t}^{mIC}$ im idealen Modell (links) und unter Kontamination (rechts);

Linientypen, Skala, Stutz- und Kontaminationsmarkierung wie in Abbildung 4.2;

also mit einer ungünstigeren Signal-to-Noise-Ratio $Q/V = 25$ als in Beispiel 4.6.1. Wir verwenden ein festes Fenster der Länge $m = 25$ und kalibrieren den mIC auf $\delta = 100\%$. Als Kontamination betrachten wir folgende Konstruktion:

$$\hat{\beta}_t = F\hat{\beta}_{t-1} + 200I_{\{t=199\}} + v_t, \quad \hat{\beta}_0 = \beta_0, \quad \hat{y}_t = Z\hat{\beta}_t + \varepsilon_t, \quad (4.44)$$

das heißt der Zustand zum Zeitpunkt 199 wird um 200 erhöht, ansonsten entwickeln sich Zustand und Beobachtung gemäß dem idealen Modell. Unter einer solchen Kontamination entstand Abbildung 4.5.

Wie in Beispiel 4.6.1 liegen die mIC-Filter, d.h. $\beta_{t|t}^{mIC}$ in den Varianten mICbtu und mICbsu, der tatsächliche Zustand β_t und der klassische Kalman-Filter $\beta_{t|t}^{KK}$ im idealen Modell in etwa im gleichen Bereich.

Am Niveausprung des Zustandes in $t = 199$ bei den kontaminierten Daten

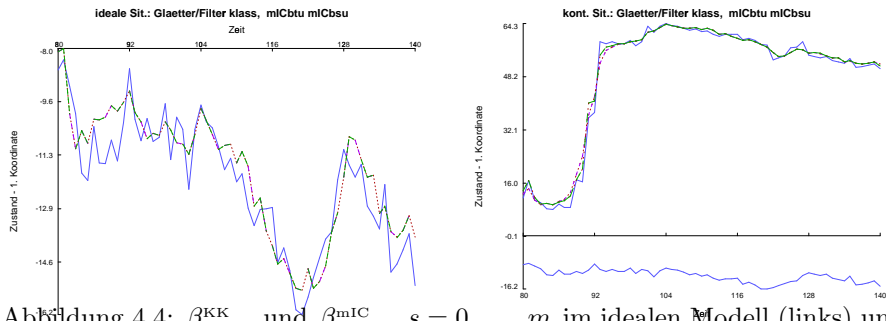


Abbildung 4.4: $\beta_{t-s|t}^{\text{KK}}$ und $\beta_{t-s|t}^{\text{mIC}}$, $s = 0, \dots, m$ im idealen Modell (links) und unter Kontamination (rechts);
Linientypen wie in Abbildung 4.2; man beachte auch hier, die unterschiedlichen Ordinatenachsen links und rechts;

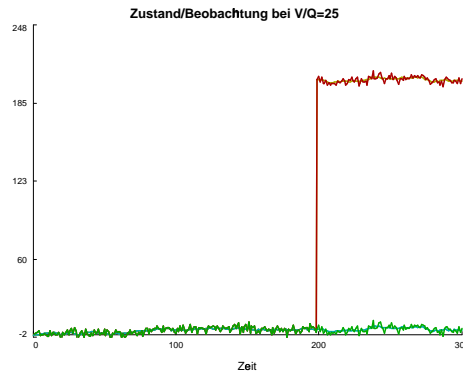


Abbildung 4.5: Zustände in Beispiel 4.6.2 bei Ausreißer in $t = 199$;
Farbgebung wie in Abbildung 4.1

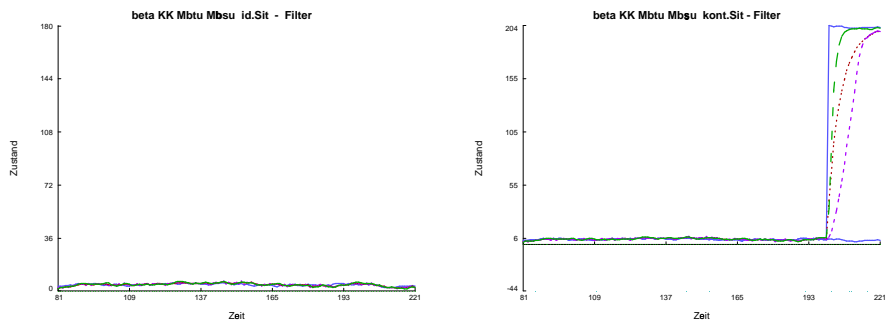


Abbildung 4.6: mIC- und KK-Filter in Beispiel 4.6.2 im idealen Modell (links) und unter Kontamination (rechts);
Linientypen wie in Abbildung 4.2;

fördern die Filter aber dann ein unterschiedliches Verhalten zutage. Der unspezialisierte mICbtu folgt dem Niveausprung am langsamsten; anders als in Beispiel 4.6.1 schneidet aber auch der klassische Kalman-Filter — wohl wegen der ungünstigeren Signal-to-Noise-Ratio — wesentlich schlechter ab als der mICbsu, der auch in dieser Situation noch ansprechende Werte liefert. Dies ist in Abbildung 4.6 dargestellt. Ganz ähnlich ist das Bild — vgl. Abbildung 4.7

— bei den “extremen” Glättern, $\beta_{t-m|t}^{KK}$ und $\beta_{t-m|t}^{mIC}$, wobei hier der unspezialisierte mICbtu unter Kontamination sogar zum klassischen Kalman–Glätter aufschließt.

In Abbildung 4.8 schließlich stellen wir wieder zu einem festem Endzeitpunkt

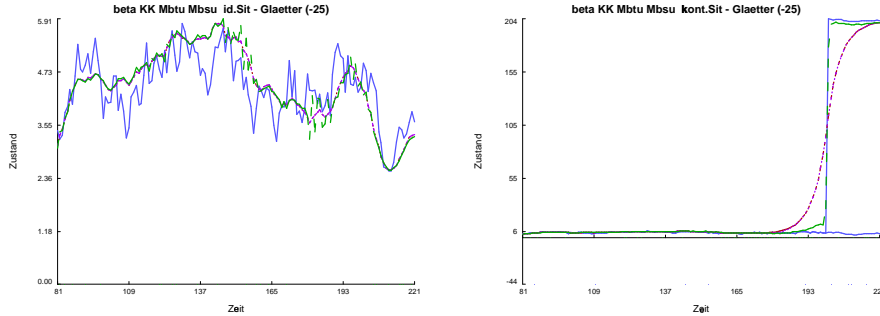


Abbildung 4.7: $\beta_{t-m|t}^{KK}$ und $\beta_{t-m|t}^{mIC}$ in Beispiel 4.6.2 im idealen Modell (links) und unter Kontamination (rechts);
Linientypen wie in Abbildung 4.2; man beachte die unterschiedlichen Ordinaten–Skalen!

$T = 140$, den Verlauf von $\beta_{t-s|t}^{KK}$ und $\beta_{t-s|t}^{mIC}$, für $s = 0, \dots, m = 60$ dar.

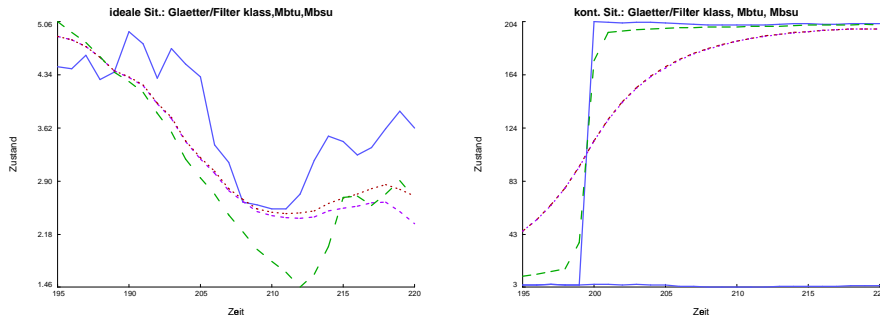


Abbildung 4.8: $\beta_{t-s|t}^{KK}$ und $\beta_{t-s|t}^{mIC}$, $s = 0, \dots, m$ in Beispiel 4.6.2 im idealen Modell (links) und unter Kontamination (rechts);
Linientypen wie in Abbildung 4.2; man beachte die unterschiedlichen Ordinaten–Skalen!

4.7 Bewertung

Im Vergleich zu den strikt rekursiven rLS– und rIC–Filtern ist der mIC–Glätter und –Filter sowohl in der Auswertung als auch in der Kalibrierung aufwendiger. Dadurch daß man aber eine detailliertere Information über die Filtervergangenheit verwendet, ist dieses Verfahren als einziges der bisher eingeführten in der Lage auch bei einer ungünstigen Signal–to–Noise–Ratio mit Strukturbrüchen/IO’s zurechtzukommen.

kein “batch”–Verfahren

Indem man nicht die gesamte Vergangenheit $y_{1:T}$ verwendet, sondern sich auf

ein Fenster der festen Länge m beschränkt, handelt es sich bei dieser Variante des mIC dennoch nicht um ein “batch”-Verfahren³, ja man kann nach der notwendigen “Einschwingphase” der Länge m von einem generalisiert-rekursiven Verfahren sprechen. Diese Eigenschaft grenzt den mIC insbesondere vom Posterior-Modus-Schätzer ab, wie er sich in Künstler (1995), Fahrmeir und Künstler (1999) dargestellt findet.

Kalibrierung “off-line”, Auswertung “on-line”

Wie beim rIC-Filter kann bei bekannten Systemmatrizen die rechentechnisch noch aufwendigere Bestimmung der Influenzkurven ψ_t^{mIC} “off-line” geschehen. Die eigentliche Auswertung des Filters an den Beobachtungen y_t ist bei Verwendung der Variante mit fester Fensterbreite m aber sogar “on-line” möglich.

Berechnung von b_t (und A_t^{mIC}) für wenige t

Für den zeitinvarianten Fall gilt das beim rIC-Filter Angemerkte unverändert weiter.

Vergleich KK : mIC bei IO’s

Im Gegensatz zu der Trägheit, die man im allgemeinen dem klassischen Kalman-Filter beim Folgen von Strukturbrüchen nachsagt, konnten wir diese Eigenschaft bei bekannten Hyper-Parametern nur in Fällen mit sehr ungünstiger Signal-to-Noise-Ratio reproduzieren. Anders stellt sich die Frage dieser Trägheit allerdings, wenn man die Hyper-Parameter mitschätzt, und dann ein Strukturbruch durch Hyper-Parameter-Schätzer als ein “großes” V interpretiert wird — vgl. Künstler (1995), Simulationen 4 und 5. Dies ist aber in unserem Setup nicht vorgesehen.

4.8 Verfügbarkeit

ISP-Routinen zur Bestimmung des mIC-Glätters und -Filters sowie zu dessen Kalibrierung können bezogen werden unter

<http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/diss/ispmacros/mIC>

³für den Begriff “batch” siehe Kassam und Poor (1985), p.469, zweite Spalte

Kapitel 5

Mini-Studie

5.1 Motivation

Im folgenden wollen wir untersuchen, wie sich die bisher vorgestellten, rekursiven, robusten Filter-Varianten rIC und rLS im Vergleich mit den u.a. in Künstler (1995), bzw. Fahrmeir und Künstler (1999) vorgeschlagenen Posterior-Modus-Schätzern schlagen.

Dazu werten wir die Schätzer sowohl in einer idealen Situation¹ als auch in verschiedenen kontaminierten Situationen aus.

Die Posterior-Modus-Schätzer benötigen weiterhin die Angabe einer Eichverteilung \hat{P}^ε für die Fehler ε_t — mit möglicherweise höheren Flankenwahrscheinlichkeiten —, die nicht dem idealen Modell P^ε oder dem real vorliegenden Modell \hat{P}^ε entsprechen muß. Auch hier werden verschiedene Wahlen für \hat{P}^ε diskutiert. All diese Schätzer werden dann in den verschiedenen Situationen mithilfe des MSE-Kriteriums verglichen.

5.2 Notation und Definitionen

5.2.1 Modell

Steady-State Modell

Um die Berechnungen in einer vernünftigen Genauigkeit und Zeit durchführen zu können, beschränken wir uns auf das Steady-State Modell aus Beispiel 1.4.1. Mit den in Kapitel 1 eingeführten Bezeichnungen setzen wir

$$\begin{aligned} p, q &= 1 \\ a_0 &= 0, \quad Q_0 = 0, \quad Q_t = Q = 1, \quad V_t = V = 1 \\ F_t &= F = 1, \quad Z_t = Z = 1 \end{aligned} \tag{5.1}$$

Beschränkung auf $t = 1$

Wir wollen den bedingten Erwartungswert und den Modus der Posteriorverteilung als (Mit-)Konkurrenten für die Schätzung des Zustandes β_t auf Basis der

¹ohne Kontamination, die Daten stammen aus den entsprechend unterstellten Verteilungen

Beobachtung y_s $s \leq t$ betrachten, wobei ε_t nicht notwendig normalverteilt ist. Hierfür sind numerische Verfahren notwendig. Um nun den bedingten Erwartungswert und den Posterior-Modus auf einem hinreichend dichten Gitter von y_t -Werten auswerten zu können, beschränken wir uns daher auf den Fall $t = 1$, denn hier müssen wir nur die gemeinsame Verteilung von (β_1, y_1) betrachten, was auf Auswertungen längs eines (nur) 2-dimensionalen Gitters hinausläuft, das man relativ dicht wählen kann (c.f. Abschnitt 5.4), ohne dabei die Effizienz und Geschwindigkeit durch (langsame) Zwischenspeicherung [auf der Festplatte] zu beeinträchtigen.

5.2.2 Verteilungen

In den bisherigen Ansätzen zur Robustifizierung im Filter-Problem wird häufig die Strategie verfolgt, die Normalitätsannahme in dem Sinn zu verallgemeinern, daß man diese durch vorgegebene, fettschwänzigere Verteilungen ersetzt, wie Konvexkombinationen von Normalverteilungen oder t -Verteilungen mit geringer Zahl an Freiheitsgraden. Dies aber verschiebt das Problem nur auf die Wahl neuer Idealverteilungen.

Die hier anvisierte Robustheit bleibt dagegen im idealen Modell bei der Normalverteilungsannahme, versucht aber, den Einfluß möglicherweise erratischer Beobachtungen zu begrenzen, und dies gleichmäßig über alle Verteilungen, aus denen die Beobachtung in der realen Situation stammen könnte.

Typen von Ausreißern

In diesem Abschnitt wollen wir nun die Performance unserer Verfahren bei den in Abschnitt 1.4.1 eingeführten AO's untersuchen, i.e. wir lassen Abweichungen von der idealen Verteilung der ε_1 zu; v_1 , hingegen unterliegt keiner Kontamination und bleibt in jedem Fall unabhängig von ε_1 .

Verteilungen in der idealen Situation

$$\varepsilon_1, v_1 \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}(0, 1) \quad (5.2)$$

Konvexkombinationen

Als eine Modellabweichung in der Verteilung von ε_1 betrachten wir Konvexkombinationen — andernorts auch "gross error models", vgl. Huber (1964) — der idealen Verteilung mit einer kontaminierenden Verteilung. Dabei setzen wir für die kontaminierende Verteilung wieder eine Normalverteilung mit folgenden Parametern an:

$$P^\varepsilon = (1 - r)\mathcal{N}(0, 1) + r\mathcal{N}(a, s^2) \quad (5.3)$$

mit $r \in (0, 1)$, $a \in \mathbb{R}$, $s \in \mathbb{R}_{\geq 0}$. Genauer wählen wir²

$$\begin{aligned} [\text{cv1}] & 0.9\mathcal{N}(0, 1) + 0.1\mathcal{N}(4, 1) \\ [\text{cv2}] & 0.9\mathcal{N}(0, 1) + 0.1\mathcal{N}(0, 9) \\ [\text{cv3}] & 0.8\mathcal{N}(0, 1) + 0.2\mathcal{N}(0, 9) \end{aligned}$$

²[cv2] entspricht einem Vorschlag von Tukey (1960).

t -Verteilungen

Daneben betrachten wir auch die folgenden t -Verteilungen als Verteilung von ε_1 :

[t1]	t_1 (1 Freiheitsgrad)
[t3]	t_3 (3 Freiheitsgrade)

5.3 Schätzer

5.3.1 der klassische Kalman–Filter in diesem Zusammenhang

Im Steady–State Modell 5.2.1 identifizieren wir:

$$\begin{aligned}
 \beta_{0|0}^{\text{KK}} &= a_0 = 0 & \Sigma_{0|0}^{\text{KK}} &= 0 \\
 \beta_{1|0}^{\text{KK}} &= 0 & \Sigma_{1|0}^{\text{KK}} &= 0 + 1 = 1 \\
 \beta_{1|1}^{\text{KK}} &= \hat{M}_1^{\text{KK}} \Delta^{\text{KK}} y_1 & \Sigma_{1|1}^{\text{KK}} &= 1 - 1/2 = 1/2 \\
 \Delta y_1^{\text{KK}} &= y_1 - 0 = y_1 & M_1^{\text{KK}} &= 1\{1 + 1\}^{-1} = 1/2
 \end{aligned} \tag{5.4}$$

Also ist der klassische Kalman–Filter in unserem Kontext $f^{\text{KK}}(y_1) = y_1/2$. Er minimiert den MSE sogar unter allen y_1 -meßbaren Schätzern in der idealen Situation, weil wegen der Normalität der bedingte Erwartungswert $E[\beta_1|y_1]$ linear in y_1 ist.

5.3.2 geeichter Posterior–Modus

Definition

Wenn wir eine gewisse, absolutstetige Verteilung \check{P}^ε mit differenzierbarer Lebesgue–Dichte \check{p}^ε für ε_1 annehmen, können wir als Schätzer für β_1 auch den Modus der Posteriorverteilung, zulassen³, i.e.

$$f_{\check{P}^\varepsilon}^{\text{PM}}(y_1) := \{\beta : \check{p}^{\beta_1|Y_1=y_1}(\beta) = \max_{\beta}!\}$$
 \tag{5.5}

mit $\check{p}^{\beta_1|Y_1=y_1}(\beta)$ der faktorisierten, bedingten Dichte von $\mathcal{L}(\beta|Y_1 = y_1)$. Weil die Maximalstelle nicht vom normierenden Integral $\int \check{p}^{\beta_1, Y_1}(\beta, y_1) \lambda(d\beta)$ abhängt, wobei wir $\check{p}^{\beta, y}$ für die unterstellte λ^2 -Dichte von (β_1, Y_1) schreiben, ist dies äquivalent zu

$$f_{\check{P}^\varepsilon}^{\text{PM}}(y_1) = \{\beta : \check{p}^{\beta_1, Y_1}(\beta, y_1) = \max_{\beta}!\}$$
 \tag{5.6}

Benutzen wir die Unabhängigkeit von v_1 und ε_1 und ihre Lebesguedichten, so erhalten wir

$$\check{p}^{\beta_1, Y_1}(\beta, y) = \check{p}^\varepsilon(y - \beta) p^{v_1}(\beta)$$
 \tag{5.7}

mit $p_{v_1} = \varphi$ der Dichte von $\mathcal{N}(0, 1)$. Weil hier \check{p}^ε , p_{v_1} differenzierbar sind, wird ein Maximum in (6) nur dann angenommen, falls

$$\frac{\partial}{\partial \beta} \check{p}^{\beta_1, Y_1}(\beta, y_1) = 0$$
 \tag{5.8}

³Dies entspricht einer Eichung des Posterior–Modus an dieser Verteilung.

Lösung im Fall $\varepsilon_1 \sim t_\nu$

Für $\varepsilon_1 \sim t_\nu$ gilt

$$\frac{\partial}{\partial \beta} \check{p}^{\beta_1, Y_1}(\beta, y_1) = -\check{p}^{\beta_1, Y_1}(\beta, y_1) \left[\beta + (\nu + 1) \frac{\beta - y_1}{\nu + (\beta - y_1)^2} \right]. \quad (5.9)$$

Daher ist $f_{\check{P}^\varepsilon}^{\text{PM}}(y_1)$ die Nullstelle von

$$Q(\beta, y_1) := \beta^3 - 2y_1\beta^2 + (2\nu + 1 + y_1^2)\beta - (\nu + 1)y_1 \quad (5.10)$$

Ersetzen wir $z \rightsquigarrow \beta - 2y_1/3$, so erhalten wir das kubische Polynom in Normalform $\tilde{Q}(z) = z^3 + pz + q$ mit $p = 2\nu + 1 - y_1^2/3$ und $q = 2/27y_1^3 + (\nu - 1)/3y_1$. Durch Einsetzen der Terme in die Diskriminante $D = q^2/4 + p^3/27$ erhalten wir, daß $D > 0$ für jedes $\nu \in \mathbb{N}$. Daher existiert nach Cardano nur eine reelle Nullstelle, und diese muß dann auch ein Maximum von (8) sein, denn als Dichte ist $\check{p}^{\beta_1, Y_1}(\beta, y_1) \geq 0$ und $\check{p}^{\beta_1, Y_1}(\beta, y_1) \rightarrow 0$ für $|\beta| \rightarrow \infty$. Mit der Cardanoschen Formel kann man diese Nullstelle gewinnen als $f_{\check{P}^\varepsilon}^{\text{PM}}(y_1) = 2y_1/3 + v_1 + v_2$ mit $u_{1,2} = -q/2 \pm \sqrt{D}$ und $v_{1,2} = \text{sign}(u_{1,2}) \sqrt[3]{|u_{1,2}|}$. Das Resultat scheint jedoch numerisch nicht stabil zu sein, so daß wir für nicht allzu große Werte von $|y|$ stattdessen die Methoden in Abschnitt 5.4 verwenden, um $f_{\check{P}^\varepsilon}^{\text{PM}}$ zu bestimmen.

Lösung im Fall $\varepsilon_1 \sim (1 - r)\mathcal{N}(0, 1) + r\mathcal{N}(a, s^2)$

Für $\varepsilon_1 \sim \check{P}^\varepsilon = (1 - r)\mathcal{N}(0, 1) + r\mathcal{N}(a, s^2)$, definieren wir

$$R(\beta, y_1) := \frac{(1 - r)(\beta - y_1) \exp(-\frac{(\beta - y_1)^2}{2}) + \frac{r}{s} \frac{(\beta - y_1 + a)}{s^2} \exp(-\frac{(\beta - y_1 + a)^2}{2s^2})}{(1 - r) \exp(-\frac{(\beta - y_1)^2}{2}) + \frac{r}{s} \exp(-\frac{(\beta - y_1 + a)^2}{2s^2})} \quad (5.11)$$

Dann ist

$$\frac{\partial}{\partial \beta} \check{p}^{\beta_1, Y_1}(\beta, y_1) = -\check{p}^{\beta_1, Y_1}(\beta, y_1) [\beta + R(\beta, y_1)] \quad (5.12)$$

Leider sind die Nullstellen von (12) nur numerisch zu finden, so daß wir diese iterativ auf einem Gitter von y_1 -Werten berechnen; weiterhin muß man die Entscheidung treffen, welches der lokalen Maxima das globale ist, da die Dichten nicht mehr unimodal zu sein brauchen. In jedem Fall kann diese Berechnung nicht parallel in Vektorschreibweise geschehen. Daher benützen wir für nicht zu große Werte von $|y|$ stattdessen die in Abschnitt 5.4 beschriebenen Maximierungs- und Interpolationsmethoden zur Berechnung von $f_{\check{P}^\varepsilon}^{\text{PM}}$.

Asymptotik für $|y_1| \rightarrow \infty$

Für große Werte von $|y|$ sind die Interpolationsmethoden in Abschnitt 5.4 ungeeignet. In diesem Fall sind analytische Ausdrücke vorzuziehen, um einen zu großen Interpolationsbereich zu vermeiden. Daher diskutieren wir die Asymptotik der geeichten Posterior-Modi etwas ausführlicher.

$\varepsilon_1 \sim t_\nu$

Um eine Asymptotik für $|y_1| \rightarrow \infty$ zu bekommen, schreiben wir D explizit auf als

$$D = \frac{\sqrt{3\nu}}{9} y_1^2 \sqrt{1 - \frac{13\nu^2 + 22\nu + 1}{4\nu} y_1^{-2} + \frac{(2\nu + 1)^3}{\nu} y_1^{-4}} = \frac{\sqrt{3\nu}}{9} y_1^2 (1 + o(y_1^{-1})) \quad (5.13)$$

Daher erhalten wir

$$\begin{aligned} u_{1,2} &= -\frac{1}{27}y_1^3 - \frac{\nu-1}{6}y_1 \pm \frac{\sqrt{3\nu}}{9}y_1^2(1 + o(y_1^{-1})) \\ v_{1,2} &= -\frac{1}{3}y_1 \sqrt[3]{1 + \frac{9(\nu-1)}{2}y_1^{-2} \mp 3\sqrt{3\nu}y_1^{-1}(1 + o(y_1^{-1}))} = \\ &= -\frac{1}{3}y_1 + \frac{3\nu-1}{2}y_1^{-1} \mp \frac{1}{3}\sqrt{3\nu} + o(y_1^{-1}) \\ f_{\tilde{P}^\varepsilon}^{\text{PM}}(y_1) &= (3\nu-1)y_1^{-1} + o(y_1^{-1}) \end{aligned}$$

$$\varepsilon_1 \sim (1-r)\mathcal{N}(\mathbf{0}, 1) + r\mathcal{N}(\mathbf{a}, s^2)$$

Weil die Exponentialfunktion für $y \rightarrow \infty$ jedes Polynom dominiert, entscheiden die $\exp(\cdot)$ -Ausdrücke über das asymptotische Verhalten. Unter den zwei sich mischenden $\exp(\cdot)$ -Dichten dominiert schließlich diejenige mit dem betragsmäßig kleineren Argument die andere. Daher gibt es die beiden Fälle

$$\begin{aligned} (\beta - y_1)^2 &< \frac{(\beta - y_1 + a)^2}{s^2} \\ R(\beta, y_1) &= \frac{(\beta - y_1) + \frac{r}{s(1-r)} \frac{(\beta - y_1 + a)}{s^2} \exp\left(\frac{(\beta - y_1)^2 - \frac{(\beta - y_1 + a)^2}{s^2}}{2}\right)}{1 + \frac{r}{s(1-r)} \exp\left(\frac{(\beta - y_1)^2 - \frac{(\beta - y_1 + a)^2}{s^2}}{2}\right)} = \\ &= \frac{(\beta - y_1) + o(y_1^0)}{1 + o(y_1^0)} = (\beta - y_1) + o(y_1^0) \end{aligned}$$

Darum müssen wir $0 \stackrel{!}{=} \beta + (\beta - y_1) + o(y_1^0)$ lösen, was asymptotisch $f_{\tilde{P}^\varepsilon}^{\text{PM}}(y_1) = y_1/2$ ergibt.

$$\begin{aligned} (\beta - y_1)^2 &> \frac{(\beta - y_1 + a)^2}{s^2} \\ \text{Wir schließen analog, daß } R(\beta, y_1) &= \frac{(\beta - y_1 + a)}{s^2} + o(y_1^0), \text{ so daß asymptotisch, } f_{\tilde{P}^\varepsilon}^{\text{PM}}(y_1) = \frac{y_1 - a}{s^2 + 1}. \end{aligned}$$

5.3.3 geeichter bedingter Erwartungswert

Definition

Es ist bekannt, daß der optimale Schätzer bezogen auf das MSE-Kriterium in unserem Kontext der bedingte Erwartungswert $E[\beta_1|y_1]$ ist. Um diesen auswerten zu können, müssen wir wie beim Posterior-Modus die Verteilung \tilde{P}^ε von ε_1 festlegen (eichen). Wie dort nehmen wir an, \tilde{P}^ε besitze eine Lebesgue-Dichte \tilde{p}^ε . Mit dieser Verteilung indizieren wir den Erwartungswert und erhalten

$$f_{\tilde{P}^\varepsilon}^{\text{bE}} := E_{\tilde{P}^\varepsilon}[\beta_1|Y_1 = y_1] = \frac{\int \beta \tilde{p}^\varepsilon(y - \beta) p^{v_1}(\beta) d\beta}{\int \tilde{p}^\varepsilon(y - \beta) p^{v_1}(\beta) d\beta} \quad (5.14)$$

Lösung in den Fällen [t1], [t3]

Für die t -Verteilungen als Eichverteilung konnten wir keine analytischen Ausdrücke ermitteln; stattdessen verwenden wir die numerischen Extrapolationsverfahren aus Abschnitt 5.4.

Lösung in den Fällen [cv1], [cv2] und [cv3]

In diesen Fällen kann man die Integration explizit durchführen und erhält

$$f_{\hat{P}^\varepsilon}^{\text{bE}} = \frac{\frac{1-r}{\sqrt{2}} \frac{y_1}{2} \exp(-\frac{y_1^2}{2}) + \frac{r}{\sqrt{1+s^2}} \frac{y_1-a}{s^2+1} \exp(-\frac{(y_1-a)^2}{2(s^2+1)})}{\frac{1-r}{\sqrt{2}} \exp(-\frac{y_1^2}{2}) + \frac{r}{\sqrt{1+s^2}} \exp(-\frac{(y_1-a)^2}{2(s^2+1)})} \quad (5.15)$$

Asymptotik für $|y_1| \rightarrow \infty$

$$\varepsilon_1 \sim t_\nu$$

Hier haben wir keine analytischen Ausdrücke für das asymptotische Verhalten, aber unsere graphischen Resultate in Abschnitt 5.5.2 zusammen mit den Extrapolationstechniken, die in Abschnitt 5.4 diskutiert werden, legen es nahe, daß die Asymptotik dieselbe ist wie im Fall des Posterior Modus.

$$\varepsilon_1 \sim (1-r)\mathcal{N}(0,1) + r\mathcal{N}(a,s^2)$$

Wir können wie beim Posterior-Modus argumentieren, und erhalten asymptotisch, daß $f_{\hat{P}^\varepsilon}^{\text{bE}} \rightarrow \frac{y_1-a}{s^2+1}$ oder $f_{\hat{P}^\varepsilon}^{\text{bE}} \rightarrow \frac{y_1}{2}$, je nachdem, ob $(\beta - y_1)^2$ größer ist als $\frac{(\beta - y_1 + a)^2}{s^2}$ oder nicht.

5.3.4 rLS**Konkrete Gestalt**

Wir erinnern daran, daß wir in Proposition 2.3.1 nachgewiesen haben, daß man in unserem Kontext, also für $t = 1$ und normales (v_1, ε_1) , M^{rLS} explizit als $M^{\text{rLS}} = M^{\text{KK}} = 1/2$ erhält.

Um b zu bestimmen, haben wir das Verfahren auf einen relativen Effizienzverlust gegenüber dem optimalen Verfahren f^{KK} von $\delta = 10\%$ kalibriert, indem wir ein Bisektionsverfahren verwendet haben. Dabei haben wir für (5) Monte-Carlo-Integration bei einem Stichprobenumfang von 50000 benutzt und einen Wert von $b = 0.82812500$ erhalten, so daß unser rLS-Schätzer lautet

$$f^{\text{rLS}}(y_1) = \min(0.82812500, \max(y_1/2, -0.82812500)) \quad (5.16)$$

SO-Optimalität

In Kapitel 8 werden wir sehen, daß

$$\hat{f}^{\text{SO}}(y_1) = \min\{b, \max\{E_{P^\varepsilon}[\beta_1|Y_1], -b\}\}$$

für ein Modell mit idealen Verteilungen P^ε und P^β von ε_1 und β_1 minimax-optimal⁴ in einer Umgebung von dort noch zu spezifizierenden SO-Ausreißern⁵ ist, sowie optimal für das Problem, den MSE im idealen Modell unter einer Schranke an den Bias⁶ zu minimieren, ebenfalls in einer SO-Umgebung.

Weil aber für $P^\varepsilon, P^\beta = \mathcal{N}(0,1)$ gerade $E_{P^\varepsilon}[\beta_1|Y_1] = M^{\text{KK}} y_1 = y_1/2$ ist, ist also der rLS-Filter hier SO-optimal. Allerdings sind diese SO-Umgebungen — zumindest in unserem Kontext⁷ — “größer”, enthalten also zu gegebenem

⁴c.f. Abschnitt 8.3

⁵c.f. Abschnitt 8.1.3

⁶c.f. Abschnitt 8.2

⁷vgl. Bemerkung 8.1.2(c)

Radius mehr Verteilungen als die AO-Umgebungen. Deswegen ist nicht klar, ob der rLS-Filter auch in diesem Setup optimal ist.

5.3.5 rIC

Wir spezifizieren die Größen im Steady-State Modell:

Die Influenzkurve im total-geclippten Fall schreibt sich als

$$\psi = \min(-b, \max(A\lambda, b)). \quad (5.17)$$

Mit Φ und φ , Verteilungsfunktion und Lebesgue-Dichte von $\mathcal{N}(0, 1)$, $\sigma^2 := \Sigma_{1|1} = 0.5$ und $c := \frac{b\sigma}{A}$, können wir (17) und (19) in unserem Fall explizit ausrechnen als Nullstellen von

$$g_1(A, b) := A(2\Phi(c) - 1) - \sigma^2 \stackrel{!}{=} 0 \quad (5.18)$$

$$g_2(A, b) = [2\Phi(c) - 1 - 2c\varphi(c)]\frac{A^2}{\sigma^2} - \frac{1 + \delta}{\sigma^2} + 2b^2(1 - \Phi(c)) \stackrel{!}{=} 0 \quad (5.19)$$

Daher können wir die Konstanten A und b mit einem gemeinsamen Newton-Verfahren für $f = (g_1, g_2)^T$ ermitteln. Die dafür nötigen Ableitungen lauten

$$\frac{\partial}{\partial A} g_1(A, b) = 2\Phi(c) - 1 - 2c\varphi(c), \quad (5.20)$$

$$\frac{\partial}{\partial b} g_1(A, b) = 2\frac{\sigma}{A}\varphi(c), \quad (5.21)$$

$$\frac{\partial}{\partial A} g_2(A, b) = 2\frac{A}{\sigma^2}(2\Phi(c) - 1) - 4\frac{b}{\sigma}\varphi(c), \quad (5.22)$$

$$\frac{\partial}{\partial b} g_2(A, b) = 4b(1 - \Phi(c)). \quad (5.23)$$

Konkrete Gestalt

Für die Zwecke dieser Untersuchung setzen wir den ‘relativen Effizienzverlust’ gemäß Anscombe-Kriterium (19) $\delta = 0.1$ und erhalten

$$A = 0.71452731 \quad b = 1.0467970, \quad (5.24)$$

so daß unser rIC-Filter von folgender Gestalt ist:

$$f^{\text{rIC}}(y_1) = \min(1.0467970, \max(0.71452731y_1, -1.0467970)). \quad (5.25)$$

5.4 Berechnungen

5.4.1 numerische Verfahren

Posterior-Modus: Maximierung

Im Fall des Posterior-Modus’ mit Eichverteilung [cv2] oder [t1], werten wir die gemeinsame Dichte von β_1 und y_1 , $\check{p}^{\beta_1, Y_1}(\beta_1, y_1)$, auf einem achsenparallelen Gitter von 350 äquidistanten Punkten in y_1 -Richtung zwischen -20 und 20 und 500 äquidistanten Punkten in β_1 -Richtung zwischen -7 und 7 aus. In der resultierenden 350×500 Matrix berechnen wir dann durch zeilenweise

Maximierung, eine (numerische) Schätzung für den Posterior-Modus auf einem äquidistanten Gitter von 350 Werten von y_1 . Weil wir von theoretischer Seite her wissen⁸, daß der Posterior-Modus in unseren Beispielen beliebig glatt ist, glätten wir die Ergebnisse noch mithilfe von Spline-Glättung — siehe auch weiter unten.

bedingter Erwartungswert: Integration

Um den bedingten Erwartungswert zu ermitteln, werten wir die beiden Integrale $I_1(y_1) := \int \beta_1 p(\beta_1, y_1) d\beta_1$ und $I_2(y_1) := \int p(\beta_1, y_1) d\beta_1$ numerisch längs eines äquidistanten Gitters von 800 Punkten für y_1 zwischen -20 und 20 aus, indem wir simultan Rombergverfahren jeweils für β_1 zwischen -7 und 7 durchführen, wobei die Integranden jeweils auf einem äquidistanten Gitter von 128 Punkten ausgewertet werden. Die Verwendung der Rombergintegration ist wegen der hochgradigen Glattheit der Integranden gerechtfertigt. Dies wiederum zieht eine exzellente Interpolierbarkeit durch Polynome nach sich, die Voraussetzung für ein gutes Resultat bei der Rombergintegration.

Im Fall der Konvexkombinationen als Eichverteilung für ϵ_1 können wir die numerische Integration umgehen, da I_1 und I_2 analytisch mit Formeln (15) berechnet werden können. Dies liefert aber auch einen cross-check mit der numerischen Integration, in dem Differenzen im Bereich $1 \text{ E} - 7$ aufgedeckt werden.

Interpolation/Glättung

Ergebnisse für den Posterior-Modus / bedingten Erwartungswert für Zwischenstellen von y_1 erreichen wir durch kubische Spline-Interpolation, die bereits in Form von Macros in ISP zur Verfügung steht. In Anbetracht möglicher numerischer Fehler aufgrund von Rundung oder Integration lassen wir zu, daß die Splines für den Posterior-Modus / bedingten Erwartungswert nicht exakt durch die Gitterpunkte gehen, indem wir einen Glättungsparameter von 0.005 für den Parameter `smpar` im ISP-Macro `smspline` verwenden. Dadurch liegt der MSE der Spline-Interpolation bei $1/200$ des MSE der besten linearen Anpassung nach dem Kleinste-Quadrate-Kriterium.

Extrapolation

Außerhalb des y_1 -Gitters ist Spline-Interpolation ungeeignet, da die führenden Koeffizienten der kubischen Polynome der Rand-Abschnitte kaum 0 sein werden, so daß wir ein asymptotisches Verhalten wie ay_1^3 für betragsmäßig große Werte von y_1 erhalten. Daher benutzen wir im Fall von Konvexkombinationen als Eichverteilung die Asymptotik, wie sie in Abschnitten 5.3.2 und 5.3.3 hergeleitet worden ist, und sehen, daß $E[\beta_1|y_1]$ asymptotisch gleich $(y_1 - a)/\sqrt{s^2 + 1}$ ist. Diese Beziehung verwenden wir für $|y| > 20$. Bei einer Eichung an t -Verteilungen erhält man eine exzellente Näherung des $f(y_1)$ -Gitters für $|y| > 4$ durch Funktionen ay^γ , $\gamma \leq -1$, wobei f der Posterior-Modus oder der bedingte Erwartungswert ist. Genauer haben wir ein lineares Regressionsmodell für $\log f = \log a + \gamma \log y_1 + \delta$ angepasst. Im Fall [t1] lieferten $a = 2$ und $\gamma = -1$ einen MSE von 28.2 und ein r^2 von 0.993 für den Posterior-Modus

⁸im wesentlichen der Satz über implizite Funktionen

und einen MSE von 6.02 und ein r^2 of 0.993 für den bedingten Erwartungswert, beides jeweils auf 860 Beobachtungen ausgewertet. Im Fall [t3] lieferten $a = \sqrt{3}$, $\gamma = -1.2$ einen MSE von 15.1 und ein r^2 von 0.977 für den bedingten Erwartungswert.

5.4.2 geschichtete Monte–Carlo Simulation für den MSE

Indem wir Inter–/Extrapolationstechniken, wie im vorigen Abschnitt beschrieben, verwenden, sind wir in der Lage, den Posterior–Modus und den bedingten Erwartungswert geeicht an den diskutierten Verteilungen sehr schnell längs hochdimensionaler Vektoren von y_1 -Werten auszuwerten.

Wir arbeiten mit Monte–Carlo–Stichproben $(\beta_1, y_1)_i$ vom Umfang 20000, die wir gemäß den in Abschnitt 5.2.2 diskutierten Situationen simulieren. Um die Genauigkeit zu erhöhen, haben wir dies 200 mal wiederholt und längs dieser geschätzten MSE's gemittelt, wodurch wir die Genauigkeit eines Stichprobenumfangs von 4 E6 erreichen.

Einen cross–check liefert der MSE des klassischen Kalman–Filters in der idealen Situation, von dem wir den theoretischen Wert von 0.5 kennen, was einen Simulationsfehler der Größenordnung $1 \text{ E} - 3$ in diesem Fall liefert.

5.5 Ergebnisse

Im folgenden stellen wir die Ergebnisse dieser Mini–Studie in Form von Tabellen und Grafiken dar.

5.5.1 Tabellen

Tabelle 5.1: MSE bei der Schätzung von β_1 wissend y_1

Schätzer	klass.	post–M–	post–M–	rIC	rLS	$E_{P^\varepsilon}[\beta_1 y_1]$
Situation: P^ε	Kalm.–F.	[cv1]	[t1]			(optimal)
ideal	0.4992 E0	0.5091 E0	0.5700 E0	0.5498 E0	0.5494 E0	0.4992 E0
conv: [cv1]	0.8972 E0	0.7107 E0	0.6711 E0	0.6565 E0	0.6953 E0	0.6449 E0
[cv2]	0.7043 E0	0.6025 E0	0.6303 E0	0.6069 E0	0.6305 E0	0.5923 E0
[cv3]	0.8973 E0	0.6906 E0	0.6865 E0	0.6606 E0	0.7041 E0	0.6548 E0
t-Vert. [t1]	0.1407 E6	0.5629 E4	0.7964 E0	0.8334 E0	0.9386 E0	0.7710 E0
[t3]	0.9974 E0	0.6848 E0	0.6780 E0	0.6513 E0	0.6919 E0	0.6457 E0

In Tabelle 5.1 sehen wir, daß der MSE des klassischen Kalman–Filters ausgewertet im idealen Modell — theoretisch 0.5, simulativ 0.4992 — optimal unter allen betrachteten Schätzern über alle Situationen hinweg ist; das Verhalten des Kalman–Filters wird jedoch indiskutabel unter den anderen nicht–idealen Situationen, insbesondere unter der Cauchy–Verteilung [t1].

An dieser Situation “scheitert” auch der mit [cv1] geeichte Posterior–Modus. Alle anderen Verfahren sind über alle betrachteten Situationen robust in dem Sinn, daß sie hinsichtlich MSE in allen Situationen “vernünftige” Werte liefern. Der an der Cauchy–Verteilung geeichte Posterior–Modus allerdings verschenkt

Tabelle 5.2: $\text{MSE}_{P^\varepsilon}(\mathbf{E}_{\tilde{P}^\varepsilon}[\beta_1 | \mathbf{y}_1])$

Eichverteilung \tilde{P}^ε		[ideal]	[cv1]	[cv2]	[cv3]	[t1]	[t3]
Situation: P^ε							
ideal		0.4992 E0	0.5604 E0	0.5175 E0	0.5384 E0	0.5934 E0	0.5504 E0
conv:	[cv1]	0.8972 E0	0.6449 E0	0.6723 E0	0.6563 E0	0.6758 E0	0.7587 E0
	[cv2]	0.7043 E0	0.7078 E0	0.5923 E0	0.5982 E0	0.6352 E0	0.6059 E0
	[cv3]	0.8973 E0	0.8473 E0	0.6625 E0	0.6548 E0	0.6764 E0	0.6605 E0
t-Vert.	[t1]	0.1407 E6	0.3104 E1	0.1015 E1	0.9817 E0	0.7709 E0	0.7931 E0
	[t3]	0.9974 E0	0.8991 E0	0.6556 E0	0.6503 E0	0.6677 E0	0.6457 E0

Tabelle 5.3: rel. Abw. zu $\text{MSE}_{P^\varepsilon}(\mathbf{E}_{P^\varepsilon}[\beta_1 | \mathbf{y}_1])$ in Tabelle 5.1

Schätzer	klass.	post-M-	post-M-	rIC	rLS
Situation: P_0^ε	Kalm.-F.	[cv2]	[t1]		
ideal	0.00 %	1.98 %	14.18 %	10.14 %	10.05 %
conv:	[cv1] 39.12 %	10.22 %	4.07 %	1.79 %	7.81 %
	[cv2] 18.91 %	1.71 %	6.40 %	2.47 %	6.45 %
	[cv3] 37.04 %	5.47 %	4.85 %	0.89 %	7.53 %
t-Vert.	[t1] 1.83 E7 %	7.30 E5 %	3.31 %	8.11 %	21.76 %
	[t3] 54.46 %	6.06 %	5.01 %	0.84 %	7.15 %

Tabelle 5.4: rel. Abw. $\text{MSE}_{P^\varepsilon}(\mathbf{E}_{\tilde{P}^\varepsilon}[\beta_1 | \mathbf{y}_1]) : \text{MSE}_{P^\varepsilon}(\mathbf{E}_{P^\varepsilon}[\beta_1 | \mathbf{y}_1])$

Eichverteilung \tilde{P}^ε		[ideal]	[cv1]	[cv2]	[cv3]	[t1]	[t3]
Situation: P^ε							
ideal		0.00 %	12.25 %	3.66 %	7.84 %	18.86 %	10.25 %
conv:	[cv1]	39.12 %	0.00 %	4.25 %	1.77 %	4.79 %	17.65 %
	[cv2]	18.91 %	19.49 %	0.00 %	0.99 %	7.24 %	2.29 %
	[cv3]	37.04 %	29.39 %	1.18 %	0.00 %	3.30 %	0.88 %
T-distr.	[t1]	1.83 E7 %	3.03 E2 %	31.61 %	27.35 %	0.00 %	2.88 %
	[t3]	54.46 %	39.24 %	1.54 %	0.71 %	3.41 %	0.00 %

— mit Ausnahme seiner eigenen Eichsituation — gegenüber dem rIC-Filter stets und gegenüber dem rLS in der idealen Situation an Effizienz, wie auch noch in Tabelle 5.3 explizit dargestellt ist.

Eine Erklärung für dieses (relativ) schlechte Abschneiden des rLS-Filters ist, daß er auf einen relativen Effizienzverlust von $\delta = 10 \%$, gemessen in der Verschlechterung des MSE, und nicht auf einen (SO-)Radius von $r_{\text{SO}} = 10 \%$ kalibriert ist. Dies entspricht in dieser Situation einem (SO-)Radius von $r_{\text{SO}} = 7.1 \%$, und ein (SO-)Radius von 10% würde durch $\delta = 13.8 \%$ erreicht⁹.

Zur Kontrolle sind die in der Praxis unpraktikablen, geeichten bedingten Erwar-

⁹Diese Tatsache stand erst nach Erstellung der Studie zur Verfügung und wurde deshalb noch nicht verwertet.

tungswerte unter Einschluß von [t3], [cv1] und [cv3] in Tabelle 5.2 abgetragen, von denen wir aus der Theorie wissen, daß sie jeweils in ihrer Eichsituation optimal sind. Diese Tatsache spiegelt sich dann auch in Tabelle 5.4 wider.

Es zeigt sich ein ähnliches Bild wie bei den Posterior-Modi: Der bedingte Erwartungswert, geeicht an der idealen Verteilung, reagiert bereits sensibel bei moderaten Kontaminationen ([cv*], [t3]) und bricht bei [t1] völlig zusammen. Bei [t1] brechen die unbeschränkten, bedingten Erwartungswerte unter [cv*] ebenfalls zusammen, während [t3] hier noch vernünftige Werte liefert. Unter den [cv*]-geeichten Erwartungswerten zeichnet sich der [cv3]-geeichte als der noch robusteste aus.

Im Hinblick auf Effizienz im idealen Modell erscheint der [t3]-geeichte Erwartungswert dem [t1]-geeichten überlegen, dies kehrt sich jedoch in Situation [cv1] beträchtlich zu ungunsten des [t3]-geeichten um.

Über alle betrachteten Situationen hinweg erweisen sich rLS- und rIC-Filter den an [cv*] geeichten Posterior-Modi als überlegen, dem an [t1] geeichten Posterior-Modus zumindest ebenbürtig. Zieht man nun auch die Tatsache heran, daß nach einer "off-line" erfolgten Kalibrierung zur Festlegung von b bzw. A und b rLS- und rIC-Filter extrem einfach auszuwerten sind, muß man sie als den Posterior-Modi überlegen erachten.

5.5.2 graphische Darstellung

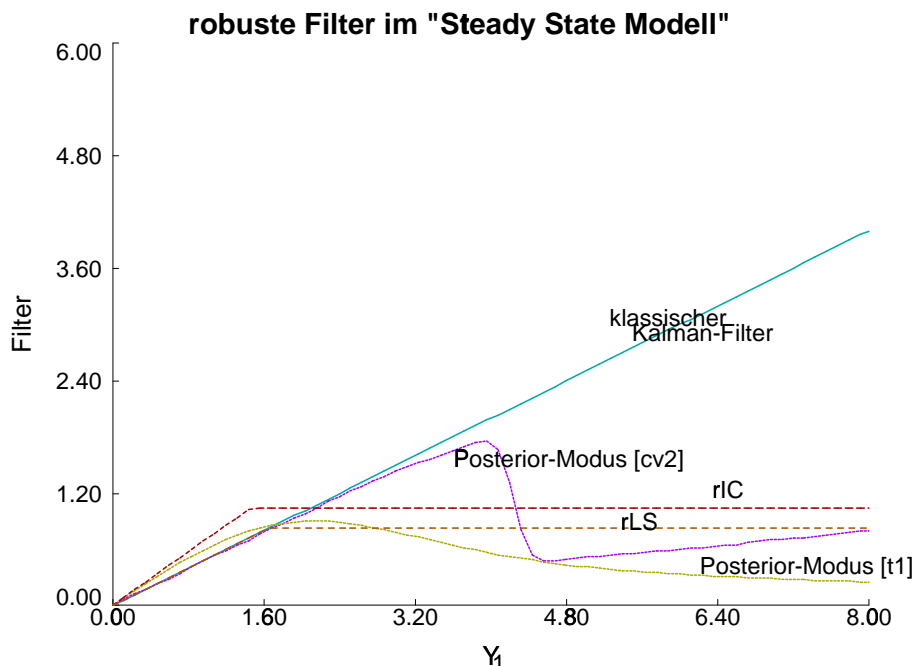


Abbildung 5.1: Schätzer für β_1 auf Basis der Beobachtung von y_1

In Abbildung 5.1 sind alle "realisierbaren" Schätzer vergrößert auf den interessanten Bereich dargestellt; wir beobachten, daß der klassische Kalman-Filter und der Posterior-Modus, geeicht an der Konvexkombination [cv1], unbeschränkt sind, während der Posterior-Modus, geeicht an einer Cauchy-Verteilung,

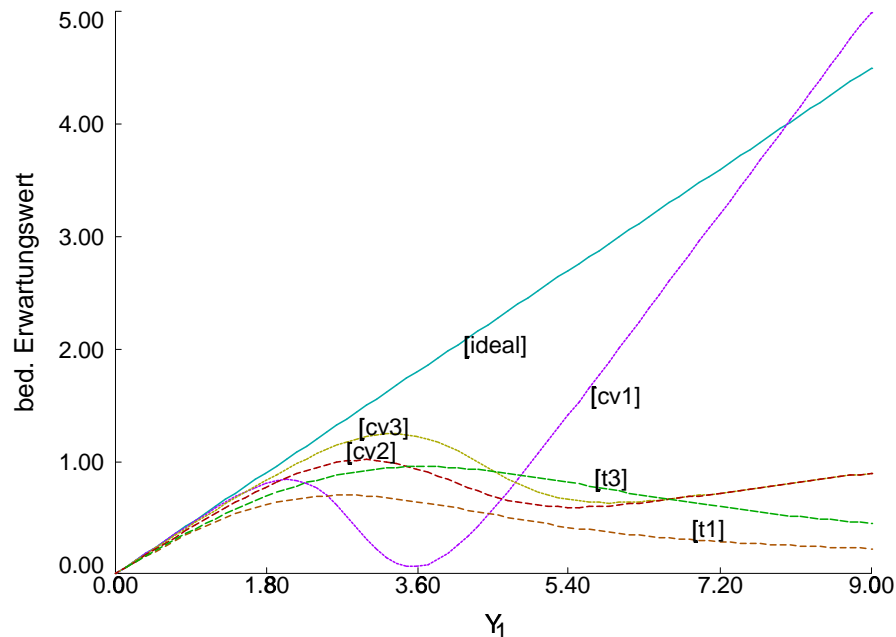


Abbildung 5.2: $E[\beta_1|y_1]$ unter verschiedenen Verteilungsannahmen für ϵ_1

sich einer Hyperbel nähert und rIC / rLS jeweils auf einer gewissen Höhe geclippt werden.

Zur Kontrolle bilden wir die theoretisch optimalen, geeichten bedingten Erwartungswerte, $E_{\tilde{P}^\epsilon}[\beta_1|y_1]$ für $\tilde{P}^\epsilon = [cv1], [cv2], [cv3], [t1]$ und $[t3]$, in Abbildung 5.2 ab. Wir erkennen, daß die Eichung an $[cv^*]$, wie theoretisch hergeleitet, für großes y_1 gegen eine lineare Funktion in y_1 tendiert. Bei der unsymmetrischen $[cv1]$ ergibt sich für $y_1 > E_{[cv1]}[\epsilon_1] = 4$ ein bedingter Erwartungswert $E_{[cv1]}[\beta_1|y_1]$, dessen Steigung sogar noch größer ist als von f^{KK} , während die symmetrischen Konvex-Kontaminationen mit einer größeren Varianz, ($[cv2]$, $[cv3]$) zu einer asymptotisch geringeren Steigung führen.

Bei den t -Verteilungen zieht ein kleinerer Freiheitsgrad ein schnelleres Abklingen gegen 0 nach sich.

Zum Vergleich von bedingtem Erwartungswert und Posterior-Modus haben wir in Abbildung 5.3 für die Eichverteilungen $[cv2]$ und $[t1]$ die Kurven in Abhängigkeit von y_1 geplottet; wir beobachten, daß für kleine Werte von y_1 in beiden Fällen der Posterior-Modus größere Schätzwerte für β_1 liefert als der entsprechende bedingte Erwartungswert, bei $[t1]$ sogar größere als f^{KK} . Im Fall $[cv2]$ beobachten wir um 4 herum ein abruptes Umschalten von “Beobachtung aus idealer Situation unterstellt” auf “Beobachtung aus kontaminierte Situation unterstellt”, wohingegen dieser Übergang beim bedingten Erwartungswert glatter ist.

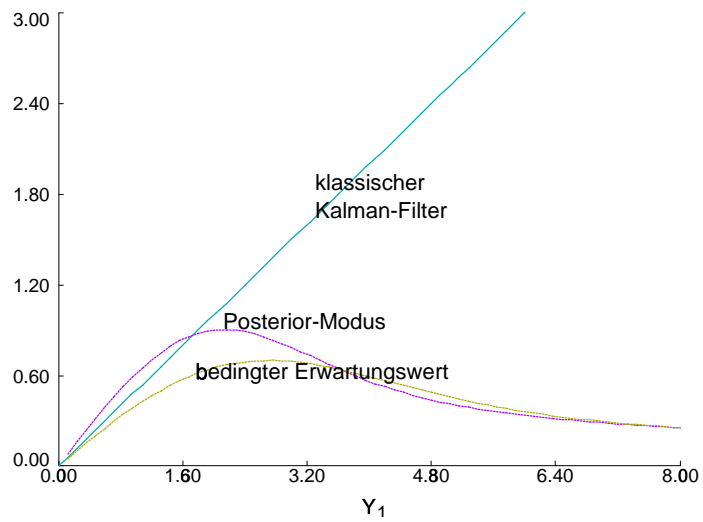
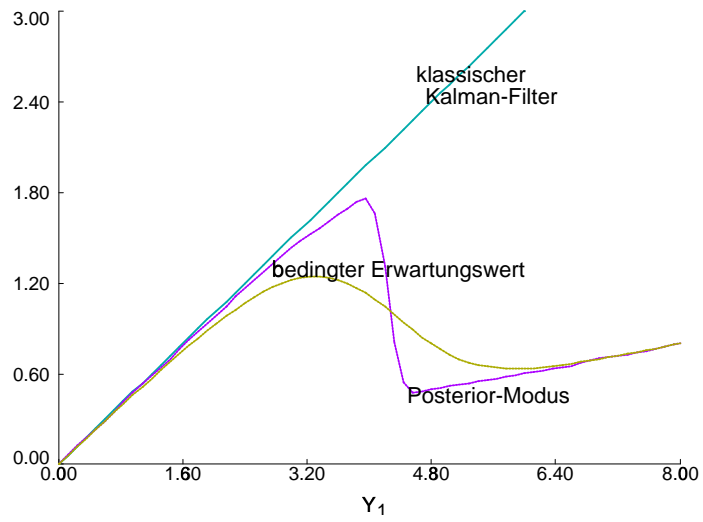


Abbildung 5.3: Posterior-Modus und bedingter Erwartungswert:
Darstellung von $E[\beta_1|y_1]$ und Posterior-Modus unter Annahme von $[cv2]$ (links) resp. $[t1]$ (rechts)

Kapitel 6

einige statistische Eigenschaften von rLS, rIC und mIC

Bei den bisher vorgestellten Verfahren, mIC, rLS und rIC, haben wir noch nicht die Frage ihrer Unverzerrtheit, ihrer (asymptotischen) Normalität und die Frage der Lebesgue-Stetigkeit ihrer Fehlerterme beantwortet. Dies wollen wir nun tun.

6.1 Symmetrie und Unverzerrtheit der robusten Verfahren

6.1.1 ein Hilfssatz zu symmetrischen Verteilungen

Lemma 6.1.1 *Seien $X \sim P^X$, $\varepsilon \sim P^\varepsilon$ stochastisch unabhängige \mathbb{R}^p - bzw. \mathbb{R}^q -wertige Zufallsgrößen, $\mathcal{L}(X) = \mathcal{L}(-X)$, $\mathcal{L}(\varepsilon) = \mathcal{L}(-\varepsilon)$, $Z \in \mathbb{R}^{q \times p}$ eine bekannte Matrix und $Y = ZX + \varepsilon$ und $\tilde{X} = f(Y) + X$ mit $f: \mathbb{R}^q \rightarrow \mathbb{R}^p$ Y -meßbar und $f(y) = -f(-y)$. Dann ist*

- (a) $\mathcal{L}(\tilde{X}) = \mathcal{L}(-\tilde{X})$
- (b) $\mathcal{L}(E[X|Y]) = \mathcal{L}(-E[X|Y])$

BEWEIS: (a): Für $A \in \mathbb{B}^p$ gilt

$$\begin{aligned} P(\tilde{X} \in A) &= P(f(X + \varepsilon) + X \in A) = P(f(-X - \varepsilon) - X \in A) = \\ &= P(-f(Y) - X \in A) = P(-\tilde{X} \in A) \end{aligned}$$

(b): Wir zeigen zunächst, daß für $A \in \mathbb{B}^q$ gilt

$$E[E[X|Y] \mathbf{I}_{\{Y \in A\}}] = -E[E[X|Y] \mathbf{I}_{\{Y \in -A\}}] \quad (6.1)$$

Sei dazu $A \in \mathbb{B}^q$ und $P^{ZX} = \mathcal{L}(ZX)$, $P^Y = P^{ZX} * P^\varepsilon$. Dann gilt

$$\begin{aligned} E[E[X|Y] \mathbf{I}_{\{Y \in A\}}] &= \int \mathbf{I}_{\{y \in A\}} E[X|Y = y] P^Y(dy) = \\ &= \int \mathbf{I}_{\{Zx + \varepsilon \in A\}} x P^\varepsilon(d\varepsilon) P^X(dx) = - \int \mathbf{I}_{\{-Zx - \varepsilon \in A\}} x P^\varepsilon(d\varepsilon) P^X(dx) = \end{aligned}$$

$$= - \int I_{\{y \in -A\}} E[X|Y=y] P^Y(dy) = - E[E[X|Y] I_{\{Y \in -A\}}].$$

Die Behauptung folgt dann mit der Tatsache, daß für \mathbb{R}^k -wertige Zufallsgrößen U mit $E|U| < \infty$ folgende Äquivalenz gilt:

$$\begin{aligned} & E[U I_{\{U \in A\}}] = - E[U I_{\{U \in -A\}}] \quad \forall A \in \mathbb{B}^P \\ \iff & P(U \in A) = P(U \in -A) \quad \forall A \in \mathbb{B}^P \end{aligned} \quad (6.2)$$

////

6.1.2 Folgerungen für rekursive Filter

Proposition 6.1.2 *Wenn im Zustandsraummodell für die treibenden Prozesse v_t , ε_t , sowie für den Anfangswert β_0 folgende Symmetrievoraussetzungen gelten,*

$$\mathcal{L}_{\text{id}}(\varepsilon) = \mathcal{L}_{\text{id}}(-\varepsilon), \quad \mathcal{L}_{\text{id}}(v) = \mathcal{L}_{\text{id}}(v), \quad \mathcal{L}_{\text{id}}(\beta_0 - E[\beta_0]) = \mathcal{L}_{\text{id}}(-(\beta_0 - E[\beta_0])), \quad (6.3)$$

so sind der rLS-, der rIC- und der mIC-Filter symmetrisch um den Erwartungswert verteilt, genauer gilt im rLS- und rIC-Fall

$$\mathcal{L}_{\text{id}}(\Delta\beta_t^{\text{rLS}/\text{rIC}}) = \mathcal{L}_{\text{id}}(-\Delta\beta_t^{\text{rLS}/\text{rIC}}) \quad (6.4)$$

und für den mIC-Fall

$$\mathcal{L}_{\text{id}}(\Delta\beta_{t:t-m}^{\text{mIC}}) = \mathcal{L}_{\text{id}}(-\Delta\beta_{t:t-m}^{\text{mIC}}). \quad (6.5)$$

Insbesondere sind alle Verfahren unverzerrt.

BEWEIS: Zunächst der rLS-Filter: Für $t = 1$ ist $\Delta\beta_1^{\text{rLS}} = \Delta\beta_1^{\text{KK}}$, und weil $\Delta\beta_1^{\text{KK}}$ linear in y_1 ist und $\mathcal{L}_{\text{id}}(Y) = \mathcal{L}_{\text{id}}(-Y)$, ist $\mathcal{L}_{\text{id}}(\Delta\beta_1^{\text{rLS}}) = \mathcal{L}_{\text{id}}(-\Delta\beta_1^{\text{rLS}})$. Wegen der Symmetrie von ε folgt nach Lemma 6.1.1 mit den Identifikationen $X \rightsquigarrow \Delta\beta_1^{\text{rLS}}$, $Y \rightsquigarrow \Delta y_1^{\text{rLS}}$, $f(y) = M^{\text{rLS}} y \min\{1, \frac{b}{|M^{\text{rLS}} y|}\}$, $\tilde{X} \rightsquigarrow \Delta\beta_{1|1}^{\text{rLS}}$, daß $\mathcal{L}_{\text{id}}(E[\Delta\beta_1^{\text{rLS}} | \Delta y_1^{\text{rLS}}]) = \mathcal{L}_{\text{id}}(-E[\Delta\beta_1^{\text{rLS}} | \Delta y_1^{\text{rLS}}])$, wobei $\Delta\beta_{1|1}^{\text{rLS}} = \beta_1 - \beta_{1|1}^{\text{rLS}}$. Die Multiplikation mit F und anschließende Faltung mit dem symmetrisch verteilten v_2 erhält die Symmetrie, so daß auch $\mathcal{L}_{\text{id}}(\Delta\beta_2^{\text{rLS}}) = \mathcal{L}_{\text{id}}(-\Delta\beta_2^{\text{rLS}})$. Daher ist insbesondere $E[\beta_{1|1}^{\text{rLS}}] = E[\beta_1]$. Der Induktionsschluß erfolgt mit denselben Argumenten.

Zum rIC-Filter: Wir betrachten die gemeinsame Verteilung von $(\Delta\beta_t^{\text{rIC}}, \Delta\beta_t^{\text{KK}})$. Im Fall $t = 1$ besteht dieser Vektor aus zwei identischen Komponenten und ist daher wie im rLS-Fall symmetrisch. Wegen der Symmetrie von ε folgt nach Lemma 6.1.1 mit den Identifikationen

$$X \rightsquigarrow \begin{pmatrix} \Delta\beta_t^{\text{rIC}} \\ \Delta\beta_1^{\text{KK}} \end{pmatrix}, \quad Y \rightsquigarrow \begin{pmatrix} \Delta\beta_1^{\text{KK}} - \Delta\beta_1^{\text{rIC}} \\ \Delta y_1^{\text{rIC}} \\ \Delta y_1^{\text{KK}} \end{pmatrix},$$

$$f(Y) = f \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \left(\begin{pmatrix} A_1(\Sigma_{1|0}^{\text{KK}})^{-1}(Y_2 - Y_3) \min\{1, \frac{b_2}{|A_1(\Sigma_{1|0}^{\text{KK}})^{-1}Y_2|}\} + \\ + A_1 Z_1^T V_1^{-1} Y_1 \min\{1, \frac{b_1}{|A_1 Z_1^T V_1^{-1} Y_1|}\} \\ M_1^{\text{KK}}(Y_3 - Z_1 \Delta\beta_1^{\text{KK}}) \end{pmatrix} \right),$$

daß $\mathcal{L}_{\text{id}}(\Delta\beta_{1|1}^{\text{rIC}}) = \mathcal{L}_{\text{id}}(-\Delta\beta_{1|1}^{\text{rIC}})$ mit $\Delta\beta_{1|1}^{\text{rIC}} = \beta_1 - \beta_{1|1}^{\text{rIC}}$. Wie beim rLS-Fall folgt die Symmetrie von $\mathcal{L}_{\text{id}}(\Delta\beta_2^{\text{rLS}}) = \mathcal{L}_{\text{id}}(-\Delta\beta_2^{\text{rLS}})$, und der Induktionsschluß erfolgt wieder mit denselben Argumenten.

Der Schluß im mIC-Fall erfolgt analog, wird hier aber nicht präsentiert.

In allen Fällen ist jeweils wegen der Symmetrie $\text{E}_{\text{id}}[\Delta\beta_{t|t[-1]}] = 0$, also $\text{E}_{\text{id}}[\beta_{t|t[-1]}] = \text{E}_{\text{id}}[\beta_t]$. ////

6.2 Dichten der rekursiven Filter

In diesem Abschnitt untersuchen wir die Lebesgue-Stetigkeit der Vorhersagefehler $\Delta\beta_t$ bzw. der Filterfehler $\Delta\beta_{t|t}$ für den rLS- und den rIC-Filter und werden jeweils deren Dichten bestimmen.

Um auch den rIC-Filter erfassen zu können, führen wir zunächst mit

$$\sigma(y_{1:0}^0) := (\emptyset, \Omega), \quad \sigma(y_{1:t}^0) := \sigma(y_{1:t-1}^0, y_t, \Delta\beta_t^{\text{KK}}) \quad (6.6)$$

die um die Kenntnis von $\Delta\beta_t^{\text{KK}}$ augmentierte σ -Algebra ein. Wir beginnen mit einem Resultat über allgemeine $\sigma(y_{1:t}^{[0]})$ -adaptierte Filter, deren Beweis klar ist und daher nicht präsentiert wird.

Betrachten wir dazu einen wie folgt rekursiv definierten Filter

$$\begin{aligned} \tilde{\beta}_{0|0} &= \text{E}[\beta_0] = a_0 \\ \text{und für } t > 0 \quad \tilde{\beta}_{t|t-1} &= F_t \tilde{\beta}_{t-1|t-1} \\ \tilde{\beta}_{t|t} &= \tilde{\beta}_{t|t-1} + f_t \end{aligned} \quad (6.7)$$

mit f_t irgendeiner $y_{1:t}^{[0]}$ -meßbaren Funktion.

Proposition 6.2.1 *Sind die Innovationen v_t , die Fehler ε_s und die Startverteilung β_0 (als Prozesse) unabhängig mit unabhängigen Randverteilungen, und besitzen v_t , ε_s q - resp. p -dimensionale Lebesgue-Dichten p^{v_t} resp. p^{ε_s} , so besitzen bei Verwendung eines $y_{1:t}^{[0]}$ -adaptierten Filters f_t gemäß (7) auch die Randverteilungen von $\Delta\beta_{t|t}$ p -dimensionale Lebesgue-Dichten*

$$p^{\Delta\beta_1}(u) = \int p^{v_1}(u - F_1(\beta - \beta_{0|0})) P^{\beta_0}(d\beta) \quad (6.8)$$

$$p^{\Delta\beta_{t|t}}(u) = \int p^{\varepsilon_t}(y_t - Z_t(u + f_t)) p^{\Delta\beta_t|y_{1:t-1}}(u + f_t) P^{y_{1:t-1}}(dy_{1:t-1}) \lambda^q(dy_t) \quad (6.9)$$

$$p^{\Delta\beta_{t+1}}(u) = \int p^{v_{t+1}}(u - F_{t+1}\beta) P^{\Delta\beta_{t|t}}(d\beta) \quad (6.10)$$

BEWEIS: Wir zeigen die Behauptung per Induktion. (8) folgt unmittelbar aus der vorausgesetzten Lebesgue-Stetigkeit von $\mathcal{L}(\beta_0)$. In Lemma 6.3.4 werden wir sehen, daß $\sigma(\Delta y_{1:t}^{[0]}) = \sigma(y_{1:t}^{[0]})$. Wir machen daher keinen unterschied, ob f_t als Funktion in $y_{1:t}^{[0]}$ oder $\Delta y_{1:t}^{[0]}$ aufzufassen ist. Setzen wir formal $y_{1:0}$ als die leere Bedingung, so folgt aus $\Delta\beta_{1|1} = \Delta\beta_1 - f_1$, $\Delta y_1 = Z_1 \Delta\beta_1 + \varepsilon_1$, daß $\Delta\beta_1 = \Delta\beta_{1|1} + f_1$ und $\varepsilon_1 = \Delta y_1 - Z_1(\Delta\beta_{1|1} + f_1)$, und damit wegen der Unabhängigkeit von ε_1 und $\Delta\beta_1$, daß

$$\begin{aligned} P(\Delta\beta_{1|1} \in A) &= \int_{\Delta\beta_{1|1} \in A} p^{\varepsilon_1}(\varepsilon) \lambda^q(d\varepsilon) p^{\Delta\beta_1} \lambda^p(d\beta) = \\ &= \int_{u \in A} p^{\varepsilon_1}(y_1 - Z_1(u + f_1)) p^{\Delta\beta_1}(u + f_1) \lambda^q(dy_1) \lambda^p(du) \end{aligned}$$

Aus der Unabhängigkeit von v_{t+1} und $\Delta\beta_t, y_{1:t}$, folgt unmittelbar daß die bedingte Verteilung $\mathcal{L}(\Delta\beta_{t+1}|y_{1:t} = y_{1:t}^{\natural})$ eine λ^p -Dichte besitzt, und zwar

$$p^{\Delta\beta_{t+1}|y_{1:t}=y_{1:t}^{\natural}}(u) = \int p^{v_{t+1}}(u - F_{t+1}\beta) P^{\Delta\beta_t|y_{1:t}=y_{1:t}^{\natural}}(d\beta)$$

Hieraus folgt sofort (10). Für $t > 1$ müssen wir noch die Vergangenheit $y_{1:t-1}$ berücksichtigen. Ansonsten erhalten wir mit denselben Argumenten — und weil $\mathcal{L}(\Delta\beta_{t+1}|y_{1:t} = y_{1:t}^{\natural})$ eine λ^p -Dichte besitzt

$$\begin{aligned} P(\Delta\beta_t|t \in A) &= \int_{\Delta\beta_{t|t} \in A} p^{\varepsilon_t}(\epsilon) \lambda^q(d\epsilon) P^{\Delta\beta_t, y_{1:t+1}}(d\beta, dy_{1:t+1}) = \\ &= \int_{\Delta\beta_{t|t} \in A} p^{\varepsilon_t}(\epsilon) \lambda^q(d\epsilon) p^{\Delta\beta_t|y_{1:t+1}=y_{1:t+1}^{\natural}}(\beta) \lambda^p(d\beta) P^{y_{1:t+1}}(dy_{1:t+1}) = \\ &= \int_{u \in A} p^{\varepsilon_t}(y_t - Z_t(u + f_t)) p^{\Delta\beta_t|y_{1:t+1}=y_{1:t+1}^{\natural}}(u + f_t) P^{y_{1:t+1}}(dy_{1:t+1}) \\ &\quad \lambda^q(dy_t) \lambda^p(du) \end{aligned}$$

///

6.2.1 Dichte von $\Delta\beta_{t|t}^{\text{rLS}}$

Um die Terme beim rLS-Filter analytisch genauer untersuchen zu können, geben wir uns für $p = 1$ Werte von b , M , S , sowie Matrizen Z und V vor, spezifizieren $\mathcal{L}(\varepsilon_t) = \mathcal{N}_q(0, V)$ und tun so, als ob $\mathcal{L}(\Delta\beta_t^{\text{rLS}}) = \mathcal{N}_1(0, S)$. Mit diesen Vorgaben wollen wir die Dichte von $\Delta\beta_{t|t}^{\text{rLS}}$ berechnen.

Bemerkung 6.2.2 Der Fall höherdimensionaler Beobachtungen ist durch die Untersuchung für $q = 1$ abgedeckt, denn zu gegebenen $Z \in \mathbb{R}^{1 \times p}$ und $M \in \mathbb{R}^{p \times 1}$ gibt es Zahlen $z, m \in \mathbb{R}$ mit

$$\mathcal{L}((X, MZX + M\varepsilon)) = \mathcal{L}((X, mzX + m\tilde{\varepsilon}))$$

mit $\tilde{\varepsilon} \sim \mathcal{N}_1(0, \tilde{V})$, unabhängig von X und $\tilde{V} > 0$ beliebig, denn setzen wir $m = \sqrt{MVM^\tau/\tilde{V}}$ und $z = M^\tau Z/m$, so gilt:

$$\begin{aligned} \mathcal{L}((X, MZX + M\varepsilon)) &= \mathcal{N}_2\left(0, \begin{pmatrix} S & SM^\tau Z \\ SM^\tau Z & S(M^\tau Z)^2 + MVM^\tau \end{pmatrix}\right) = \\ &= \mathcal{N}_2\left(0, \begin{pmatrix} S & Smz \\ Smz & S(mz)^2 + m^2\tilde{V} \end{pmatrix}\right) \end{aligned}$$

Daher dürfen wir uns im folgenden auf $q = 1$ beschränken.

6.2.2 Dichte von $\Delta\beta_{t|t}^{\text{rLS}}$ bei normalen, 1-dim. v_t, ε_s

Im weiteren wollen wir $\Delta\beta_t^{\text{rLS}}$ durch X , $\Delta\beta_{t|t}^{\text{rLS}}$ durch U abkürzen und alle Abhängigkeiten in t unterdrücken. Es gilt

Proposition 6.2.3

$$\begin{aligned}
p^U(u) &= \frac{1}{\sqrt{S}} \left[1 - \Phi\left(\frac{MZu + (1 - MZ)b}{\sqrt{V}|M|}\right) \right] \varphi\left(\frac{u - b}{\sqrt{S}}\right) + \\
&\quad + \frac{1}{\gamma} \varphi\left(\frac{u}{\gamma}\right) \left[\Phi\left(\frac{(1 - \mu)u + b}{\sigma}\right) - \Phi\left(\frac{(1 - \mu)u - b}{\sigma}\right) \right] + \\
&\quad + \frac{1}{\sqrt{S}} \Phi\left(\frac{MZu - (1 - MZ)b}{\sqrt{V}|M|}\right) \varphi\left(\frac{u + b}{\sqrt{S}}\right)
\end{aligned}$$

mit

$$\gamma^2 := (1 - MZ)^2 S + M^2 V, \quad \sigma^2 := \frac{M^2 V S}{\gamma^2}, \quad \mu := \frac{S(1 - MZ)}{\gamma^2}, \quad (6.11)$$

BEWEIS: Offenbar ist

$$U = \begin{cases} (1 - MZ)X - M\varepsilon & \text{für } -b/|M| - ZX \leq \varepsilon \leq b/|M| - ZX \\ X - b \operatorname{sign}(M) & \varepsilon \geq b/|M| - ZX \\ X + b \operatorname{sign}(M) & \varepsilon \leq -b/|M| - ZX \end{cases} \quad (6.12)$$

Dazu definieren wir

$$y_1 = y_1(X) = -b/M - ZX \quad (6.13)$$

$$y_2 = y_2(X) = b/M - ZX \quad (6.14)$$

$$y_3 = y_3(u, X) = \frac{u - (1 + MZ)X}{M} \quad (6.15)$$

$$x_1 = x_1(u) = u - b \quad (6.16)$$

$$x_2 = x_2(u) = u + b \quad (6.17)$$

Dann gilt

$$P(U \leq u) = P(X \leq x_1) + P(x_1 \leq X \leq x_2, \varepsilon \operatorname{sign}(M) \geq y_3 \operatorname{sign}(M)) \quad (6.18)$$

und somit haben wir mit $\Phi(\cdot)$ der Verteilungsfunktion von $\mathcal{N}(0, 1)$ und φ deren Dichte

$$P(U \leq u) = \phi\left(\frac{x_1}{\sqrt{S}}\right) + \int_{x_1 \leq x \leq x_2} \frac{1}{\sqrt{S}} \left[1 - \Phi\left(\operatorname{sign}(M) \frac{y_3}{\sqrt{V}}\right) \right] \varphi\left(\frac{x}{\sqrt{S}}\right) dx$$

Damit erhalten wir als Dichte

$$\begin{aligned}
p^U(u) &= \frac{1}{\sqrt{S}} \left[\Phi\left(\operatorname{sign}(M) \frac{y_3(u, x_1)}{\sqrt{V}}\right) \varphi\left(\frac{x_1}{\sqrt{S}}\right) + \Phi\left(-\operatorname{sign}(M) \frac{y_3(u, x_2)}{\sqrt{V}}\right) \varphi\left(\frac{x_2}{\sqrt{S}}\right) \right] + \\
&\quad + \frac{1}{\sqrt{V} S M} \int_{x_1}^{x_2} \varphi\left(\frac{y_3(u, x)}{\sqrt{V}}\right) \varphi\left(\frac{x}{\sqrt{S}}\right) dx
\end{aligned}$$

Daher folgt mit quadratischer Ergänzung

$$\frac{1}{M\sqrt{S}\sqrt{V}} \varphi\left(\frac{y_3(u, x)}{\sqrt{V}}\right) \varphi\left(\frac{x}{\sqrt{S}}\right) = \frac{1}{\sigma} \varphi\left(\frac{x - \mu u}{\sigma}\right) \frac{1}{\gamma} \varphi\left(\frac{u}{\gamma}\right)$$

und damit

$$\begin{aligned}
p^U(u) &= \frac{1}{\sqrt{S}} \left[\Phi\left(\operatorname{sign}(M) \frac{y_3(u, x_1)}{\sqrt{V}}\right) \varphi\left(\frac{x_1}{\sqrt{S}}\right) + \Phi\left(-\operatorname{sign}(M) \frac{y_3(u, x_2)}{\sqrt{V}}\right) \varphi\left(\frac{x_2}{\sqrt{S}}\right) \right] + \\
&\quad + \frac{1}{\gamma} \varphi\left(\frac{u}{\gamma}\right) \left[\Phi\left(\frac{x_2 - \mu u}{\sigma}\right) - \Phi\left(\frac{x_1 - \mu u}{\sigma}\right) \right]
\end{aligned}$$

Setzen wir nun wieder die Werte für x_i und $y_i(x_i)$ ein, so ergibt dies

$$\begin{aligned} p^U(u) &= \frac{1}{\sqrt{S}} \left[1 - \Phi\left(\frac{MZu + (1 - MZ)b}{\sqrt{V}|M|}\right) \right] \varphi\left(\frac{u - b}{\sqrt{S}}\right) + \\ &\quad + \frac{1}{\gamma} \varphi\left(\frac{u}{\gamma}\right) \left[\Phi\left(\frac{(1 - \mu)u + b}{\sigma}\right) - \Phi\left(\frac{(1 - \mu)u - b}{\sigma}\right) \right] + \\ &\quad + \frac{1}{\sqrt{S}} \Phi\left(\frac{MZu - (1 - MZ)b}{\sqrt{V}|M|}\right) \varphi\left(\frac{u + b}{\sqrt{S}}\right) \end{aligned}$$

////

Bemerkung 6.2.4 Offenbar gilt für $b \rightarrow \infty$, daß $\varphi\left(\frac{u \pm b}{\sqrt{S}}\right) \rightarrow 0$, wobei die beiden zugehörigen Faktoren beschränkt sind, und daß gleichzeitig

$$\text{sign}(M) \left[\Phi\left(\frac{(1 - \mu)u + b}{\sigma}\right) - \Phi\left(\frac{(1 - \mu)u - b}{\sigma}\right) \right] \rightarrow 1,$$

so daß wir im Grenzfall die Dichte einer $\mathcal{N}(0, \gamma^2)$ -Verteilung erhalten.

Als Beispiel haben wir in Abbildung 6.1 die Dichten von $\Delta\beta_{t|t}^{\text{KK}}$ und $\Delta\beta_{t|t}^{\text{rLS}}$ im Steady-State Modell aus Beispiel 1.4.1 im Grenzfall, also

$$S = \lim_t \Sigma_{t|t-1}^{\text{KK}} = (\sqrt{5} + 1)/2 \doteq 1.618, \quad \text{resp. } M = (\sqrt{5} - 1)/2 \doteq 0.618$$

für $b = 1.14$ abgetragen, was einem Radius¹ von 10% und $\delta = 17.8\%$ entspricht. Daß die Dichte von $\Delta\beta_{t|t}^{\text{rLS}}$ für betragsmäßig große β schließlich oberhalb von der von $\Delta\beta_{t|t}^{\text{KK}}$ verläuft, ist auf dieser Skalierung nicht zu erkennen.

6.2.3 Dichte von $\Delta\beta_{t|t}^{\text{rIC}}$

Im rIC-Fall ist die Lage etwas komplizierter, da neben der Vergangenheit des Filters selbst auch immer der klassische Kalman-Filter mitbetrachtet werden muß. Hier stellen wir nur für $p = 1$ eine Aussage auf, der Fall $q \geq 1$ kann wie beim rLS-Filter behandelt werden, und wir beschränken uns auf $q = 1$.

Zur Erinnerung: Der Filter-Schritt beim rIC besagt für $p = q = 1$, daß

$$\beta_{t|t}^{\text{rIC}} = \beta_{t|t-1}^{\text{rIC}} + \max\{-b_t, \min\{A_t(Z/V\Delta y_t^{\text{rIC}} + (\beta_{t|t-1}^{\text{rIC}} - \beta_{t|t-1}^{\text{KK}})/\Sigma_{t|t-1}^{\text{KK}}), b_t\}\}$$

mit $\beta_{t|t-1}^{\text{KK}}$ der Ein-Schritt-Vorhersage aus dem klassischen Kalman-Filter.

Für die Untersuchung geben wir uns Clippingschwellen b_t , klassische Kalman-Gains $M_t = M_t^{\text{KK}}$, Lagrange-Multiplikatoren A_t und klassische Fehlervarianzen $S_t = \Sigma_{t|t-1}^{\text{KK}}$, sowie Hyper-Parameter F , $Z > 0$, Q und V vor und kürzen $\Delta\beta_{t|t}^{\text{rIC}}$ durch U_t , $\Delta\beta_{t|t}^{\text{rIC}}$ durch X_t , $\beta_{t|t}^{\text{rIC}} - \beta_{t|t}^0$ durch R_t und $\beta_{t|t-1}^{\text{rIC}} - \beta_{t|t-1}^0$ durch W_t ab.

Wie im rLS-Filter spezifizieren wir $\mathcal{L}(\beta_0) = \mathcal{L}(v_t) = \mathcal{N}_1(0, Q)$, $\mathcal{L}(\varepsilon_t) = \mathcal{N}_q(0, V)$. Mit diesen Vorgaben wollen wir die Dichte von $\Delta\beta_{t|t}^{\text{rIC}}$ berechnen.

¹vgl. Kapitel 8

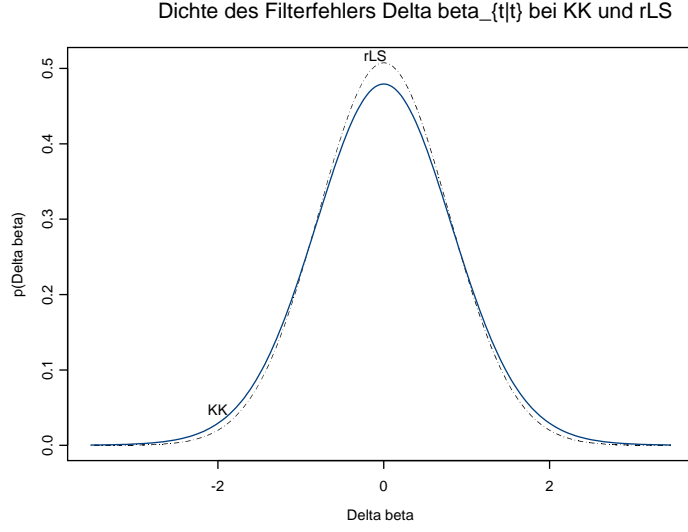


Abbildung 6.1: Dichten von $\Delta\beta_{t|t}^{\text{KK}}$ und $\Delta\beta_{t|t}^{\text{rLS}}$ im Steady-State Modell für $\delta = 17.8\%$

Proposition 6.2.5 *Sind die Innovationen v_t , die Fehler ε_s und die Startverteilung β_0 (als Prozesse) unabhängig mit unabhängigen Randverteilungen, und sind β_0 , v_t , ε_s absolutstetig mit Lebesgue-Dichten p^{β_0} , p^{v_t} resp. p^{ε_s} , so besitzen die Paare (U_t, R_t) und (X_t, W_t) aus der Definition des rIC-Filters für $t \geq 1$ resp. $t \geq 2$ 2-dimensionale Lebesgue-Dichten*

$$p^{\Delta\beta_1^0}(u) = \int p^{v_1}(u - F(w - \beta_{0|0}))p^{\beta_0}(w) dw \quad (6.19)$$

$$\begin{aligned} p^{U_1, R_1}(u, r) &= \mathbb{I}_{\{r_{1,1} \leq r \leq r_{2,1}\}} p^{\varepsilon_1}\left(\frac{V-Z^2 A_1}{Z A_1 - V M_1} r - Z u\right) p^{\Delta\beta_1^0}\left(u + \frac{V}{Z A_1 - V M_1} r\right) + \\ &\quad + \mathbb{I}_{\{r \geq r_{1,1}\}} p^{\varepsilon_1}\left(\frac{-b_1 - r}{M_1} - Z(u - b_1)\right) p^{\Delta\beta_1^0}(u - b_1) + \\ &\quad + \mathbb{I}_{\{r \leq r_{2,1}\}} p^{\varepsilon_1}\left(\frac{b_1 - r}{M_1} - Z(u + b_1)\right) p^{\Delta\beta_1^0}(u + b_1) \end{aligned} \quad (6.20)$$

$$p^{X_t, W_t}(x, w) = \frac{1}{|F|} \int p^{v_t}(x - F u) p^{U_{t-1}, R_{t-1}}(u, w/F) du \quad (6.21)$$

$$\begin{aligned} p^{U_t, R_t}(u, r) &= \frac{V}{Z A_t - V M_t} \int_{w_{1,t}(r)}^{w_{2,t}(r)} p^{\varepsilon_t}\left(\frac{V-Z^2 A_t}{V} y_{3,t}(r, w) - Z u - Z \frac{A_t}{S_t} w\right) \times \\ &\quad \times p^{X_t, W_t}\left(u + \frac{A_t Z}{V} y_{3,t}(r, w) + \frac{A_t}{S_t} w, w\right) dw + \\ &\quad + \frac{1}{M_t} \int_{-\infty}^{w_{2,t}(r)} p^{\varepsilon_t}(y_4(r, w) - Z(u - b_t)) p^{X_t, W_t}(u - b_t, w) dw + \\ &\quad + \frac{1}{M_t} \int_{w_{1,t}(r)}^{\infty} p^{\varepsilon_t}(y_5(r, w) - Z(u + b_t)) p^{X_t, W_t}(u + b_t, w) dw + \\ &\quad + p^{X_t, W_t}(u - b_t, w_{2,t}(r)) \int_{y_{4,t}(r, w_{2,t}(r))}^{y_{1,t}(w_{2,t}(r))} p^{\varepsilon_t}(y - Z(u - b_t)) dy + \\ &\quad + p^{X_t, W_t}\left(u + \frac{A_t Z}{V} y + \frac{A_t}{S_t} w_{2,t}(r), w_{2,t}(r)\right) \times \end{aligned}$$

$$\begin{aligned}
& \times \int_{y_{1,t}(w_{2,t}(r))}^{y_{3,t}(r,w_{2,t}(r))} p^{\varepsilon_t}(y - Z[u + A_t(Z/Vy + w_{2,t}(r)/S_t)]) dy + \\
& + p^{X_t, W_t}(u + \frac{A_t Z}{V}y + \frac{A_t}{S_t}w_{1,t}(r)), w_{1,t}(r)) \times \\
& \times \int_{y_{3,t}(r,w_{1,t}(r))}^{y_{2,t}(w_{1,t}(r))} p^{\varepsilon_t}(y - Z[u + A_t(Z/Vy + w_{1,t}(r)/S_t)]) dy + \\
& + p^{X_t, W_t}(u + b_t, w_{1,t}(w)) \int_{y_{2,t}(w_{1,t}(r))}^{y_{5,t}(r,w_{1,t}(r))} p^{\varepsilon_t}(y - Z(u + b_t)) dy. \quad (6.22)
\end{aligned}$$

mit

$$\begin{aligned}
r_{1,1} &= -b(1 - \frac{M_1 V}{A_1 Z}) & r_{2,1} &= b(1 - \frac{M_1 V}{A_1 Z}) \\
w_{1,t} &= w_{1,t}(r) = r - b(1 - \frac{M_t V}{A_t Z}) & w_{2,t} &= w_{2,t}(r) = r + b(1 - \frac{M_t V}{A_t Z}) \\
y_{1,t} &= y_{1,t}(w) = -b_t \frac{V}{A_t Z} - \frac{V}{Z S_t} w, & y_{2,t} &= y_{2,t}(w) = b_t \frac{V}{A_t Z} - \frac{V}{Z S_t} w \\
y_{3,t} &= y_{3,t}(r, w) = \frac{V[r - (A_t/S_t + M_t Z)w]}{Z A_t - V M_t} \\
y_{4,t} &= y_{4,t}(r, w) = Z w - \frac{r + b_t}{M_t} & y_{5,t} &= y_{5,t}(r, w) = Z w - \frac{r - b_t}{M_t}
\end{aligned}$$

BEWEIS: Im ersten Schritt ist $\Delta\beta_1^{\text{rIC}} \equiv \Delta\beta_1^{\text{rK}}$. Dementsprechend ist, wenn wir y für Δy_0 schreiben,

$$(U_1, R_1)^\tau = \begin{cases} (X_0 + b_1, -b_1 - M_1 y)^\tau & \text{für } y \leq -b_1 \frac{V}{A_1 Z} \\ (X_0 - A_1 Z/V y, (A_1 Z/V - M_1) y)^\tau & \text{für } \frac{-b_1 V}{A_1 Z} \leq y \leq \frac{b_1 V}{A_1 Z} \\ (X_0 - b_1, b_1 - M_1 y)^\tau & \text{für } y \geq b_1 \frac{V}{A_1 Z} \end{cases}$$

Also ist

$$\begin{aligned}
& (U_1 \leq u, R_1 \leq r) = (X_0 + b_1 \leq u, -b_1 - M_1 y \leq r, y \leq -b_1 \frac{V}{A_1 Z}) \dot{\cup} \\
& \dot{\cup} (X_0 - A_1 Z/V y \leq u, (A_1 Z/V - M_1) y \leq r, \frac{-b_1 V}{A_1 Z} \leq y \leq \frac{b_1 V}{A_1 Z}) \dot{\cup} \\
& \dot{\cup} (X_0 - b_1 \leq u, b_1 - M_1 y \leq r, y \geq b_1 \frac{V}{A_1 Z}) \\
& = (X_0 + b_1 \leq u, r \geq r_{1,1}, \frac{-b_1 - r}{M_1} \leq y \leq -b_1 \frac{V}{A_1 Z}) \dot{\cup} \\
& \dot{\cup} (X_0 - A_1 Z/V y \leq u, r_{1,1} \leq r \leq r_{2,1}, \frac{-b_1 V}{A_1 Z} \leq y \leq r \frac{V}{A_1 Z - M_1 V}) \dot{\cup} \\
& \dot{\cup} (X_0 - A_1 Z/V y \leq u, r \geq r_{2,1}, -b_1 \frac{V}{A_1 Z} \leq y \leq b_1 \frac{V}{A_1 Z}) \dot{\cup} \\
& \dot{\cup} (X_0 - b_1 \leq u, r \geq r_{2,1}, y \geq b_1 \frac{V}{A_1 Z}) \dot{\cup} \\
& \dot{\cup} (X_0 - b_1 \leq u, r \leq r_{2,1}, y \geq \frac{b_1 - r}{M_1})
\end{aligned}$$

und so erhalten wir

$$\begin{aligned}
P(U_1 \leq u, R_1 \leq r) &= \int \mathbf{I}_{\{x+b_1 \leq u\}} \mathbf{I}_{\{r \geq r_{1,1}\}} \mathbf{I}_{\{\frac{-b_1 - r}{M_1} \leq y \leq -b_1 \frac{V}{A_1 Z}\}} + \\
& + \mathbf{I}_{\{x - A_1 Z/V y \leq u\}} \mathbf{I}_{\{r_{1,1} \leq r \leq r_{2,1}\}} \mathbf{I}_{\{-b_1 \frac{V}{A_1 Z} \leq y \leq r \frac{V}{A_1 Z - M_1 V}\}} + \\
& + \mathbf{I}_{\{x - A_1 Z/V y \leq u\}} \mathbf{I}_{\{r \geq r_{2,1}\}} \mathbf{I}_{\{-b_1 \frac{V}{A_1 Z} \leq y \leq b_1 \frac{V}{A_1 Z}\}} + \\
& + \mathbf{I}_{\{x - b_1 \leq u\}} \mathbf{I}_{\{r \leq r_{2,1}\}} \mathbf{I}_{\{y \geq \frac{b_1 - r}{M_1}\}} + \\
& + \mathbf{I}_{\{x - b_1 \leq u\}} \mathbf{I}_{\{r \geq r_{2,1}\}} \mathbf{I}_{\{y \geq b_1 \frac{V}{A_1 Z}\}} p^{\varepsilon_1}(y - Zx) p^{\Delta\beta_1^0}(x) dy dx = \\
& = \int \mathbf{I}_{\{r \geq r_{1,1}\}} \mathbf{I}_{\{\tilde{x} \leq u\}} \mathbf{I}_{\{\frac{-b_1 - r}{M_1} \leq y \leq -b_1 \frac{V}{A_1 Z}\}} p^{\varepsilon_1}(y - Z(\tilde{x} - b_1)) \times
\end{aligned}$$

$$\begin{aligned}
& \times p^{\Delta\beta_1^0}(\tilde{x} - b_1) dy d\tilde{x} + \\
& + \int \mathbf{I}_{\{\tilde{x} \leq u\}} \left\{ \mathbf{I}_{\{r_{1,1} \leq r \leq r_{2,1}\}} \mathbf{I}_{\{-b_1 \frac{V}{A_1 Z} \leq y \leq r \frac{V}{A_1 Z - M_1 V}\}} + \right. \\
& \quad \left. + \mathbf{I}_{\{r \geq r_{2,1}\}} \mathbf{I}_{\{-b_1 \frac{V}{A_1 Z} \leq y \leq b_1 \frac{V}{A_1 Z}\}} \right\} \times \\
& \times p^{\varepsilon_1}(y - Z(\tilde{x} + A_1 Z/Vy)) p^{\Delta\beta_1^0}(\tilde{x} + A_1 Z/Vy) dy d\tilde{x} + \\
& + \int \mathbf{I}_{\{\tilde{x} \leq u\}} \left\{ \mathbf{I}_{\{r \leq r_{2,1}\}} \mathbf{I}_{\{y \geq \frac{b_1 - r}{M_1}\}} + \mathbf{I}_{\{r \geq r_{2,1}\}} \mathbf{I}_{\{y \geq b_1 \frac{V}{A_1 Z}\}} \right\} \times \\
& \times p^{\varepsilon_1}(y - Z(\tilde{x} + b_1)) p^{\Delta\beta_1^0}(\tilde{x} + b_1) dy d\tilde{x},
\end{aligned}$$

so daß wir als Dichte wie behauptet erhalten

$$\begin{aligned}
P^{U_1, R_1}(u, r) &= \mathbf{I}_{\{r \geq r_{1,1}\}} \frac{1}{M_1} p^{\varepsilon_1}\left(\frac{-b_1 - r}{M_1} - Z(u - b_1)\right) p^{\Delta\beta_1^0}(u - b_1) + \\
& + \mathbf{I}_{\{r_{1,1} \leq r \leq r_{2,1}\}} \frac{V}{A_1 Z - M_1 V} p^{\varepsilon_1}\left(r \frac{V - A_1 Z^2}{A_1 Z - M_1 V} - Zu\right) p^{\Delta\beta_1^0}\left(u + r \frac{A_1 Z}{A_1 Z - M_1 V}\right) + \\
& + \mathbf{I}_{\{r \leq r_{2,1}\}} \frac{1}{M_1} p^{\varepsilon_1}\left(\frac{b_1 - r}{M_1} - Z(u + b_1)\right) p^{\Delta\beta_1^0}(u + b_1)
\end{aligned}$$

Wie sich die p^{X_t, W_t} -Dichte ergibt, ist klar. Für $t > 1$ ist mit $y = \Delta y_t^{\text{ric}}$ und $w = W_t$, sowie mit

$$r_{1,t} = r_{1,t}(w) = -b(1 - \frac{M_t V}{A_t Z}) + w \quad r_{2,t} = r_{2,t}(w) = b(1 - \frac{M_t V}{A_t Z}) + w$$

$$(U_t, R_t)^\tau = \begin{cases} (X_t + b_t, -b_t - M_t y + M_t Z w)^\tau & \text{für } y \leq y_{1,t} \\ \left(\begin{array}{l} X_t - A_t(Z/Vy + w/S_t), \\ (\frac{A_t Z}{V} - M_t)y + (\frac{A_t}{S_t} + M_t Z)w \end{array} \right) & \text{für } y_{1,t} \leq y \leq y_{2,t} \\ (X_t - b_t, b_t - M_t y + M_t Z w)^\tau & \text{für } y \geq y_{2,t} \end{cases}$$

Daher kann man die Niveaumengen umformen gemäß

$$\begin{aligned}
& (U_t \leq u, R_t \leq r) = (X_t + b_t \leq u, -b_t - M_t y + M_t Z w \leq r, y \leq y_{1,t}(w)) \dot{\cup} \\
& \dot{\cup} \left(\begin{array}{l} X_t - A_t(Z/Vy + w/S_t) \leq u, (A_t Z/V - M_t)y + \frac{A_t}{S_t} w \leq r, \\ y_{1,t}(w) \leq y \leq y_{2,t}(w) \end{array} \right) \dot{\cup} \\
& \dot{\cup} (X_t - b_t \leq u, b_t - M_t y + M_t Z w \leq r, y \geq y_{2,t}(w)) \\
& = (X_t + b_t \leq u, r \geq r_{1,t}(w), y_{4,t}(r, w) \leq y \leq y_{1,t}(w)) \dot{\cup} \\
& \dot{\cup} \left(\begin{array}{l} X_t - A_t(Z/Vy + w/S_t) \leq u, r_{1,t}(w) \leq r \leq r_{2,t}(w), \\ y_{1,t}(w) \leq y \leq y_{3,t}(r, w) \end{array} \right) \dot{\cup} \\
& \dot{\cup} (X_t - A_t(Z/Vy + w/S_t) \leq u, r \geq r_{2,t}(w), y_{1,t}(w) \leq y \leq y_{2,t}(w)) \dot{\cup} \\
& \dot{\cup} (X_t - b_t \leq u, r \geq r_{2,t}(w), y \geq y_{2,t}(w)) \dot{\cup} \\
& \dot{\cup} (X_t - b_t \leq u, r \leq r_{2,t}(w), y \geq y_{5,t}(r, w)),
\end{aligned}$$

wobei wir ausnutzen, daß $M_t = ZS_t/(Z^2S_t + V)$.

Setzen wir ein, daß

$$(r \leq r_{2,t}(w)) = (w \geq w_{1,t}(r)), \quad \text{resp. } (r \leq r_{1,t}(w)) = (w \geq w_{2,t}(r)),$$

heißt das

$$(U_t \leq u, R_t \leq r) = (X_t + b_t \leq u, w \leq w_{2,t}(r), y_{4,t}(r, w) \leq y \leq y_{1,t}(w)) \dot{\cup}$$

$$\begin{aligned}
& \dot{\cup} \left(X_t - A_t(Z/Vy + w/S_t) \leq u, w_{1,t}(r) \leq w \leq w_{2,t}(r), \right. \\
& \quad \left. y_{1,t}(w) \leq y \leq y_{3,t}(r, w) \right) \dot{\cup} \\
& \dot{\cup} (X_t - A_t(Z/Vy + w/S_t) \leq u, w \leq r_{1,t}(r), y_{1,t}(w) \leq y \leq y_{2,t}(w)) \dot{\cup} \\
& \dot{\cup} (X_t - b_t \leq u, r \leq w_{1,t}(r), y \geq y_{2,t}(w)) \dot{\cup} \\
& \dot{\cup} (X_t - b_t \leq u, r \geq w_{1,t}(r), y \geq y_{5,t}(r, w)),
\end{aligned}$$

und so schreibt sich die Wahrscheinlichkeit als

$$\begin{aligned}
P(U_t \leq u, R_t \leq r) &= \int \mathbf{I}_{\{w \leq w_{2,t}(r)\}} \mathbf{I}_{\{\tilde{x} \leq u\}} \mathbf{I}_{\{y_{4,t}(r,w) \leq y \leq y_{1,t}(w)\}} \times \\
& \quad \times p^{\varepsilon_t}(y - Z(\tilde{x} - b_t)) p^{X_t, W_t}(\tilde{x} - b_t, w) dy d\tilde{x} dw + \\
& + \int \mathbf{I}_{\{\tilde{x} \leq u\}} \left\{ \mathbf{I}_{\{w_{1,t}(r) \leq w \leq w_{2,t}(r)\}} \mathbf{I}_{\{y_{1,t}(w) \leq y \leq y_{3,t}(r,w)\}} + \right. \\
& \quad \left. \mathbf{I}_{\{w \leq w_{1,t}(r)\}} \mathbf{I}_{\{y_{1,t}(w) \leq y \leq y_{2,t}(w)\}} \right\} \times \\
& \quad \times p^{\varepsilon_t}(y - Z[\tilde{x} + A_t(Z/Vy + w/S_t)]) \times \\
& \quad \times p^{X_t, W_t}(\tilde{x} + A_t(Z/Vy + w/S_t), w) dy d\tilde{x} dw + \\
& + \int \mathbf{I}_{\{\tilde{x} \leq u\}} \left\{ \mathbf{I}_{\{w \leq w_{1,t}(w)\}} \mathbf{I}_{\{y \geq y_{2,t}(w)\}} + \mathbf{I}_{\{w \geq w_{1,t}(w)\}} \mathbf{I}_{\{y \geq y_{5,t}(r,w)\}} \right\} \times \\
& \quad \times p^{\varepsilon_t}(y - Z(\tilde{x} + b_t)) p^{X_t, W_t}(\tilde{x} + b_t, w) dy d\tilde{x} dw,
\end{aligned}$$

so daß wir als Dichte wie behauptet erhalten

$$\begin{aligned}
P^{U_t, R_t}(u, r) &= \frac{V}{Z A_t - V M_t} \int_{w_{1,t}(r)}^{w_{2,t}(r)} p^{\varepsilon_t} \left(\frac{V - Z^2 A_t}{V} y_{3,t}(r, w) - Zu - Z \frac{A_t}{S_t} w \right) \times \\
& \quad \times p^{X_t, W_t} \left(u + \frac{A_t Z}{V} y_{3,t}(r, w) + \frac{A_t}{S_t} w, w \right) dw + \\
& + \frac{1}{M_t} \int_{-\infty}^{w_{2,t}(r)} p^{\varepsilon_t} (y_4(r, w) - Z(u - b_t)) p^{X_t, W_t}(u - b_t, w) dw + \\
& + \frac{1}{M_t} \int_{w_{1,t}(r)}^{+\infty} p^{\varepsilon_t} (y_5(r, w) - Z(u + b_t)) p^{X_t, W_t}(u + b_t, w) dw + \\
& + p^{X_t, W_t}(u - b_t, w_{2,t}(r)) \int_{y_{4,t}(r, w_{2,t}(r))}^{y_{1,t}(w_{2,t}(r))} p^{\varepsilon_t}(y - Z(u - b_t)) dy + \\
& + p^{X_t, W_t} \left(u + \frac{A_t Z}{V} y + \frac{A_t}{S_t} w_{2,t}(r), w_{2,t}(r) \right) \times \\
& \quad \times \int_{y_{1,t}(w_{2,t}(r))}^{y_{3,t}(r, w_{2,t}(r))} p^{\varepsilon_t}(y - Z[u + A_t(Z/Vy + w_{2,t}(r)/S_t)]) dy + \\
& + p^{X_t, W_t} \left(u + \frac{A_t Z}{V} y + \frac{A_t}{S_t} w_{1,t}(r), w_{1,t}(r) \right) \times \\
& \quad \times \int_{y_{3,t}(r, w_{1,t}(r))}^{y_{2,t}(w_{1,t}(r))} p^{\varepsilon_t}(y - Z[u + A_t(Z/Vy + w_{1,t}(r)/S_t)]) dy + \\
& + p^{X_t, W_t}(u + b_t, w_{1,t}(r)) \int_{y_{2,t}(w_{1,t}(r))}^{y_{5,t}(r, w_{1,t}(r))} p^{\varepsilon_t}(y - Z(u + b_t)) dy.
\end{aligned}$$

////

6.3 (asymptotische) Normalität der robusten Verfahren

6.3.1 Resultate als Folgerung von Proposition A.2.4

Wir verwenden nun Proposition A.2.4, um eine Reihe von Konsequenzen herzuleiten. Diese stehen insofern nicht im Widerspruch zu Heyde und Johnstone (1979), als daß deren Annahme (A3), die im wesentlichen ein unbeschränktes Anwachsen der Information über den gesuchten Parameter sichert, bei uns nicht erfüllt sein wird.

Nicht-Normalität von rLS, rIC

Unmittelbar aus Proposition A.2.4 folgt

Korollar 6.3.1 *Unter Annahme der Unabhängigkeit von $\{\varepsilon_t\}$ und $\{v_t\}$ als Prozesse ist im Zustandsraummodell (1), (2) die asymptotische Normalität des v_t -Prozeß notwendig für die asymptotische Normalität der β_t , y_t und jedes Filterfehlerprozesses $\Delta\beta_t$ gemäß (7). Dasselbe gilt für Δy_t , falls dieses nicht stochastisch gegen 0 konvergiert. Asymptotische Normalität der ε_t ist notwendig für die asymptotische Normalität der y_t und Δy_t . Weil der klassische Kalman-Filter in Δy_t linear ist, ist für dessen asymptotische Normalität die asymptotische Normalität der y_t notwendig, falls $M_t^{\text{KK}} \not\rightarrow 0$ — mit den Implikationen für ε_t und v_t .*

Proposition 6.3.2 *Sogar im idealen Modell (V3) bzw. (V4) können wir für jeden rekursiven Filter wie in (7) mit einem beschränkten f_t , i.e. $\|f_t\|_\infty \leq b$ für ein gewisses b , wobei $f_t \not\rightarrow 0$ stochastisch, die asymptotische Normalität der Vektoren $(\beta_{t|t}^\tau, \beta_{t-1|t-1}^\tau)^\tau$ und $(\Delta\beta_t^\tau, \Delta\beta_{t-1}^\tau)^\tau$ ausschließen, wobei wir für letztere zusätzlich fordern, daß $\infty > K > \|F_t\| > \delta > 0$ für alle t und $\text{rk} E[f_t f_t^\tau] = p$.*

BEWEIS: Nehmen wir die asymptotische Normalität der $(\beta_{t|t}^\tau, \beta_{t-1|t-1}^\tau)^\tau$ an. Dann muß auch $f_t = \beta_{t|t} - F_t \beta_{t-1|t-1}$ als eine lineare Transformation wieder asymptotisch normalverteilt sein, was im Widerspruch steht zur Beschränktheit von f_t . Was $(\Delta\beta_t^\tau, \Delta\beta_{t-1}^\tau)^\tau$ anlangt, so gilt, daß $-F_t f_{t-1} + v_t = \Delta\beta_t - F_t \Delta\beta_{t-1}$, was asymptotisch normalverteilt sein muß. Daher gilt mit Proposition A.2.4, daß $F_t f_{t-1}$ asymptotisch normalverteilt ist. Aber $F_t \not\rightarrow 0$, so daß $\text{rk} \ker F_t < p$, und weil $\text{rk} E[f_t f_t^\tau] = p$, folgt $\text{rk} E[F_t f_{t-1} (F_t f_{t-1})^\tau] \geq 1$ für alle t . Daher müßte der Träger des nichtdegenerierten Teils von $F_t f_{t-1}$ unbeschränkt sein, um normalverteilt sein zu können, was wieder der Beschränktheit von f_t widerspricht.

////

Normalität der Posterior-Verteilung im Zustandsraummodell

Unter recht allgemeinen Bedingungen für das rekursive Schema (7) sind $\Delta\beta|y_{1:t}^{[0]}$ und $\beta_t|y_{1:t-1}^{[0]}$ verstanden als Markov-Kerne, faktorisiert nach einer Realisation $y_{1:t[-1]}^{[0]}$ von $y_{1:t[-1]}^{[0]}$, normalverteilt:

Proposition 6.3.3 *Seien ε_t, v_t gemäß (V3) verteilt. Dann gilt für beliebige $y_{1:t}^{[0]}$ -meßbare Funktionen f_t , daß für jeden wie in (7) definierten Filter der bedingte Filterfehler $\Delta\beta_t|y_{1:t-1}^{[0]}$ normalverteilt ist.*

Vor dem Beweis zeigen wir ein Lemma, das auch allgemein von Interesse ist:

Lemma 6.3.4

$$\sigma(\Delta y_{1:t}^{[0]}) = \sigma(y_{1:t}^{[0]})$$

BEWEIS ZU LEMMA 6.3.4: Als Konstante ist $\beta_{0|0}$ sowohl $y_{1:0}^{[0]}$ - als auch $\Delta y_{1:0}^{[0]}$ -meßbar. Nehmen wir nun induktiv an, daß $\beta_{t-1|t-1}$ sowohl $y_{1:t-1}^{[0]}$ - als auch $\Delta y_{1:t-1}^{[0]}$ -meßbar ist, so folgt aus

$$\Delta y_t = y_t - Z_t F_t \beta_{t-1|t-1} \quad \text{resp.} \quad y_t = \Delta y_t + Z_t F_t \beta_{t-1|t-1}$$

daß $\sigma(\Delta y_{1:t}^{[0]}) = \sigma(y_{1:t}^{[0]})$, und somit tatsächlich auch, daß $\beta_{t|t} = F_t \beta_{t-1|t-1} + f_t$ sowohl $y_{1:t-1}^{[0]}$ - als auch $\Delta y_{1:t-1}^{[0]}$ -meßbar ist. ////

BEWEIS ZU PROPOSITION 6.3.3: $\Delta\beta_1 = F_1(\beta_0 - a_0) + v_1$ ist verteilt gemäß $\mathcal{N}(0, Q_1 + F_1 Q_0 F_1^T)$, und daran ändert auch das Bedingen nach der trivialen σ -Algebra nichts. Nehmen wir nun an, daß nach Induktion $\Delta\beta_t|y_{1:t-1}^{[0]}$ normalverteilt ist, dann gilt dies wegen der Unabhängigkeit von ε_t auch für die lineare Transformation aus ε_t und Δy_t^r , $(\Delta\beta_t^r, \Delta y_t^r)^T|y_{1:t-1}^{[0]}$. Aber die (regulären) bedingten Verteilungen eines gemeinsam normalverteilten Vektors sind nach Lemma A.2.1 wieder normalverteilt, daher ist $\Delta\beta_t|y_{1:t}^{[0]}$ normalverteilt. Betrachten wir nun $\Delta\beta_{t+1} = F_{t+1}\Delta\beta_t - F_{t+1}f_t + v_{t+1}$, so sehen wir, daß, bedingt nach $\Delta y_t^{[0]}$, man $F_{t+1}f_t$ als von $\Delta\beta_t$ und $\Delta y_t^{[0]}$ unabhängige Konstante auffassen kann, die nichts an der Normalität von $\Delta\beta_t|y_{1:t}^{[0]}$ und v_{t+1} ändert, und so bleibt auch die Summe normalverteilt. Mit der Äquivalenz der σ -Algebren, die wir oben gezeigt haben, folgt die Behauptung. ////

Falls v_t nicht normalverteilt ist, gilt diese Proposition nicht mehr, wie man sofort mit Proposition A.2.4 schließen kann: Die asymptotische Normalität bricht zusammen, wenn wir mit der Verteilung von $\Delta\beta_{t+1} = F_{t+1}\Delta\beta_t - F_{t+1}f_t + v_{t+1}$ bedingt nach $\Delta y_t^{[0]}$ argumentieren, ja es gilt sogar

Proposition 6.3.5 *Auch für die asymptotische Normalität der Posterior-Verteilung der β_t gegeben $y_{1:t-1}$ benötigt man die asymptotische Normalität der v_t , sofern die eindimensionalen Randverteilungen des v_t -Prozesses zumindest straff sind.*

BEWEIS: Nehmen wir die asymptotische Normalität der $\beta_t|y_{1:t-1}^{[0]}$ an. Wir befinden uns in einem polnischen Raum und können uns daher auf die faktorisierten Versionen bedingter Erwartungswerte zurückziehen. Die Annahme bedeutet, daß

$$\varphi_{\beta_t|y_{1:t-1}^{[0]}}(t, y_{1:t-1}^{[0]}) = \mathbb{E}[\exp(it\beta_t)|y_{1:t-1}^{[0]}] \rightarrow \varphi_{\mathcal{N}(a,C)},$$

wobei a und C von $y_{1:t-1}^{[0]}$ abhängen dürfen. Nun ist $\beta_t = F_t \beta_{t-1} + v_t$, und daher gilt wegen der Unabhängigkeit von v_t von sowohl β_{t-1} als auch $y_{1:t-1}^{[0]}$

$$\begin{aligned} \varphi_{\beta_t | y_{1:t-1}^{[0]}}(t, y_{1:t-1}^{[0]}) &= \mathbb{E}[\exp(it(F_t \beta_{t-1} + v_t)) | y_{1:t-1}^{[0]}] = \\ &= \mathbb{E}[\exp(it F_t \beta_{t-1}) | y_{1:t-1}^{[0]}] \mathbb{E}[\exp(it v_t) | y_{1:t-1}^{[0]}] = \\ &= \varphi_{F_t \beta_{t-1} | y_{1:t-1}^{[0]}}(t, y_{1:t-1}^{[0]}) \mathbb{E}[\exp(it v_t)] = \\ &= \varphi_{F_t \beta_{t-1} | y_{1:t-1}^{[0]}}(t, y_{1:t-1}^{[0]}) \varphi_{v_t}(t) \end{aligned}$$

Hier nun dürfen wir mit Proposition A.2.4 argumentieren und erhalten die Behauptung. ////

Bemerkung 6.3.6 Diese letzte Proposition sagt nichts über die Möglichkeit einer anderen Standardisierung: Wir schließen den Fall aus, daß $\beta_t | y_{1:t-1}^{[0]}$ nicht straff ist, das heißt $\|C\| = \infty$. Wenn wir auch andere Standardisierungen zulassen, ist es möglich, eine simultane Straffheit der standardisierten $s_t \beta_t | y_{1:t-1}^{[0]}$ und die stochastische Konvergenz der $s_t v_t$ gegen 0 zu erreichen, wie z.B. im Fall der Irrfahrt, und daher ist es auch nicht ausgeschlossen, daß die $s_t \beta_t | y_{1:t-1}^{[0]}$ asymptotisch normal sind, ohne daß dies auch die v_t sind. Man kann aber keine nicht-triviale Normalität erreichen, sofern $\|F_t \beta_{t-1} | y_{1:t-1}^{[0]}\| / \|v_t\|$ straff ist, Dies ist z.B. der Fall bei $\limsup_t \rho(F_t) < 1$.

Wir fassen zusammen:

Für ε_t , v_t nicht normal konnte man mithilfe des Cramér-Lévy-Theorems leicht nachweisen, daß dann keine asymptotische Normalität für die Filter- resp. Vorhersagefehler eines rekursiven Filters gemäß (7) vorliegen kann. Ebenso konnte gezeigt werden, daß bereits für den Prozeß der $(\Delta \beta_t, \Delta \beta_{t-1})_t$ zu einem solchen Filter die gemeinsame asymptotische Normalität widerlegt werden kann. Es bleibt aber die Frage, ob nicht die eindimensionalen Randverteilungen von $\Delta \beta_t$ asymptotisch normal sein können. Dies klären wir für Dimension $p = 1$ im rLS-Fall für b hinreichend groß mithilfe von Theorem 1.3.1.

6.3.2 Folgerungen aus den Dichte-Darstellungen aus Abschnitt 6.2

Mithilfe der Dichtedarstellungen aus Abschnitt 6.2 können wir zumindest für $p = 1$ und b hinreichend groß die Lücke um die asymptotische Normalität der eindimensionalen Randverteilungen des rLS/rIC-Filter-/Vorhersagefehlers schließen; wir vermuten stark, daß entsprechendes auch für die höherdimensionalen Modelle gilt.

Proposition 6.3.7 *Auch im idealen Modell (V3) sind für $p = 1$, $b_0 < b_t < \infty$ bzw. $\delta(b_0) < \delta < \text{tr} \Sigma_{t|t-1}^{\text{rLS}} / \text{tr} \Sigma_{t|t}^{\text{rLS}}$ mit*

$$b_0 := \sup\{b > 0 \mid 2 \frac{\mathbb{E}[\Delta \beta_t^{\text{rLS}} \Delta y_t^{\text{rLS}} w_b]}{\mathbb{E}[\Delta \beta_t^{\text{rLS}} \Delta y_t^{\text{rLS}}]} \leq \frac{\mathbb{E}[(\Delta y_t^{\text{rLS}})^2 w_b^2]}{\mathbb{E}[(\Delta y_t^{\text{rLS}})^2]}\}, \quad (6.23)$$

wobei

$$w_b := \min\left\{1, \frac{b}{|\tilde{M}_t^{\text{rLS}} \Delta y_t^{\text{rLS}}|}\right\}, \quad \tilde{M}_t^{\text{rLS}} = \frac{\mathbb{E}[\Delta \beta_t^{\text{rLS}} \Delta y_t^{\text{rLS}}]}{\mathbb{E}[(\Delta y_t^{\text{rLS}})^2]} \quad (6.24)$$

bereits die eindimensionalen Randverteilungen des rLS-Filterfehlers, resp. -Vorhersagefehlers nicht (asymptotisch) normalverteilt sind.

Bemerkung 6.3.8 Wegen der Stetigkeit von

$$u(b) := 2 \frac{\mathbb{E}[\Delta\beta_t^{\text{rLS}} \Delta y_t^{\text{rLS}} w_b]}{\mathbb{E}[\Delta\beta_t^{\text{rLS}} \Delta y_t^{\text{rLS}}]} - \frac{\mathbb{E}[(\Delta y_t^{\text{rLS}})^2 w_b^2]}{\mathbb{E}[(\Delta y_t^{\text{rLS}})^2]}$$

in $b \in [0, \infty]$ und weil $u(\infty) = 1$ ist $b_0 < \infty$.

BEWEIS: Wir weisen die Aussage nur für den Filterfehler nach; die Aussage für den Vorhersagefehler folgt mit dem Cramér-Lévy-Theorem.

Zunächst zeigen wir, daß

$$\text{Var } \Delta\beta_{t|t}^{\text{rLS}} = \Sigma_{t|t}^{\text{rLS}} = \mathbb{E}[(\Delta\beta^{\text{rLS}} - \tilde{M}\Delta y^{\text{rLS}} w_{b_t})^2] < \Sigma_{t|t-1}^{\text{rLS}} = S = \mathbb{E}[(\Delta\beta^{\text{rLS}})^2].$$

Dazu müssen wir zeigen, daß

$$2 \mathbb{E}[\tilde{M}_t^{\text{rLS}} \Delta\beta_t^{\text{rLS}} \Delta y_t^{\text{rLS}} w_{b_t}] > (\tilde{M}_t^{\text{rLS}})^2 \mathbb{E}[(\Delta y_t^{\text{rLS}})^2 w_{b_t}^2];$$

dies aber ist für $b_t > b_0$ nach (23) wegen der Definition von \tilde{M}_t^{rLS} der Fall. Andererseits dominiert aber in Proposition 6.2.3 für $u \rightarrow -\infty$ der Term

$$\frac{1}{\sqrt{S}} [1 - \Phi(\frac{MZu - (1 - MZ)b}{\sqrt{V}|M|})] \varphi(\frac{u+b}{\sqrt{S}}),$$

für $u \rightarrow \infty$ der Term

$$\frac{1}{\sqrt{S}} [1 - \Phi(\frac{MZu + (1 - MZ)b}{\sqrt{V}|M|})] \varphi(\frac{u-b}{\sqrt{S}})$$

die jeweils übrigen exponentiell, so daß die Schwänze der Dichte schließlich von der Größenordnung $\frac{c}{\sqrt{S}} \varphi(\frac{u}{\sqrt{S}})$ sind, so daß, wenn $\Delta\beta_{t|t}^{\text{rLS}}$ normalverteilt wäre, dann gemäß einer $\mathcal{N}(0, S)$ -Verteilung — anstelle einer $\mathcal{N}(0, \Sigma_{t|t})$ -Verteilung.

////

Kapitel 7

Stabilität und Stationarität der Filter

In diesem Kapitel behandeln wir die (asymptotische) Stabilität und Stationarität der bisher eingeführten Filter, rLS, rIC und mIC. Ausgangspunkt dieses Kapitels war die Diplomarbeit von Rausch (1999) und die dort erfolgten Untersuchungen über die Stationarität des klassischen Kalman-Filters.

7.1 Folgerungen aus Sylvesterschem Trägheitssatz für den Kalman-Filter

Zunächst halten wir t fest und unterdrücken diese Abhängigkeit.

Lemma 7.1.1 *Sei \hat{M} eine Matrix von der Bauart des Kalman-Gain, also $\hat{M} = \Sigma Z^\tau (Z \Sigma Z^\tau + V)^{-1}$ mit gewissen positiv (semi)definiten, symmetrischen Matrizen $\Sigma \in \mathbb{R}^{p \times p}$ und $V \in \mathbb{R}^{q \times q}$ und einer Matrix $Z \in \mathbb{R}^{q \times p}$ so daß $Z \Sigma Z^\tau + V$ vollen Rang hat; dann gilt*

- (a) $\hat{M}Z$ ist diagonalisierbar.
- (b) $\hat{M}Z \succeq 0$
- (c) Ist $\Sigma \succ 0$, so $\mathbb{I}_p - \hat{M}Z \succeq 0$, mit “ \succ ” falls $V \succ 0$.
- (d) Ist $\Sigma \succ 0$, so $\rho(\mathbb{I}_p - \hat{M}Z) \leq 1$.

BEWEIS: Wir schreiben $A \sim B$, falls sich A in B durch Konjugation mit $T \in \text{GL}(p)$ überführen läßt, also $B = T^{-1}AT$. Die Diagonalisierbarkeit einer Matrix $A \in \mathbb{R}^{p \times p}$ bleibt erhalten unter Konjugation mit $T \in \text{GL}(p)$. Damit folgt (a), denn es gilt

$$\hat{M}Z \sim \Sigma^{\frac{1}{2}} Z^\tau (Z \Sigma Z^\tau + V)^{-1} Z \Sigma^{\frac{1}{2}} \quad (7.1)$$

zusammen mit Korollar C.1.6 folgt damit aber auch (b) Weiter gilt

$$\begin{aligned} & (Z \Sigma Z^\tau + V) \succ Z \Sigma Z^\tau & (7.2) \\ \Rightarrow & (Z \Sigma Z^\tau + V)^{-1} \prec (Z \Sigma Z^\tau)^{-1} \\ \Rightarrow & \Sigma^{\frac{1}{2}} Z^\tau (Z \Sigma Z^\tau + V)^{-1} Z \Sigma^{\frac{1}{2}} \prec \Sigma^{\frac{1}{2}} Z^\tau (Z \Sigma Z^\tau)^{-1} Z \Sigma^{\frac{1}{2}} \end{aligned}$$

7.1. FOLGERUNGEN AUS SYLVESTERSCHEM TRÄGHEITSSATZ FÜR DEN KALMAN-FILTER 91

Sei $y \in \mathbb{R}^p$. Dann besitzt y die Darstellung $y = y_0 + y_1$ mit $y_0 \in \ker(Z\Sigma^{\frac{1}{2}})$, $y_1 \in \ker(Z\Sigma^{\frac{1}{2}})^\perp$. Nach dem Isomorphiesatz gilt $\ker(Z\Sigma^{\frac{1}{2}})^\perp \cong \Sigma^{\frac{1}{2}}Z^\tau(\mathbb{R}^p)$, also existiert ein $x_1 \in \mathbb{R}^q$ mit $y_1 = \Sigma^{\frac{1}{2}}Z^\tau x_1$. Daher gilt

$$\begin{aligned} & y^\tau \Sigma^{\frac{1}{2}} Z^\tau (Z\Sigma Z^\tau + V)^{-1} Z\Sigma^{\frac{1}{2}} y \preceq & (7.3) \\ & \preceq y^\tau \Sigma^{\frac{1}{2}} Z^\tau (Z\Sigma Z^\tau)^{-1} Z\Sigma^{\frac{1}{2}} y \\ & = y_1^\tau \Sigma^{\frac{1}{2}} Z^\tau (Z\Sigma Z^\tau)^{-1} Z\Sigma^{\frac{1}{2}} y_1 = \\ & = x_1^\tau Z\Sigma Z^\tau (Z\Sigma Z^\tau)^{-1} Z\Sigma Z^\tau x_1 = \\ & = x_1^\tau Z\Sigma Z^\tau x_1 = \|y_1\|^2 \leq \|y\|^2 \end{aligned}$$

Daher ist

$$\Sigma^{\frac{1}{2}} Z^\tau (Z\Sigma Z^\tau + V)^{-1} Z\Sigma^{\frac{1}{2}} \prec \mathbb{I}_p \quad (7.4)$$

und damit folgt durch Konjugation mit $\Sigma^{-\frac{1}{2}}$ und Lemma C.1.6 der erste Teil der Behauptung (c), wobei in (3) “ \prec ” gilt, falls $V \succ 0$, woraus der zweite folgt. (d) folgt mit Theorem C.1.3 aus (4) durch Konjugation mit $\Sigma^{\frac{1}{2}}$:

$$0 \prec \mathbb{I}_p - \Sigma^{\frac{1}{2}} Z^\tau (Z\Sigma Z^\tau + V)^{-1} Z\Sigma^{\frac{1}{2}} \iff 0 \prec \mathbb{I}_p - \hat{M}Z$$

////

Lemma 7.1.2 Sei zu $b > 0$, $\tilde{M} = \tilde{M}(y) := \hat{M} \min\{1, \frac{b}{|\hat{M}y|}\}$. Dann sind für \tilde{M} anstelle von \hat{M} punktweise in y die Aussagen von Lemma 7.1.1 in Kraft. Gilt für $\hat{M}Z$ in Aussage (b) “ \succ ”, so auch für $\tilde{M}Z$.

BEWEIS: Für festes y ist $\tilde{M}Z$ als skalares Vielfaches von $\hat{M}Z$ diagonalisierbar, also (a). Da der skalare Faktor $w = \min\{1, \frac{b}{|\hat{M}y|}\} > 0$ ist, folgt (b). Es gilt

$$\hat{M}Z - \tilde{M}Z = (1 - \frac{b}{|\hat{M}y|})_+ \hat{M}Z \succ 0,$$

so daß

$$\hat{\mathbb{I}}_p - \tilde{M}Z = \mathbb{I}_p - \hat{M}Z + (1 - \frac{b}{|\hat{M}y|})_+ \hat{M}Z \succ \mathbb{I}_p - \hat{M}Z,$$

also (c). (d) folgt wie in Lemma 7.1.1.

////

Bisher kranken die Ergebnisse noch an der Frage, wie für die Matrizen des Kalman-Filters in Aussage (b) von Lemma 7.1.1 “ \succ ” erzwungen werden kann. Für den Fall $\text{rk } Z^\tau V^{-1} Z = \text{rk } Q = p$ folgt dies allerdings sofort. Mit Induktion sieht man, daß damit auch $\text{rk } \Sigma_{t|t-1}^{\text{KK}} = p$ und damit $\mathbb{I}_p - M_t^{\text{KK}} Z$ geschrieben werden kann als $\Sigma_{t|t}^{\text{KK}} \Sigma_{t|t-1}^{\text{KK}^{-1}}$, was aber dann vollen Rang hat.

Immer noch nicht klar ist allerdings, ob auch für F mit mindestens einem Eigenwert λ vom Betrag größer oder gleich 1 erreicht werden kann, daß alle Eigenwerte von $F(\mathbb{I}_p - MZ)$ einen Betrag kleiner 1 haben. Hierzu benötigen wir einige Resultate aus der Kontrolltheorie.

7.2 Begriffe aus der Kontrolltheorie für den Kalman-Filter

Der folgende Abschnitt stellt die wesentlichen Begriffe aus der Kontroll- und Lyapunovtheorie diskreter dynamischer Modelle zusammen. Er ist weitestgehend Standard und ist in dieser Darstellung dem Anhang zu Anderson und Moore (1979) entnommen.

7.2.1 Lyapunov-Stabilität

Bei der Untersuchung der asymptotischen Eigenschaften unserer Filter und Glätter erweisen sich die Begriffsbildungen aus der Lyapunov-Theorie für diskrete, dynamische Systeme als nützlich.

Dazu wollen wir das diskrete dynamische System

$$x_0 = a_0, \quad x_k = f_k(x_{k-1}), \quad k \in \mathbb{N} \quad (7.5)$$

mit $a_0 \in \mathbb{R}^p$, $f_k : \mathbb{R}^p \rightarrow \mathbb{R}^p$ bekannt betrachten, wobei $f_k(0) = 0$ für alle k , so daß $x_e := 0$ ein Gleichgewichtszustand ist, also $f_k(x_e) = x_e$.

Definition 7.2.1 Ein Gleichgewichtszustand x_e heißt stabil, falls zu beliebigen k_0 , $\varepsilon > 0$ ein $\delta = \delta(\varepsilon, k_0)$ existiert, so daß aus $|x_{k_0} - x_e| < \delta$ folgt $|x_k - x_e| < \varepsilon$ für alle $k \geq k_0$.

Definition 7.2.2 Ein Gleichgewichtszustand x_e heißt asymptotisch stabil, falls er stabil ist und zu beliebigen k_0 ein $\delta = \delta(k_0)$ existiert, so daß aus $|x_{k_0} - x_e| < \delta$ folgt $\lim_k |x_k - x_e| = 0$.

Er heißt global asymptotisch stabil, falls δ unabhängig von k_0 und beliebig groß gewählt werden kann.

Er heißt exponentiell asymptotisch stabil, falls mit einem $\alpha = \alpha(k_0) > 0$ und $0 < \rho < 1$ gilt

$$|x_k - x_e| < \alpha \rho^{(k-k_0)}.$$

Definition 7.2.3 Ein Gleichgewichtszustand x_e heißt beschränkt oder auch Lagrange-stabil, falls zu k_0 und x_{k_0} ein $\delta = \delta(k_0, x_{k_0})$ existiert, so daß $|x_k| \leq \delta$ für alle $k \geq k_0$.

Definition 7.2.4 Eine Funktion $V : \mathbb{R}^p \rightarrow \mathbb{R}$ heie Lyapunov-positiv-(semi)-definit¹, geschrieben $V \succ 0$ [$V \succeq 0$], falls $V(x) \geq 0$ für alle $x \in \mathbb{R}^p$, $V(0) = 0$ und $V(x) = 0 \implies x = 0$ — letzteres nicht im Fall "semi".

Eine Funktion $V : \mathbb{R}^p \rightarrow \mathbb{R}$ heie Lyapunov-negativ-(semi)definit, geschrieben $V \prec 0$ ($V \preceq 0$), falls $-V \succ 0$ [$-V \succeq 0$].

Zunächst betrachten wir zeitinvariante Systeme, also $f_k \equiv f$. Für zwei aufeinanderfolgende Zustände x_{k+1}, x_k setzen wir

$$\Delta V(x_k) := V(x_{k+1}) - V(x_k).$$

Weiterhin steht im folgenden S für ein abgeschlossenes, beschränktes Gebiet des \mathbb{R}^p , $0 \in S$. Dann gelten folgende Theoreme

¹Das Präfix "Lyapunov-" wurde vom Autor dieser Arbeit zur Unterscheidung des Begriffs positiv (semi)definit für Funktionen $V : \mathbb{N}$ oder $\mathbb{R} \rightarrow \mathbb{C}$ im Zusammenhang mit den Sätzen von Herglotz und Bochner hinzugefügt.

Theorem 7.2.5 [Stabilität] Falls es für ein S eine Funktion $V : \mathbb{R}^p \rightarrow \mathbb{R}$ gibt mit $V|_S \succ 0$ und $\Delta V \preceq 0$ längs Trajektorien x_k gemäß (5), dann ist $x_e = 0$ stabil.

Theorem 7.2.6 [as. Stabilität] Falls es für ein S eine Funktion $V : \mathbb{R}^p \rightarrow \mathbb{R}$ gibt mit $V|_S \succ 0$ und $\Delta V|_S \prec 0$, dann ist $x_e = 0$ asymptotisch stabil.

Theorem 7.2.7 [as. Stabilität] Falls es für ein S eine Funktion $V : \mathbb{R}^p \rightarrow \mathbb{R}$ gibt mit $V|_S \succ 0$ und $\Delta V|_S \preceq 0$, und $\Delta V(x_k) = 0$ für alle k impliziert, daß $x_k \equiv 0$, dann ist $x_e = 0$ asymptotisch stabil.

Theorem 7.2.8 [globale as. Stabilität] Falls in den Theoremen 7.2.6 oder 7.2.7 $S = \mathbb{R}^p$ und falls $V(x) \rightarrow \infty$ für $|x| \rightarrow \infty$, dann ist $x_e = 0$ global asymptotisch stabil.

Eine Funktion V , mit der ein Stabilitätsresultat mithilfe der Theoreme 7.2.5–7.2.8 nachgewiesen werden kann, heißt *Lyapunovfunktion*.

7.2.2 Zustandsraummodelle in der Kontrolltheorie

Wie eingangs in Abschnitt 1.1.3 erwähnt, werden in der Kontrolltheorie ebenfalls Zustandsraummodelle betrachtet, vgl. auch Kailath (1980). Im einfachsten Fall verwendet man dabei folgende

Definition 7.2.9

$$\beta_0 = a_0, \quad (7.6)$$

$$\beta_t = F_t \beta_{t-1} + G_t v_t, \quad t \in \mathbb{N}, \quad (7.7)$$

$$y_t = Z_t \beta_t, \quad t \in \mathbb{N}, \quad (7.8)$$

$$y_t \in \mathbb{R}^q, a_0, \beta_t \in \mathbb{R}^p, v_t \in \mathbb{R}^r, F_t \in \mathbb{R}^{p \times p}, G_t \in \mathbb{R}^{p \times r}, Z_t \in \mathbb{R}^{q \times p}$$

Im Gegensatz zu der stochastischen Variante (2) und (1) werden hier die Innovationen v_t als exogen beeinflussbare Größen interpretiert, mit denen man den Prozeß der β_t steuert.

Offenbar ist für $v_t \equiv 0$ das System aus Definition 7.2.9 asymptotisch stabil im Sinn von Definition 7.2.2, falls $|F| < 1$.

Dies ist aber für Matrizen, die sich nicht durch Konjugation mit einer orthogonalen (unitären) Matrix diagonalisieren lassen, nicht notwendig; eine notwendige Bedingung hierfür ist, daß $\rho(F) < 1$ gilt.

Beschränkte Eingangsgrößen v_t haben dann auch beschränkte Ausgangsgrößen β_t und y_t zur Folge.

7.2.3 Kontrolltheorie im zeitinvarianten Fall

Zunächst geben wir die Definitionen aus Anderson und Moore (1979) im zeitinvarianten Fall, d.h. $F_t \equiv F$, $G_t \equiv G$, $Z_t \equiv Z$ wieder, wobei wir mit $\sigma(A)$ die Menge der Eigenwerte von A bezeichnen:

Definition 7.2.10 Das Paar (F, G) heißt vollständig erreichbar, falls eine der folgenden äquivalenten Bedingungen zutrifft:

- (a) $\text{rk}[G, FG, \dots, F^{p-1}G] = p$
- (b) $w^\tau F^i G = 0, i = 0, 1, \dots, p-1 \implies w = 0$
- (c) $w^\tau G = 0, w^\tau F = \lambda w^\tau$ für ein $\lambda \in \mathbb{C} \implies w = 0$
- (d) $\forall \lambda \in \mathbb{C} \exists K = K(\lambda) \in \mathbb{R}^{p \times r} : \lambda \in \sigma(F + GK^\tau)$
- (e) Zu $a_0 = 0$ in (6), $a_p \in \mathbb{R}^p \exists v_0, \dots, v_{p-1} \in \mathbb{R}^p : \beta_p = a_p$.

Definition 7.2.11 Das Paar (F, G) heißt vollständig kontrollierbar, falls eine der folgenden äquivalenten Bedingungen zutrifft:

- (a) $[G, FG, \dots, F^{p-1}G](\mathbb{R}^r) \supset F^p(\mathbb{R}^p)$
- (b) $w^\tau F^i G = 0, i = 0, 1, \dots, p-1 \implies w^\tau F^p = 0$
- (c) $w^\tau G = 0, w^\tau F = \lambda w^\tau$ für ein $\lambda \in \mathbb{C} \implies \lambda w = 0$
- (d) Zu $a_0 \in \mathbb{R}^p$ in (6), $\exists v_0, \dots, v_{p-1} \in \mathbb{R}^p : \beta_p = 0$.

Definition 7.2.12 Das Paar (F, G) heißt vollständig stabilisierbar, falls eine der folgenden äquivalenten Bedingungen zutrifft:

- (a) $w^\tau G = 0, w^\tau F = \lambda w^\tau$ für ein $\lambda \in \mathbb{C} \implies |\lambda| < 1$ oder $w = 0$
- (b) $\exists K \in \mathbb{R}^{p \times r} : |\lambda| < 1 \quad \forall \lambda \in \sigma(F + GK^\tau)$
- (c) Falls $G = [G_1^\tau, 0]^\tau$ mit $G_1 \in \mathbb{R}^{p' \times p}$ und es ein $T \in \text{GL}(p)$ gibt mit

$$TFT^{-1} = \begin{bmatrix} F_{11} & F_{12} \\ 0 & F_{22} \end{bmatrix}, \quad F_{11} \in \mathbb{R}^{p' \times p'}$$

und $[F_{11}, G_1]$ vollständig kontrollierbar, dann muß $|\lambda| < 1 \forall \lambda \in \sigma(F_{22})$ sein.

Definition 7.2.13 Das Paar (F, Z^τ) heißt

- (a) vollständig beobachtbar,
- (b) vollständig konstruierbar,
- (c) vollständig ermittelbar,

falls (F^τ, Z)

- (a) vollständig erreichbar
- (b) vollständig kontrollierbar
- (c) vollständig stabilisierbar

ist.

Bemerkung 7.2.14 Definitionen 7.2.10–7.2.12 beschäftigen sich mit der “Mächtigkeit” der Steuerung v_t :

Ist das System aus Definition 7.2.9 vollständig erreichbar, so kann ich mit meiner Steuerung v_t schließlich t jeden Zustand im \mathbb{R}^p erreichen. Wenn es vollständig kontrollierbar ist, kann ich jeden Zustand, den das System als Resultat der Transformation mit F annehmen kann, mit meiner Steuerung erreichen. Ist das System vollständig stabilisierbar, so kann ich mit der Steuerung v_t zumindest auf jede expansive Richtung im Zustand² reagieren, weil ich einen solchen auch erreiche.

Die Begriffe aus Definition 7.2.13 beschreiben die “Mächtigkeit” der Beobachtungstransformation Z :

Ist das System vollständig beobachtbar, so kann ich eindeutig jeden Punkt im \mathbb{R}^p als Zustand im Modell identifizieren. Wenn es vollständig konstruierbar ist, kann ich jeden Zustand, den das System als Resultat der Transformation mit F

²d.h. $F\beta = \lambda\beta, |\lambda| > 1$

annehmen kann, eindeutig identifizieren. Ist das System vollständig ermittelbar, so kann ich zumindest jede expansive Richtung im Zustand identifizieren und gegebenenfalls mit der Steuerung v_t darauf reagieren.

Bemerkung 7.2.15 Andere Definitionen zu diesen Begriffen finden sich in Chen (1985): Er nennt das stochastische Zustandsraummodell (1.2) und (1.1) *stochastisch beobachtbar (ermittelbar)*, falls es ein t_0 gibt, so daß

$$\lim_{\alpha \rightarrow \infty} \Sigma_{t_0|t_0}(\alpha \mathbb{I}_p) \quad \text{bzw.} \quad \lim_{\alpha \rightarrow \infty} \Sigma_{0|t_0}(\alpha \mathbb{I}_p)$$

existiert. Dabei ist $\Sigma_{i|t_0}(\alpha \mathbb{I}_p)$ die entsprechende Fehlerkovarianz für $Q_0 = \alpha \mathbb{I}_p$ und alle anderen Hyper-Parameter konstant.

Er weist nach, daß für $Q_t \equiv 0$, $V_t \equiv 0$ für $t > 0$ diese Begriffe mit denen der Kontrolltheorie zusammenfallen; allerdings ist nach unseren Definitionen der Begriff der Ermittelbarkeit durch den der Konstruierbarkeit zu ersetzen.

Definition 7.2.16 Die Funktion $z \mapsto W(z) = H(z\mathbb{I}_p - F)^{-1}G$ heißt *Transferfunktion des Systems aus Definition 7.2.9*.

Mit diesen Definitionen gelten folgende Sätze:

Theorem 7.2.17 Das System aus Definition 7.2.9 hat minimale Dimensionen p , q und r genau dann, wenn (F, Z^τ) vollständig ermittelbar und (F, G) vollständig beobachtbar ist.

Ein weiteres Tripel (F_2, G_2, Z_2) realisiert dieselbe Transferfunktion $W(z)$ genau dann, wenn es ein $T \in \text{GL}(p)$ gibt mit $F_2 = T^{-1}FT$, $Z_2 = ZT$, $G_2 = T^{-1}G$.

BEWEIS: Anderson und Moore (1979). ///

Theorem 7.2.18 Sei ein zeitinvariantes Zustandsraummodell gemäß (2) und (1) gegeben, also $F_t = F$ und $Z_t = F$ für alle t . Weiter sei die Verteilungsannahme (V2) mit $Q_t = Q$ und $V_t = Q$ für alle t erfüllt.

(a) Genau dann konvergieren die Fehlerkovarianzen $\Sigma_{t|t-1}^{\text{KK}}$ für $t \rightarrow \infty$ gegen ein Grenzelement Σ^{KK} , und damit auch $M_t^{\text{KK}} \rightarrow M^{\text{KK}}$, $\Sigma_{t|t}^{\text{KK}} \rightarrow T^{\text{KK}}$, wenn (F, Z^τ) vollständig ermittelbar ist.

(b) Der Fehler des klassischen Kalman-Filters, $\beta_t - \beta_{t|t}^{\text{KK}}$ gemäß (12), ist genau dann schwach asymptotisch stationär, wenn (F, Z^τ) vollständig ermittelbar ist.

(c) Ist zusätzlich der $\rho(F) < 1$, so ist dann auch der klassische Kalman-Filter, $\beta_{t|t}^{\text{KK}}$ gemäß (12), schwach asymptotisch stationär.

(d) Sind (F, Z^τ) vollständig ermittelbar, so ist der Filterfehler des klassischen Kalman-Filters (12) genau dann exponentiell asymptotisch stabil im Sinn von Definition 7.2.2, wenn $(F, Q^{\frac{1}{2}})$ vollständig stabilisierbar ist.

(e) Sind die Fehlerterme v_t , ε_t (strikt) stationär, so ist der Fehler des klassischen Kalman-Filters, $\beta_t - \beta_{t|t}^{\text{KK}}$, (strikt) asymptotisch stationär, und falls zusätzlich $\rho(F) < 1$, so ist dies dann auch der klassische Kalman-Filter, $\beta_{t|t}^{\text{KK}}$.

BEWEIS: (a), (b) und (d) Anderson und Moore (1979). (c): Wir schreiben

$$\begin{pmatrix} \beta_{t+1} \\ \beta_{t+1} - \beta_{t+1|t+1}^{\text{KK}} \end{pmatrix} = \begin{pmatrix} F & 0 \\ 0 & (\mathbb{I}_p - M_t^{\text{KK}}Z)F \end{pmatrix} \begin{pmatrix} \beta_t \\ \beta_t - \beta_{t|t}^{\text{KK}} \end{pmatrix} + \tilde{\eta}_t,$$

$$\tilde{\eta}_t = \begin{pmatrix} 0 \\ -M_t^{\text{KK}} \end{pmatrix} \varepsilon_t + \begin{pmatrix} \mathbb{I}_p \\ (\mathbb{I}_p - M_t^{\text{KK}} Z) \end{pmatrix} v_t$$

Offenbar ist dieses System wegen der Unkorreliertheit der Komponenten entkoppelbar, und daher schwach asymptotisch stationär genau dann, wenn es jede der Komponenten ist. Dies ist durch die Voraussetzungen gewährleistet. Da $\beta_{t|t}^{\text{KK}} = \beta_t - (\beta_t - \beta_{t|t}^{\text{KK}})$ folgt somit die Behauptung aus der gemeinsamen schwachen asymptotischen Stationarität von $(\beta_t, \beta_t - \beta_{t|t}^{\text{KK}})$.

(e): Schreiben wir nun

$$\Delta\beta_{t+1}^{\text{KK}} = F(\mathbb{I}_p - M_t^{\text{KK}} Z)\Delta\beta_t^{\text{KK}} + \eta_t, \quad \eta_t = v_t - FM_t^{\text{KK}}\varepsilon_t,$$

so erkennen wir einen VAR(1)–Prozeß mit strikt asymptotisch stationären Innovationen η_t , dessen Koeffizient $F(\mathbb{I}_p - M_t^{\text{KK}} Z)$ gegen $F(\mathbb{I}_p - M^{\text{KK}} Z)$ geht. Nun folgt aber aus der vollständigen Ermittelbarkeit, daß $\rho(F(\mathbb{I}_p - M^{\text{KK}} Z)) < 1$ und somit folgt die Behauptung für $\text{Cov}[\Delta\beta_t^{\text{KK}}]$, und damit auch für $\text{Cov}[\Delta\beta_{t|t}^{\text{KK}}] = \text{Cov}[(\mathbb{I}_p - M_t^{\text{KK}} Z)\Delta\beta_t^{\text{KK}} - M_t^{\text{KK}}\varepsilon_t]$. Die gemischten Autokovarianzterme zu festem Lag m folgen wie für Lag $m = 0$, indem wir die Glättungsformeln (17) verwenden und einsehen, daß $\rho(J_s) = \rho(F(\mathbb{I}_p - M_s^{\text{KK}} Z))$. Die Behauptung für $\beta_{t|t}^{\text{KK}}$ folgt wie in (c). ////

Korollar 7.2.19 (a) *Wenn (F, Z^τ) vollständig ermittelbar ist, konvergieren die Fehlerkovarianzen der Kalman–Glätter $\Sigma_{t|t+k}^{\text{KK}}$, $k > 0$ für $t \rightarrow \infty$ gegen ein Grenzelement $\Sigma^{(k)}$, und läßt man anschließend $k \rightarrow \infty$ gehen, so konvergiert auch $\Sigma^{(k)}$ gegen ein beschränktes $\Sigma^{(\infty)}$.*

(b) *Der Fehler des klassischen Kalman–Glätters, $\beta_t - \beta_{t|t+k}^{\text{KK}}$, $k > 0$ gemäß (17), ist schwach asymptotisch stationär, wenn (F, Z^τ) vollständig ermittelbar ist.*

(c) *Ist zusätzlich $\rho(F) < 1$, so ist dann auch der klassische Kalman–Glätter, $\beta_{t|t+k}^{\text{KK}}$ gemäß (17), schwach asymptotisch stationär.*

(d) *Sind die Fehlerterme v_t, ε_t (strikt) stationär, so ist der Fehler des klassischen Kalman–Glätters, $\beta_t - \beta_{t|t+k}^{\text{KK}}$ (strikt) asymptotisch stationär, und falls zusätzlich $\rho(F) < 1$, so ist dies dann auch der klassische Kalman–Glätter, $\beta_{t|t+k}^{\text{KK}}$.*

BEWEIS: (a): Verwenden wir die Notation aus Lemma 10.5.2, so erhalten wir

$$\beta_{t|t+k}^{\text{KK}} = \beta_{t|t}^{\text{KK}} + \sum_{l=1}^k J_{t:t+k} M_{t+l}^{\text{KK}} \Delta y_{t+l}^{\text{KK}}. \quad (7.9)$$

Damit ergibt sich $\beta_t - \beta_{t|t+k}^{\text{KK}} = \beta_t - \beta_{t|t}^{\text{KK}} + \sum_{l=1}^k J_{t:t+l} M_{t+l}^{\text{KK}} \Delta y_{t+l}^{\text{KK}}$ und

$$\text{Cov}[\beta_t - \beta_{t|t+k}^{\text{KK}}] = \Sigma_{t|t}^{\text{KK}} - \sum_{l=1}^k J_{t:t+l} M_{t+l}^{\text{KK}} (\Delta_l^{\text{KK}}) (J_{t:t+l} M_{t+l}^{\text{KK}})^\tau$$

Für $t \rightarrow \infty$ gehen nun $\Sigma_{t|t}^{\text{KK}} \rightarrow T^{\text{KK}}$, $\Sigma_{t|t-1}^{\text{KK}} \rightarrow \Sigma^{\text{KK}}$, $M_t^{\text{KK}} \rightarrow M^{\text{KK}}$, $\Delta_t^{\text{KK}} \rightarrow \Delta^{\text{KK}}$ also $J_t = \Sigma_{t|t}^{\text{KK}} F^\tau (\Sigma_{t|t-1}^{\text{KK}})^{-1} \rightarrow T^{\text{KK}} F^\tau \Sigma^{\text{KK}} =: J$ und $J_{t:t+l} \rightarrow J^l$. Da wir nur endlich viele Summanden betrachten, gilt daher

$$\text{Cov}[\beta_t - \beta_{t|t+k}^{\text{KK}}] \xrightarrow{t \rightarrow \infty} T^{\text{KK}} - \sum_{l=1}^k J^l M^{\text{KK}} (\Delta^{\text{KK}}) (J^l M^{\text{KK}})^\tau.$$

Nach Theorem C.1.3 ist $\mathbb{I}_p - J^\tau = \mathbb{I}_p - \Sigma^{\text{KK}} F T^{\text{KK}}$ genau dann positiv definit, wenn nach Konjugation mit $(\Sigma^{\text{KK}})^{-1}$ auch $\mathbb{I}_p - F T^{\text{KK}} \Sigma^{\text{KK}}$ es ist, also wenn $\rho(F T^{\text{KK}} \Sigma^{\text{KK}}) < 1$. Dies ist aber gerade die exponentiell asymptotische Stabilität aus Theorem 7.2.18. Damit gilt

$$\lim_k \lim_t \text{Cov } \beta_t - \beta_{t|t+k}^{\text{KK}} = \Sigma^{(\infty)},$$

wobei wir mit den Operatoren aus (B.1) erhalten

$$\text{svec}[\Sigma^{(\infty)}] = \text{svec}[T^{\text{KK}}] - (M^{\text{KK}} \otimes \mathbb{I}_p)(\mathbb{I}_{p^2} - J \otimes J^\tau)^{-1}(\mathbb{I}_p \otimes M^{\text{KK}}) \text{svec}[\Delta^{\text{KK}}].$$

(b) folgt unmittelbar aus (a). Ist $\rho(F) < 1$, so ist (c) aus Theorem 7.2.18 in Kraft, und damit folgt (c) mit (9). (d) schließlich folgt wie die entsprechende Aussage in Theorem 7.2.18. ////

7.2.4 aus der Kontrolltheorie im allgemeinen Fall

Im folgenden müssen wir Definitionen 7.2.10–7.2.13 auf den zeitvariablen Fall verallgemeinern. Wir verwenden hierzu die Definitionen aus Schick (1989), wobei wir aber den Begriff “kontrollierbar”, um konsistent zu Definitionen 7.2.10–7.2.13 zu bleiben, durch “erreichbar” ersetzen.

Definition 7.2.20 Die Folge der Paare (F_t, Q_t) heißt vollständig kontrollierbar, falls es ein $s_1 \in \mathbb{N}$ gibt und ein $\alpha_1 > 0$, so daß für alle $t \in \mathbb{N}$

$$\sum_{i=t-s_1}^t F_{i+1:t}^\tau Q_i F_{i+1:t} \succ \alpha_1 \mathbb{I}_p \quad (7.10)$$

mit

$$F_{t:t-1} = \mathbb{I}_p, \quad F_{t:s} := F_{t:s-1} F_s, \quad s \geq t \quad (7.11)$$

Die Folge der Tripel (F_t, Z_t, V_t) heißt vollständig beobachtbar, falls es ein $s_2 \in \mathbb{N}$ gibt und ein $\alpha_2 > 0$, so daß für alle $t \in \mathbb{N}$

$$\sum_{i=t}^{t+s_2} F_{t:i-1} Z_i^\tau V_i^{-1} Z_i F_{t:i-1}^\tau \succ \alpha_2 \mathbb{I}_p \quad (7.12)$$

Bemerkung 7.2.21 Nach dem Satz von Cayley–Hamilton können wir ohne Einschränkung der Allgemeinheit $s_1 = s_2 = p$ wählen.

Bemerkung 7.2.22 Pagan (1980) verwendet als Definition für vollständige Beobachtbarkeit / Kontrollierbarkeit gleich die Beschränktheit in (10), (12) im positiv definiten Sinn durch $\beta_i \mathbb{I}_p$, wobei er dann von *gleichmäßig vollständig kontrollierbar / beobachtbar* (UCC / UCO) spricht; dabei beruft er sich auf Jazwinski (1970).

Mit diesen Begriffen gilt

Theorem 7.2.23 Seien die Folgen der Matrizen $F_t, Z_t, Q_t, V_t, V_t^{-1}$ beschränkt. Sei die Folge der Paare (F_t, Q_t) vollständig kontrollierbar und die Folge der Tripel (F_t, Z_t, V_t) vollständig beobachtbar.

Dann ist der klassische Kalman-Filter (12) schwach asymptotisch stationär und exponentiell asymptotisch stabil im Sinn von Definition 7.2.2.

Sind die Fehlerterme v_t, ε_t (strikt) stationär, so ist der klassische Kalman-Filter dann auch (strikt) asymptotisch stationär.

BEWEIS: Teil 1 der Aussage: Schick (1989), Teil 2 wie im Beweis zu Theorem 7.2.18. ////

Der Einfachheit halber wollen wir die Stabilitäts- und Stationaritätsaussagen für die nichtlinearen, robusten Filter und Glätter nur auf den zeitinvarianten Fall beschränken.

7.3 Begriffe aus der Theorie der Markovketten auf beliebigem Zustandsraum

Da wir es mit stochastischen Systemen zu tun haben, ist es sinnvoll die Begriffe aus der Kontrolltheorie gegebenenfalls zu modifizieren bzw. eigene Begriffe anzuwenden.

Als erstes führen wir Begriffe aus der Theorie der Markovketten³ auf beliebigen Zustandsräumen ein, die wir im wesentlichen dem Buch von Meyn und Tweedie (1993) entnommen haben.

7.3.1 robuste Verfahren als Realisation einer Markovkette

Ohne schon zu viel vorwegzunehmen halten wir fest, daß alle unsere rekursiven (bzw. m -rekursiven) Verfahren sich schreiben lassen als

$$\begin{aligned} X_0 &= x_0 \\ X_t &= \mathbf{G}_t(X_{t-1}, \eta_t), \quad t \geq 1 \end{aligned} \tag{7.13}$$

mit einer Folge von Lipschitz-stetigen Funktionen $\mathbf{G}_t : \mathbb{R}^s \times \mathbb{R}^r \rightarrow \mathbb{R}^s$, einer Folge unabhängig identisch verteilter \mathbb{R}^r -wertiger Zufallsgrößen η_t und einer \mathbb{R}^r -wertigen Zufallsgröße x_0 , die unabhängig von allen η_t ist.

Bemerkung 7.3.1 (a) Im Fall des rLS-Filters wird $s = p, r = p + q, \eta_t = (v_t^\tau, \varepsilon_{t-1}^\tau)^\tau, X_t = \Delta\beta_t^{\text{rLS}}$ sein und

$$\mathbf{G}_{t+1}(X_t, \eta_{t+1}) = F\Delta\beta_t^{\text{rLS}} + v_{t+1} - FM_t^{\text{rLS}}(Z\Delta\beta_t^{\text{rLS}} + \varepsilon_t) \min\left\{1, \frac{b_t}{|M_t^{\text{rLS}}(Z\Delta\beta_t^{\text{rLS}} + \varepsilon_t)|}\right\}$$

(b) Im Fall des rIC-Filters wird $s = 2p, r = p + q$ sein, und

$$\mathbf{G}_{t+1}(X_t, \eta_{t+1}) = \tilde{\mathbf{F}}_{t+1}X_t + \eta_{t+1},$$

mit den in (29) — (31) eingeführten Größen.

Definition 7.3.2 (a) Ein stochastischer Prozeß $\{X_t\}_{t \in \mathbb{N}}, X_t : (\Omega, \mathcal{A}) \rightarrow (\Xi, \mathcal{B})$, mit Zustandsraum (Ξ, \mathcal{B}) , definiert auf dem Wahrscheinlichkeitsraum

³also diskreter Zeitbereich

(Ω, \mathcal{A}, P) , heie Markovkette, falls er die Markov-Eigenschaft besitzt:
Fr jedes t und jede X_t -mebare Funktion f_t gilt

$$\mathbb{E}[f|\sigma(\{X_s\}_{s<t})] = \mathbb{E}[f|X_{t-1}]. \quad (7.14)$$

(b) Eine Markovkette $\{X_t\}_{t \in \mathbb{N}}$ auf einem polnischen Zustandsraum Ξ , versehen mit der Borel- σ -Algebra $\mathcal{B}(\Xi)$ heit Feller-Markovkette, falls die bedingten Wahrscheinlichkeiten $P(X_t \in A | X_{t-1} = x_{t-1})$ sich durch Markovkerne $p_t : \Xi \times \mathcal{B}(\Xi) \rightarrow \mathbb{R}_{\geq 0}$ darstellen lassen, die alle schwach stetig in x sind, d.h. fr alle stetigen und beschrnkten Funktionen $f : \Xi \rightarrow \mathbb{R}$ und alle $x \in X$ gelte:

$$x_n \xrightarrow{n \rightarrow \infty} x \implies \int f(y) p_t(dy, x_n) \longrightarrow \int f(y) p_t(dy, x)$$

(c) Eine Feller-Markovkette heit (asymptotisch) homogen, falls p_t (asymptotisch) nicht von t abhngt.

Proposition 7.3.3 Mit der Identifikation $\xi_t := (X_t^\tau, \eta_{t+1}^\tau)^\tau$ in (13) erhalten wir eine Markovkette auf dem \mathbb{R}^{2s+r} .

BEWEIS: Da $\{\xi_t\}_{t \in \mathbb{N}}$ eine Markovkette ist, sehen wir, indem wir mit $\bar{\xi}_t$ die σ -Algebra bezeichnen, die durch ξ_s , $s \leq t$ erzeugt wird und uns berzeugen, da fr jedes t und jede ξ_{t+1} -mebare Funktion f (14) gilt:

$$\begin{aligned} \mathbb{E}[f|\bar{\xi}_t] &= \mathbb{E}[f(\xi_{t+1})|\bar{\xi}_t] = \mathbb{E}[f((X_{t+1}^\tau, \eta_{t+2}^\tau)^\tau)|\bar{\xi}_t] = \\ &= \mathbb{E}[f((\mathbf{G}_{t+1}(\xi_t), \eta_{t+2}^\tau)^\tau)|\bar{\xi}_t] = \mathbb{E}[f((\mathbf{G}_{t+1}(\xi_t), \eta_{t+2}^\tau)^\tau)|\xi_t] = \mathbb{E}[f|\xi_t]. \end{aligned}$$

Wegen der Stetigkeit von \mathbf{G}_t handelt es sich sogar um eine Feller-Markovkette.

////

7.3.2 Harris-Rekurrenz

Weil wir auch berabzhlbare Zustandsrume zulassen wollen, gengen die einfachen Begriffe wie Transienz, Rekurrenz, Irreduzibilitt aus der Theorie der Markovketten auf abzhلبarem Zustandsraum nicht (vgl. Doob (1953), V.2), um das asymptotische Verhalten von X_t fr $t \rightarrow \infty$ zu untersuchen, und wir bentigen allgemeinere Begriffe.

Als sehr ntzlich erweisen sich folgende Definitionen fr allgemeine, polnische Rume als Zustandsraum, die wir Meyn (1989) entnommen haben:

Wir betrachten eine Feller-Markovkette $\{X_t\}_{t \in \mathbb{N}}$, $X_t : (\Omega, \mathcal{A}) \rightarrow (\Xi, \mathcal{B}(\Xi))$, mit polnischem Zustandsraum Ξ , versehen mit der Borel- σ -Algebra $\mathcal{B}(\Xi)$, definiert auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) .

Zunchst wollen wir mit den Markovkernen p_t regulre Versionen der bedingten Verteilungen von $p(X_t \in A | X_{t-1} = x_{t-1})$ bezeichnen, sowie rekursiv

$$p^1(A, x) := p_1(A, x) \quad \text{und fr } t > 1 \quad p^{t+1}(A, x) = \int p_{t+1}(A, y) p^t(dy, x).$$

Schlielich sei fr $x \in \Xi$ und Ereignisse A aus $\mathcal{B}(\Xi)^{\otimes \mathbb{N}}$

$$P_x(A) := P(\{X_t\}_{t \geq 1} \in A | X_0 = x).$$

Mit dieser Notation kommen wir zu

Definition 7.3.4 Eine homogene Feller–Markovkette $\{X_t\}_{t \in \mathbb{N}}$ heißt

(a) φ -irreduzibel, falls es ein $A \in \mathcal{B}(\Xi)$, ein Wahrscheinlichkeitsmaß φ auf $\mathcal{B}(\Xi)$, ein $n_0 \in \mathbb{N}$ und ein $\lambda_0 > 0$ gibt mit den Eigenschaften

(i) $P_x(\{X_t \in A \text{ mindestens einmal}\}) > 0$ für alle $x \in \Xi$.

(ii) $\sum_{k=1}^{n_0} P_x(\{X_k \in E\}) \geq \lambda_0 \varphi(\{E\})$ für alle $x \in \Xi$ und alle $E \in \mathcal{B}(\Xi)$.

Kommt es auf das Maß φ nicht an, oder ist es aus dem Zusammenhang klar, so spricht man schlicht von Irreduzibilität.

(b) Harris–rekurrent, falls $\{X_t\}_{t \in \mathbb{N}}$ irreduzibel ist und für alle $x \in \Xi$ gilt

$$P_x(\{X_t \in A \text{ mindestens einmal}\}) = 1.$$

(c) positiv Harris–rekurrent, falls $\{X_t\}_{t \in \mathbb{N}}$ Harris–rekurrent ist und es höchstens ein invariantes Wahrscheinlichkeitsmaß π auf $(\Xi, \mathcal{B}(\Xi))$ gibt.

Dabei heißt ein σ -finites Maß π auf $(\Xi, \mathcal{B}(\Xi))$ invariant, falls

$$\pi(A) = \int p(A, y) \pi(dy) \quad \forall A \in \mathcal{B}(\Xi).$$

Eine Menge $A \in \mathcal{B}(\Xi)$ heißt absorbierend, falls $p(A, x) = 1$ für alle $x \in \Xi$. Für absorbierende A können wir die Einschränkung $\{X_t\}|_A$ betrachten. Ist $\{X_t\}|_A$ Harris–rekurrent, so heißt A Harris–Menge.

Falls $\{X_t\}$ irreduzibel ist, so heißt die Menge A in (a) petite; es gibt dann eine Periode $m \in \mathbb{N}$ und Ereignisse $E_1, \dots, E_m \in \mathcal{B}(\Xi)$, so daß

$$\int \mathbf{I}_{E_{(i+1) \bmod m}}(y) p(dy, \cdot) = \mathbf{I}_{E_i \bmod m}(\cdot), \quad \int_{E_i} p^m(dy, \cdot) = \mathbf{I}_{E_i}(\cdot).$$

Proposition 7.3.5 (a) Falls $\{X_t\}_{t \in \mathbb{N}}$ Harris–rekurrent ist, gibt es ein invariantes σ -finites Maß π auf $(\Xi, \mathcal{B}(\Xi))$.

(b) Falls $\{X_t\}_{t \in \mathbb{N}}$ Harris–rekurrent mit invariantem Maß π ist, so gilt für die E_i aus Definition 7.3.4 $\pi(\{\cup E_i\}^c) = 0$.

(c) Falls $\{X_t\}_{t \in \mathbb{N}}$ Harris–rekurrent und aperiodisch ($m = 1$) ist mit invariantem Wahrscheinlichkeitsmaß π , so gilt mit dem Totalvariationsabstand d_V für jedes Maß μ auf $(\Xi, \mathcal{B}(\Xi))$

$$\lim_{t \rightarrow \infty} d_V(\int p^t(\cdot, x) \mu(dx), \pi(\cdot)) = 0$$

(d) Ist $\{X_t\}_{t \in \mathbb{N}}$ irreduzibel mit einer offenen petiten Menge, so ist $\{X_t\}_{t \in \mathbb{N}}$

(i) Harris–rekurrent genau dann, falls es für jeden Anfangswert $x \in \Xi$ und jedes $\varepsilon > 0$ ein Kompaktum $K = K(x, \varepsilon) \subset \Xi$ gibt mit

$$P_x(\{X_t\} \in K \text{ nur endlich oft}) < \varepsilon. \tag{7.15}$$

(ii) positiv Harris–rekurrent genau dann, falls es für jeden Anfangswert $x \in \Xi$ und jedes $\varepsilon > 0$ ein Kompaktum $K = K(x, \varepsilon) \subset \Xi$ gibt mit

$$P_x(\{X_t\} \notin K \text{ unendlich oft}) < \varepsilon. \tag{7.16}$$

Offenbar ist dies nichts anderes als die Straffheit der Bildmaße $P_x^{X_t}$ von $\{X_t\}$ unter P_x zu jedem beliebigen $x \in \Xi$. Wir sagen dann, $\{X_t\}$ sei straff.

BEWEIS: Meyn (1989); dieser verweist dabei zum Teil auf Orey (1971), Nummelin (1984) und Athreya und Ney (1982). ////

7.3.3 Begriffe der stochastischen Kontrolltheorie

Bei der Untersuchung der asymptotischen L_2 -Theorie der linearen Verfahren waren die Begriffe der (deterministischen) Kontrolltheorie nützlich; für unsere nichtlinearen, robusten Verfahren benötigen wir leistungsfähigere Begriffe, wie wir sie aus der stochastischen Kontrolltheorie beziehen können.

Sie sind in dieser Darstellung Meyn und Caines (1991) entnommen⁴.

Definition 7.3.6 *Wir betrachten ein Markovsystem $\{X_t\}$ der Form (13). Zusätzlich seien die Übergangsfunktionen \mathbf{G}_t einmal stetig differenzierbar und $\mu_\eta := \mathcal{L}(\eta_t)$ besitze eine unterhalbstetige Lebesgue-Dichte $\nu(\eta)$.*

(a) *Man nennt einen Punkt $y \in \mathbb{R}^r$ erreichbar von $x \in \mathbb{R}^r$, falls es ein $T \in \mathbb{N}$ gibt und Realisationen $\eta_1^*, \dots, \eta_T^*$ von η_1, \dots, η_T , so daß $X_T = y$, $X_0 = x$.*

(b) *Falls für jedes $x \in \mathbb{R}^r$ die Menge der erreichbaren Punkte von x aus ein nichtleeres Inneres besitzt, heißt $\{X_t\}$ vorwärts akzessibel.*

(c) *$\{X_t\}$ heißt*

(i) *schwach stochastisch kontrollierbar, falls es zu jeder Anfangsbedingung $X_0 = x_0$ ein $T = T(x_0) \in \mathbb{N}$ gibt und eine offene Menge $\mathcal{O}_{x_0} \subset \mathbb{R}^r$, so daß auf \mathcal{O}_{x_0} gilt $p^T(dy, x_0) \gg \lambda(dy)$.*

(ii) *lokal stochastisch kontrollierbar, falls $\{X_t\}$ schwach stochastisch kontrollierbar ist und zusätzlich auf \mathcal{O}_{x_0} gilt $p^T(dy, x_0) \ll \lambda(dy)$.*

(d) *Die Funktion $\mathbf{G}_{0:k}^x : \mathbb{R}^{ks} \rightarrow \mathbb{R}^r$, definiert als*

$$\begin{aligned} \mathbf{G}_{0:0}^x &= x \\ \mathbf{G}_{0:k}^x &= \mathbf{G}_{0:k}^x(z_1, \dots, z_k) = \mathbf{G}_k(\mathbf{G}_{0:k-1}^x(z_1, \dots, z_{k-1}), z_k), \quad k \geq 1 \end{aligned} \quad (7.17)$$

heißt die zu (13) zugeordnete erweiterte Übergangsmatrix.

(e) *Die Matrix $\mathcal{C}_y^k = \mathcal{C}_y^k(\boldsymbol{\eta})$ definiert als*

$$\mathcal{C}_y^k = [A_{k-1} \cdots A_1 B_0, A_{k-1} \cdots A_2 B_1, \dots, A_{k-1} B_{k-2}, B_{k-1}] \quad (7.18)$$

mit

$$\begin{aligned} A_k &= A_k(y, \eta_1, \dots, \eta_{k+1}) = \left[\frac{\partial}{\partial X} \mathbf{G}_t \right]_{\mathbf{G}_{0:k, z_{k+1}}^x} \\ B_k &= B_k(y, \eta_1, \dots, \eta_{k+1}) = \left[\frac{\partial}{\partial \eta} \mathbf{G}_t \right]_{\mathbf{G}_{0:k, z_{k+1}}^x} \end{aligned}$$

heißt generalisierte Beobachtungs-Matrix.

Bemerkung 7.3.7 (a) Wir haben hier stochastische Beobachtbarkeit im Unterschied zu Meyn und Caines (1991) gleich für den inhomogenen Fall definiert. Für den Beweis der folgenden Propositionen spielt dies keine Rolle.

(b) Die bei der Definition zu erfüllende Glattheit der \mathbf{G}_t ist nur lokal um 0 nötig, um daraus Teil (b) des folgenden Theorems schließen zu können.

⁴Um konsistent mit den deterministischen Begriffen zu sein, müßten wir die von Meyn und Caines (1991) eingeführten Begriffe "lokal (schwach) stochastisch kontrollierbar" ersetzen durch "lokal (schwach) stochastisch erreichbar"; darauf wird aber im folgenden verzichtet.

Theorem 7.3.8 Sei $\{X_t\}$ ein Markovsystem gemäß (13), das zusätzlich noch die Bedingungen zu Definition 7.3.6 erfüllt. Weiter sei \mathcal{O}_η die Menge, auf der $\nu > 0$ ist⁵

(a) $\{X_t\}$ ist lokal stochastisch kontrollierbar genau dann, wenn es für jede Anfangsbedingung $x \in \mathbb{R}^r$ ein T gibt, so daß

$$\text{rk } \mathcal{C}_y^k(\eta) = r \quad \forall \eta \in \mathcal{O}_\eta^T \cap N_x^c, \quad \text{wobei } \lambda^{Ts}(N_x) = 0. \quad (7.19)$$

(b) $\{X_t\}$ ist schwach stochastisch kontrollierbar genau dann, wenn es für jede Anfangsbedingung $x \in \mathbb{R}^r$ ein T und ein $\eta \in \mathcal{O}_\eta^T$ gibt, so daß

$$\text{rk } \mathcal{C}_y^k(\eta) = r \quad (7.20)$$

(c) Falls (20) für ein $\eta \in \mathcal{O}_\eta^T$ gilt, so gibt es ein $c > 0$ und offene Mengen \mathcal{U}_x^η , \mathcal{V}_x^η die jeweils x und $\mathbf{G}_{0:k}^x(\eta)$ enthalten, so daß

$$p^T(A, y) \geq c\lambda^r(A \cap \mathcal{V}_x^\eta) \quad \forall A \in \mathbb{B}^r, y \in \mathcal{U}_x^\eta \quad (7.21)$$

BEWEIS: Meyn und Caines (1991) ////

Theorem 7.3.9 Sei $\{X_t\}$ ein Markovsystem gemäß (13), das zusätzlich noch die Bedingungen zu Definition 7.3.6 erfüllt, außerdem sei $0 \in \mathcal{O}_\eta$, $\{X_t\}$ schwach stochastisch kontrollierbar und straff, und $\mathbf{G}_t(0, 0) = 0$ für alle t .

Weiter gelte punktweise in (x, η) , daß $\mathbf{G}_t \rightarrow \mathbf{G}$, \mathbf{G} stetig in (x, η) in $\mathbb{R}^r \times \mathbb{R}^s$ und stetig differenzierbar in $(x, \eta) = (0, 0)$.

Zur Unterscheidung seien nun $p^t(\cdot, x)$ und $p_0^t(\cdot, x)$ die iterierten Markovkerne, die jeweils aus den \mathbf{G}_t -Übergängen bzw. aus \mathbf{G} -Übergängen hervorgehen.

Wenn nun $\lim_t d_V(p^t(\cdot, x), p_0^t(\cdot, x)) = 0$, so ist $\{X_t\}$ aperiodisch und positiv Harris rekurrent, und $\lim_t d_V(p^t(\cdot, x), \pi(\cdot)) = 0$ gleichmäßig für x auf kompakten Teilmengen des \mathbb{R}^r .

BEWEIS: der homogene Fall, das heißt $\mathbf{G}_t = \mathbf{G}$ für alle t wird in Meyn und Caines (1991) gezeigt. Der Rest folgt aus der Dreiecksungleichung für d_V und der Tatsache, daß Aperiodizität sich in π ablesen läßt, so daß auch das inhomogene System $\{X_t\}$ aperiodisch ist. ////

7.4 asymptotische Stabilität und Stationarität des rLS-Filters

Theorem 7.4.1 Sei ein zeitinvariantes Zustandsraummodell gemäß (1) und (2) gegeben, also $F_t = F$ und $Z_t = F$ für alle t . Weiter sei die Verteilungsannahme (V2) mit $Q_t = Q$ und $V_t = Q$ für alle t erfüllt. Sei (F, Z^τ) vollständig ermittelbar. Die Fehlerterme ε_t und die Innovationen v_t mögen strikt stationär sein und (im idealen Modell) (V2) genügen, und gelte zusätzlich⁶

$$Q \succ 0 \quad (7.22)$$

Dann gilt für die beiden vereinfachten rLS-Varianten,

⁵Wegen der Unterhalbstetigkeit von ν ist \mathcal{O}_η offen.

⁶also vollständige Erreichbarkeit und exponentiell asymptotische Stabilität von $\Delta\beta_{t|t}^{\text{rLS}}$

(a) der rLS-Filter, bei dem man anstelle der optimalen linearen Transformation M_t^{rLS} den Kalman-Gain M_t^{KK} verwendet

(b) der rLS-Filter, bei dem man von Anfang an nur $M^{\text{KK}} := \lim_t M_t^{\text{KK}}$ als lineare Transformation verwendet

wenn wir diese gemäß (5) bzw. (7) kalibrieren und die Clipping-Schwellen b_t "hinreichend schnell" gegen ein b konvergieren⁷, so gibt es ein $\delta_0 > 0$, so daß für alle $0 < \delta < \delta_0$ der Filterfehler $\Delta\beta_t$ strikt asymptotisch stationär und exponentiell asymptotisch stabil ist. Gilt weiter $|F| < 1$, so ist dann auch der Filter $\beta_{t|t}$, sowie die Vorhersage $\beta_{t|t-1}$ strikt asymptotisch stationär.

Sind die treibenden Prozesse v_t, ε_t zusätzlich ergodisch, so gilt dies auch für die Fehlerprozesse, und falls $|F| < 1$ auch für die Filter und Einschrittvorhersagen. Für den "original"-rLS-Filter, eingeführt in (4), kann zumindest die Existenz von Häufungspunkten der Verteilung nachgewiesen werden.

Bemerkung 7.4.2 Gilt $p = 1$ oder $|F| < 1$, so ist die vollständige Ermittlichkeit automatisch erfüllt, denn für $|F| < 1$ ist nach Lemma 7.1.1

$$\rho(F(\mathbb{I}_p - M^{\text{KK}} Z)) \leq |F| |(\mathbb{I}_p - M^{\text{KK}} Z)| < |\mathbb{I}_p - M^{\text{KK}} Z| \leq 1.$$

Für $p = 1$ kann man die Grenzwerte $\Sigma^{\text{KK}}, T^{\text{KK}} = \lim_t \Sigma_{t|t}^{\text{KK}}, M^{\text{KK}}, 1 - M^{\text{KK}} Z$ mithilfe

$$W := \sqrt{[QZ^2 + V(F^2 - 1)]^2 + 4QZ^2V}$$

explizit angeben als

$$\Sigma^{\text{KK}} = \frac{QZ^2 + V(F^2 - 1) + W}{2Z^2}, \quad (7.23)$$

$$T^{\text{KK}} = \frac{V QZ^2 + V(F^2 - 1) + W}{Z^2 QZ^2 + V(F^2 + 1) + W} \quad (7.24)$$

$$M^{\text{KK}} = \frac{1 QZ^2 + V(F^2 - 1) + W}{Z QZ^2 + V(F^2 + 1) + W} \quad (7.25)$$

$$(1 - M^{\text{KK}} Z) = \frac{2V}{QZ^2 + V(F^2 + 1) + W} \quad (7.26)$$

Es gilt $(1 - |F|)^2 \geq 0$ und somit $1 + F^2 \geq 2|F|$, also

$$|F| |1 - M^{\text{KK}} Z| = \frac{2|F|V}{QZ^2 + V(F^2 + 1) + W} < \frac{2|F|V}{V(F^2 + 1)} \leq 1.$$

Wir beweisen Theorem 7.4.1 zweimal: zunächst direkt (und allgemeiner) und dann mit Techniken aus Abschnitt 7.3 für absolutstetige, in der Zeit unabhängig identisch verteilte $(\varepsilon_t, v_t)_t$ mit unterhalbstetiger gemeinsamer Lebesgue-Dichte, die im Punkt 0 nicht verschwindet:

BEWEIS [1.VARIANTE]: Sei $\tilde{M}_t := M_t \min\{1, \frac{b_t}{|M_t \Delta y_t^{\text{rLS}}|}\}$, mit M_t einer Lineartransformation gemäß Varianten (a) und (b) des Theorems oder gemäß (4) der ursprünglichen Definition des rLS-Filters und $\Delta y_t^{\text{rLS}} = y_t - Z\beta_{t|t-1}^{\text{rLS}}$ mit $\beta_{t|t-1}^{\text{rLS}}$ aus der Filtervergangeheit des entsprechenden Filters.

⁷wie "schnell" wird in (27) im Beweis spezifiziert

Wir erinnern daran, daß nach (7) für die Kalibrierung der Filter gemäß (a) oder (b) aus dem Theorem bzw. (4) gilt:

$$\text{tr } \Sigma_{t|t}^{\text{rLS}} = \text{E}[\|\Delta\beta_t^{\text{rLS}} - M_t^{\text{KK}} \Delta y_t^{\text{rLS}}\|^2] \stackrel{!}{=} (1 + \delta) \text{tr } \tilde{\Sigma}_{t|t}^{\text{KK}}$$

Dabei entsteht $\tilde{\Sigma}_{t|t}^{\text{KK}}$ als Fehlerkovarianz des unter allen linearen Filtern optimalen Filters $\tilde{\beta}_{t|t}^{\text{KK}}$. Dieser wiederum wird ausgehend von einer Fehlerkovarianz $\tilde{\Sigma}_{t|t-1}^{\text{KK}}$ für $\Delta\tilde{\beta}_{t|t}^{\text{KK}}$ so bestimmt, daß dann $\tilde{\beta}_{t|t}^{\text{KK}} = \tilde{\beta}_{t|t-1} + \tilde{M}_t^{\text{KK}}(y_t - \tilde{\beta}_{t|t-1}^{\text{KK}})$ mit $\tilde{M}_t = \tilde{\Sigma}_{t|t-1}^{\text{KK}} Z^\tau (Z \tilde{\Sigma}_{t|t-1}^{\text{KK}} Z^\tau + V)^{-1}$. $\tilde{\Sigma}_{t|t-1}^{\text{KK}}$ schließlich ergibt sich als diejenige Kovarianz, die aus der Fortschreibung eines idealen Systems (\tilde{F}, Z, Q, V) mit $\tilde{F} = (1 + \delta)^{\frac{1}{2}} F$ hervorgeht.

Offenbar folgt nun aus der vollständigen Ermittelbarkeit, daß für $0 < \delta < \rho^{-1}(F(\mathbb{I}_p - M^{\text{KK}} Z)) - 1$ auch das Paar $[(1 + \delta)^{\frac{1}{2}} F, Z]$ ermittelbar ist, und somit konvergieren für ein solches δ nach Theorem 7.2.18 $\tilde{\Sigma}_{t|t-1}^{\text{KK}}$ und $\tilde{\Sigma}_{t|t}^{\text{KK}}$. Also sind insbesondere $\Sigma_{t|t-1}^{\text{rLS}}$ beschränkt, und $\text{tr } \Sigma_{t|t}^{\text{rLS}}$ konvergiert, und in jedem Fall sind $\Delta\beta_t^{\text{rLS}}$ und $\beta_t - \beta_{t|t}^{\text{rLS}}$ straff.

Für Dimension $p = 1$ haben wir damit schon die asymptotische Varianzstationarität für $\Delta\beta$.

Als nächstes zeigen wir die exponentielle asymptotische Stabilität:

Wie in Anderson und Moore (1979), p. 80 schließen wir mit einem Widerspruchsbeweis: Wir nehmen an, daß längs einer Realisation des Prozesses

$$\limsup_t \rho(\tilde{F}_t) \geq 1$$

für $\tilde{F}_t := F(\mathbb{I}_p - \tilde{M}_t Z)$. Nach Übergang zu einer Teilfolge, dürfen wir annehmen, daß $\Sigma_{t|t-1}^{\text{rLS}} \rightarrow \Sigma$, $\tilde{M}_t \rightarrow \tilde{M}$ und somit $\tilde{F}_t \rightarrow \tilde{F} = \lim_t F(\mathbb{I}_p - \tilde{M} Z)$, also

$$\Sigma^{\text{rLS}} = \tilde{F} \Sigma^{\text{rLS}} \tilde{F} + \tilde{M} V \tilde{M} + Q.$$

Wegen der stetigen Abhängigkeit der Erwartungswerte von den Matrixeinträgen folgt, daß auch $\rho(\tilde{F}) \geq 1$, und für einen Eigenvektor $0 \neq \omega \in \mathbb{C}^p$ zum Eigenwert $\lambda \in \mathbb{C}$, $|\lambda| > 1$ haben wir

$$0 \geq (1 - |\lambda|^2) \omega^H \Sigma \omega = \omega^H \tilde{M} V \tilde{M}^\tau \omega + \omega^H Q \omega \leq 0,$$

also $Q\omega = 0$ ein Widerspruch zu (22). Also ist $\limsup_t \rho(\tilde{F}_t) \leq \tilde{\rho} < 1$. Damit aber geht $\tilde{F}_{t:1}$, definiert als

$$\tilde{F}_{1:1} = \tilde{F}_1, \quad \tilde{F}_{t:1} = \tilde{F}_t \tilde{F}_{t-1:0}$$

exponentiell schnell gegen 0, und weil wir mit $\eta_k := v_{k+1} - F \tilde{M}_k \varepsilon_k$ schreiben können

$$\Delta\beta_{t+1} = \tilde{F}_{t:1} \Delta\beta_1 + \sum_{k=2}^t \tilde{F}_{t:k} \eta_k,$$

folgt hieraus die behauptete exponentielle asymptotische Stabilität.

Wegen der Straffheit der $\Delta\beta_t$ folgt aus dem Satz von Prokhorov, Theorem A.1.2, die Existenz schwacher Häufungspunkte. Wir zeigen nun zunächst, daß es unter der Annahme, daß $M_t^{\text{KK}} \equiv M^{\text{KK}}$ und $b_t \equiv b > 0$, nur einen solchen geben kann:

Sei P_t das Bildmaß von $\Delta\beta_t$. Angenommen $P \neq Q$ seien zwei Häufungspunkte.

Dann gibt es ein $\varepsilon > 0$ und ein Ereignis $A \in \mathbb{B}^p$, so daß $|P(A) - Q(A)| > 8\varepsilon$ und $P(\partial A) + Q(\partial A) = 0$. Seien P_{t_l} und P_{t_m} Teilfolgen, die schwach gegen P respektive Q konvergieren.

Sei weiter $K \subset \mathbb{R}^p$ ein Straffheitskompaktum für P_t (inklusive der Startverteilung von $\Delta\beta_0$). Wegen der asymptotischen Stabilität gibt es aber ein $t_0 \in \mathbb{N}$, so daß für alle Realisationen von $\Delta\beta_0$ aus $K \cap N^c$, N eine Nullmenge, und alle $t > t_0$ gilt $|\tilde{F}_{t_0} \Delta\beta_0| < \delta/2$. Dieses δ wählen wir so klein, daß für

$$A^\delta := \{x \in \mathbb{R}^p : \inf_{y \in A} |y - x| \leq \delta\}$$

gilt $P(A^\delta) - P(A) < \varepsilon$ und $P(\partial[A^\delta]) = 0$.

Seien nun l, m so groß, daß $t_l, t_m - t_l > t_0$ und daß

$$|P(A) - P_{t_l}(A)|, |Q(A) - P_{t_m}(A)|, |P(A^\delta) - P_{t_l}(A^\delta)| < \varepsilon.$$

Insbesondere gilt $|P_{t_m}(A) - P_{t_l}(A)| > 6\varepsilon$ und $|P_{t_l}(A^\delta) - P_{t_l}(A)| < 3\varepsilon$. Bezeichnen wir $\Delta\beta_t(\Delta\beta_0)$ als die Zufallsvariable, die sich ausgehend von $\Delta\beta_0$ entwickelt, so gilt:

$$\begin{aligned} 6\varepsilon &< |P_{t_m}(A) - P_{t_l}(A)| = |P(\Delta\beta_{t_m}(\Delta\beta_0) \in A) - P(\Delta\beta_{t_l}(\Delta\beta_0) \in A)| = \\ &= |P(\Delta\beta_{t_l}(\Delta\beta_{t_m-t_l}) \in A) - P(\Delta\beta_{t_l}(\Delta\beta_0) \in A)| \leq \\ &\leq |P(\Delta\beta_{t_l}(\Delta\beta_{t_m-t_l}) \in A, \Delta\beta_{t_m-t_l} \in K) - P(\Delta\beta_{t_l}(\Delta\beta_0) \in A)| + \varepsilon \leq \\ &\leq |P(\Delta\beta_{t_l}(\Delta\beta_0) \in A^\delta, \Delta\beta_{t_m-t_l} \in K) - P(\Delta\beta_{t_l}(\Delta\beta_0) \in A)| + \varepsilon \leq \\ &\leq |P(\Delta\beta_{t_l}(\Delta\beta_0) \in A^\delta) - P(\Delta\beta_{t_l}(\Delta\beta_0) \in A)| + 2\varepsilon = \\ &= |P_{t_l}(A^\delta) - P_{t_l}(A)| + 2\varepsilon \leq 5\varepsilon, \end{aligned}$$

ein Widerspruch. Also konvergieren die P_t schwach gegen ein P .

Da die M_t^{KK} nach Theorem 7.2.18 geometrisch gegen M^{KK} konvergieren folgt die allgemeine Aussage, falls es uns gelingt zu zeigen:

Mit dem Kompaktum K aus der vorangegangenen Argumentationskette gilt für $h > h_0 \in \mathbb{N}$

$$\forall l \in \mathbb{N} : \sup_{x \in K} |\Delta\beta_h^{(1)}(x) - \Delta\beta_h^{(2)}(x)| = o_P(t^0), \quad t \rightarrow \infty, \quad (7.27)$$

wobei sich $\Delta\beta_h^{(1)}$ aus den Rekursionen mit $\bar{M}_i = M^{\text{KK}} \min\{1, b_i/|M^{\text{KK}} \cdot |\}$ für $i = t+1, \dots, t+h$ ergibt und $\Delta\beta_h^{(2)}$ entsprechend mit $i = t+l+1, \dots, t+l+h$. Dies ist die im Theorem formulierte "hinreichend schnelle" Konvergenz der b_t . Für die asymptotische strikte Stationarität, müssen wir nun noch zeigen, daß ebenso die endlich-dimensionalen Randverteilungen des $\Delta\beta^{\text{rLS}}$ -Prozesses konvergieren.

Fassen wir die letzten $m+1$ Fehler in einem Vektor zusammen und unterdrücken das Superscript rLS,

$$\Delta\beta_{t-m:t} := (\Delta\beta_{t-m}^\tau, \dots, \Delta\beta_t^\tau)^\tau,$$

so erkennen wir

$$\Delta\beta_{t-m+1:t+1} = \mathbf{G}_{t+1} \Delta\beta_{t-m:t} - ((F\tilde{M}_t), 0, \dots, 0)^\tau \varepsilon_t + (1, 0, \dots, 0)^\tau v_{t+1}$$

mit

$$\mathbf{G}_{t+1} = \begin{pmatrix} \tilde{\mathbf{F}}_{t+1} & 0 & \cdots & 0 \\ \mathbb{I}_p & & & \\ & \mathbb{I}_p & & \\ & & \ddots & \\ & & & \mathbb{I}_p & 0 \end{pmatrix},$$

und $|\mathbf{G}_{t+1}| = |\mathbf{F}_{t+1}|$, so daß der Beweis wie für $m = 0$ geht.

Ebenso geht man für den Filterfehler $\beta_t - \beta_{t|t}$ vor.

Die Aussage für die Filter und Vorhersagen für $|F| < 1$ folgen wie im klassischen Rahmen. Die Ergodizität überträgt sich nach Lemma A.3.2, von den (ε_t, v_t) auf die Fehlerterme bzw. auf die Filter und Vorhersagen. ////

BEWEIS [2.VARIANTE]: Wir zeigen nur die asymptotische Stationarität im homogenen Fall, also $M_t^{\text{KK}} \equiv M^{\text{KK}}$ und $b_t \equiv b > 0$, den Übergang zum inhomogenen Fall erledigt man wie in der ersten Variante.

Offenbar handelt es sich bei unseren rLS-Rekursionen um ein Markovsystem gemäß (13), und nach unseren zusätzlichen Annahmen sind $(\varepsilon_t, v_t)_t$ absolutstetig, in der Zeit unabhängig identisch verteilt und besitzen eine unterhalbstetige, gemeinsame Lebesgue-Dichte, die im Punkt 0 nicht verschwindet. Die Funktion $\tilde{M}(0) = 0$ und \tilde{M} ist für $b > 0$ an der Stelle 0 differenzierbar mit Ableitung M , und überall stetig. Das System ist straff (siehe Beweis, 1. Variante) und schwach beobachtbar: Für alle $y \in \mathbb{R}^p$ gilt für die generalisierte Übergangsmatrix $\text{rk } C_y^1 = p$, denn von jedem $y \in \mathbb{R}^p$ kann ich durch Wahl von v_1 in einem Schritt \mathcal{G} den Zustand 0 erreichen. In diesem gilt für den letzten Block B_{T-1} in $\text{rk } C_0^T: B_{T-1} = [-MV^{\frac{1}{2}}, Q^{\frac{1}{2}}]$ und damit $\text{rk } C_0^T = p$ nach (22). Also sind die Voraussetzungen von Theorem 7.3.9 erfüllt und es folgt die Behauptung. ////

7.5 asymptotische Stabilität und Stationarität des rIC-Filters

Theorem 7.5.1 *Sei ein zeitinvariantes Zustandsraummodell gemäß (1) und (2) gegeben, also $F_t = F$ und $Z_t = F$ für alle t . Weiter sei die Verteilungsannahme (V2) mit $Q_t = Q$ und $V_t = Q$ für alle t erfüllt. Sei (F, Z^τ) vollständig ermittelbar. Die Fehlerterme ε_t und die Innovationen v_t mögen strikt stationär sein und (im idealen Modell) (V2) genügen, und gelte zusätzlich (22) und für die Varianten mit Ausnahme der separat-IO-gestutzten Variante des rIC-Filters gelte $|F| < 1$.*

Dann gilt für alle eingeführten rIC-Varianten, d.h. separat / gemeinsam gestutzt, bzw. un-/ informations-/selbststandardisiert:

Kalibriert man diese gemäß (19), so gibt es ein $\delta_0 > 0$, so daß für alle $0 < \delta < \delta_0$ der Filterfehler $\Delta\beta_t$ strikt asymptotisch stationär und asymptotisch stabil ist. Gilt weiter $|F| < 1$, so ist dann auch der Filter $\beta_{t|t}$, sowie die Vorhersage $\beta_{t|t-1}$ strikt asymptotisch stationär.

Sind die treibenden Prozesse v_t, ε_t zusätzlich ergodisch, so gilt dies auch für die Fehlerprozesse, und falls $|F| < 1$ auch für die Filter und Einschrittvorhersagen.

Wir beweisen Theorem 7.5.1 nur einmal; ein Beweis analog zu dem für den rLS-Fall, basierend auf den Techniken aus Abschnitt 7.3 unter den dort gemachten

Zusatzvoraussetzungen läßt sich ebenfalls gewinnen, wir verzichten hier aber darauf.

BEWEIS: Wir bemerken, daß sich die Filtergleichung jeder der Varianten schreiben läßt als

$$\beta_{i,t}^{\text{riC}} = \beta_{i,t-1}^{\text{riC}} + A_t \Lambda_{1,t} w_{1,t} + A_t \Lambda_{2,t} w_{2,t}$$

mit $\Lambda_{1,t} = Z^\tau V^{-1} (Z \Delta \beta_t^{\text{riC}} + \varepsilon_t)$, $\Lambda_{2,t} = \Sigma_{t|t-1}^{\text{KK}} (\Delta \beta_t^{\text{riC}} - \Delta \beta_t^{\text{KK}})$, wobei $\Delta \beta_t^{\text{KK}}$ und $\Sigma_{t|t-1}^{\text{KK}}$ die entsprechenden Größen aus dem klassischen Kalmanfilter sind, die Norm $|\cdot|$ für die jeweils passende⁸ steht, und für $i = 1, 2$ gilt

$$w_{i,t} = \min\left\{1, \frac{b_t^{(i)}}{|A_t \Lambda_{i,t}|}\right\} \quad \text{oder} \quad w_{i,t} = \min\left\{1, \frac{b_t}{|A_t (\Lambda_{1,t} + \Lambda_{2,t})|}\right\}. \quad (7.28)$$

Dementsprechend ergibt sich

$$\begin{pmatrix} \Delta \beta_{t+1}^{\text{riC}} \\ \Delta \beta_{t+1}^{\text{KK}} \end{pmatrix} = \tilde{F}_t \begin{pmatrix} \Delta \beta_t^{\text{riC}} \\ \Delta \beta_t^{\text{KK}} \end{pmatrix} + \eta_t \quad (7.29)$$

mit

$$\begin{aligned} \tilde{F}_t &= \begin{pmatrix} \tilde{F}_t^{(1,1)} & \tilde{F}_t^{(1,2)} \\ \tilde{F}_t^{(2,1)} & \tilde{F}_t^{(2,2)} \end{pmatrix} = \\ &= F \begin{pmatrix} \mathbb{I}_p - A_t (Z^\tau V^{-1} Z w_{1,t} + \Sigma_{t|t-1}^{\text{KK}-1} w_{2,t}) & A_t \Sigma_{t|t-1}^{\text{KK}-1} w_{2,t} \\ 0 & \mathbb{I}_p - M_t^{\text{KK}} Z \end{pmatrix}, \end{aligned} \quad (7.30)$$

$$\eta_t = \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \end{pmatrix} = -F \begin{pmatrix} A_t Z^\tau V^{-1} w_{1,t} \\ M_t^{\text{KK}} \end{pmatrix} \varepsilon_t + v_{t+1} \quad (7.31)$$

Betrachten wir den Term $C_t := \mathbb{I}_p - A_t (Z^\tau V^{-1} Z w_{1,t} + \Sigma_{t|t-1}^{\text{KK}-1} w_{2,t})$, so sehen wir

$$\begin{aligned} C_t &= \mathbb{I}_p - \Sigma_{t|t}^{\text{KK}} [Z^\tau V^{-1} Z + \Sigma_{t|t-1}^{\text{KK}-1}] + \\ &\quad + \Sigma_{t|t}^{\text{KK}} [Z^\tau V^{-1} Z (1 - w_{1,t}) + \Sigma_{t|t-1}^{\text{KK}-1} (1 - w_{2,t})] + \\ &\quad + [\Sigma_{t|t}^{\text{KK}} - A_t] (Z^\tau V^{-1} Z w_{1,t} + \Sigma_{t|t-1}^{\text{KK}-1} w_{2,t}) = \\ &= \Sigma_{t|t}^{\text{KK}} [Z^\tau V^{-1} Z (1 - w_{1,t}) + \Sigma_{t|t-1}^{\text{KK}-1} (1 - w_{2,t})] + \\ &\quad + [\Sigma_{t|t}^{\text{KK}} - A_t] (Z^\tau V^{-1} Z w_{1,t} + \Sigma_{t|t-1}^{\text{KK}-1} w_{2,t}) =: C_{1,t} + C_{2,t} \end{aligned}$$

Nun ist für δ hinreichend klein $|\Sigma_{t|t}^{\text{KK}} - A_t|$ beliebig klein, so daß dann nur $C_{1,t}$ die Norm von C_t bestimmt. Andererseits ist $[Z^\tau V^{-1} Z (1 - w_{1,t}) + \Sigma_{t|t-1}^{\text{KK}-1} (1 - w_{2,t})]$ symmetrisch positiv semidefinit und maximal im Sinn der Halbordnung “ \prec ”, falls $w_{i,t} = 0$, und dann ist

$$C_{1,t} = \Sigma_{t|t}^{\text{KK}} [Z^\tau V^{-1} Z + \Sigma_{t|t-1}^{\text{KK}-1}] = \mathbb{I}_p.$$

Im Fall der separat-IO-Stützung ist $C_{1,t}$ maximal, falls $w_{2,t} = 0$, und dann ist

$$C_{1,t} = \Sigma_{t|t}^{\text{KK}} \Sigma_{t|t-1}^{\text{KK}-1} = \mathbb{I}_p - M_t^{\text{KK}} Z.$$

⁸un-/ informations-/ selbststandardisiert

Diese Anordnung überträgt sich nach Korollar C.1.8, so daß in jedem Fall für $\delta \rightarrow 0$ gilt, daß $|FC_t| \leq |FC_{1,t}| + o(\delta^0) < 1$ ist. Aus Stetigkeitsgründen gilt dies bereits für $0 \neq \delta$ hinreichend klein, und somit $|\tilde{\mathbf{F}}_t| < 1$ schließlich t , was wiederum die Straffheit von

$$\underline{\Delta\beta}_t := (\Delta\beta_t^{\text{rIC } \tau}, \Delta\beta_t^{\text{KK } \tau})^\tau$$

bedeutet.

Genau wie im ersten Beweis zu Theorem 7.4.1 fassen wir die letzten $m+1$ Fehler in einem Vektor zusammen,

$$\underline{\Delta\beta}_{t-m:t} := (\underline{\Delta\beta}_{t-m}^\tau, \dots, \underline{\Delta\beta}_t^\tau)^\tau,$$

und erkennen:

$$\underline{\Delta\beta}_{t-m+1:t+1} = \mathbf{G}_{t+1} \underline{\Delta\beta}_{t-m:t} + (\eta_t, 0, \dots, 0)^\tau$$

mit

$$\mathbf{G}_{t+1} = \begin{pmatrix} \mathbf{F}_{t+1} & 0 & \cdots & 0 \\ \mathbb{I}_{2p} & & & \\ & \mathbb{I}_{2p} & & \\ & & \ddots & \\ & & & \mathbb{I}_{2p} & 0 \end{pmatrix},$$

und $|\mathbf{G}_{t+1}| = |\mathbf{F}_{t+1}|$, so daß Straffheit folgt wie für $m=0$. Analog zum ersten Beweis zu Theorem 7.4.1 setzen wir für $0 \leq k \leq t$ $\tilde{\mathbf{F}}_{t:k} := \tilde{\mathbf{F}}_t \cdots \tilde{\mathbf{F}}_k$, sowie $\tilde{\mathbf{F}}_{t:k}^{(1,1)} := \tilde{\mathbf{F}}_t^{(1,1)} \cdots \tilde{\mathbf{F}}_k^{(1,1)}$. Damit können wir schreiben

$$\underline{\Delta\beta}_{t+1} = \tilde{\mathbf{F}}_{t:1} \underline{\Delta\beta}_1 + \sum_{k=2}^t \tilde{\mathbf{F}}_{t:k} \eta_k.$$

Nach (22) gilt $\text{Var } \eta_{1,k} \succeq Q \succ 0$. Wie im Beweis zu Theorem 7.4.1 folgt $E|\tilde{\mathbf{F}}_{t:0}^{(1,1)}| \rightarrow 0$, gleiches gilt nach Voraussetzung für $\mathbf{F}_{t:0}$ aus diesem Beweis. Weil aber $\tilde{\mathbf{F}}_{t:0}^{(1,2)}$ sicher beschränkt ist, folgt somit auch $E|\tilde{\mathbf{F}}_{t:0}| \rightarrow 0$, also asymptotische Stabilität und fast sichere Konvergenz gegen 0.

Genauso zeigen wir auch die Konvergenz gegen 0 von $\tilde{\mathbf{F}}_{t+h:h}$ für $h > 0$.

Wegen gemeinsamer Straffheit von $\underline{\Delta\beta}_t$ folgt nach dem Satz von Prokhorov die Existenz schwacher Häufungspunkte. Wie im Beweis zu Theorem 7.4.1 folgt unter der Annahme, daß $A_t \equiv A$ und $b_t^{[(i)]} \equiv b^{[(i)]} > 0$, daß es nur einen solchen geben kann.

Da die Kalibrierung im idealen, normalverteilten Modell erfolgt, konvergieren auch die b_t geometrisch, denn diese sind im normalverteilten Modell stetig differenzierbar in der Kovarianz $\Sigma_{t|t-1}^{\text{KK}}$ mit einer beschränkten Ableitung (vgl.), und dies ist die einzige Abhängigkeit von b_t in t ; gleiches gilt für die A_t .

Ob allerdings die Geschwindigkeit dieser Konvergenz hinreichend groß ist, wissen wir hier noch nicht; sie ist es, falls es uns gelingt, (27) zu zeigen, wobei wir an die Stelle der $\Delta\beta_h^{(i)}(x)$ die Paare $\underline{\Delta\beta}_h^{(i)}$ verwenden.

Im rIC-Fall können wir (27) für δ hinreichend klein tatsächlich nachweisen: Wie im Beweis zu Theorem 7.4.1 folgt, daß es zu ε ein t_0 gibt, so daß für $h > t_0$ gilt $|\tilde{\mathbf{F}}_{t+h:t} \Delta\beta_0| < \varepsilon/2$ für alle $\Delta\beta_0$ im Straffheitskompaktum K . Daher können

wir uns bei der Fragestellung, ob $\sup_{x \in K} |\underline{\Delta\beta}_h^{(1)}(x) - \underline{\Delta\beta}_h^{(2)}(x)|$ hinreichend klein wird, auf $h = t_0$ beschränken. Weiter wählen wir den Anfangszeitpunkt t so spät, daß für ein $0 < \rho < 1$ und $L > 0$ gilt: $|b_s - b| < \varepsilon$ und $|A_s - A| < \rho^s L$ für alle $s > t$.

Schließlich ist, da $\text{rk} \Sigma_{t|t}^{\text{KK}} = p$, auch $\text{rk} A_t = p$ und damit können wir A_t schreiben als $A(1 + \rho^t L_t)$ mit L_t einer beschränkten Folge von Matrizen.

Aus der Straffheit von $\underline{\Delta\beta}_{t-m:t}$ für $m = t_0$ folgt auch die Straffheit von $\underline{\Delta y}_{t-t_0:t}$, und somit können wir, indem wir δ hinreichend klein wählen, erreichen, daß in einem t_0 -Abschnitt unabhängig vom "Aufpunkt" des t_0 -Abschnitts mit Wahrscheinlichkeit $1 - \varepsilon$ kein Clipping durchgeführt wird, denn aus der Beschränktheit von A_t und b_t folgt, daß dies uniform geschehen kann. Somit unterscheiden sich $\underline{\Delta\beta}_h^{(i)}$ auf diesem Straffheitskompaktum nur in den A_t . Für hinreichend spätes t als Anfangszeitpunkt gilt nun für feste Realisationen aus dem Straffheitskompaktum $\underline{\Delta\beta}_t$, ε_{t+j} , v_{t+j+1} , $j = 1, \dots, t_0$ und l beliebig:

$$\begin{aligned} & \underline{\Delta\beta}_{t+t_0}^{(1)} - \underline{\Delta\beta}_{t+t_0}^{(2)} = \\ & = (\tilde{\mathbf{F}}_{t+t_0:t} - \tilde{\mathbf{F}}_{t+l+t_0:l+t}) \underline{\Delta\beta}_t + \sum_{k=2}^{t_0} (\tilde{\mathbf{F}}_{t+t_0:k+t} - \tilde{\mathbf{F}}_{t+l+t_0:k+l+t}) v_{t+1+k} - \\ & \quad - \sum_{k=2}^{t_0} \left(\tilde{\mathbf{F}}_{t+t_0:k+t} \begin{pmatrix} A_{t+k} Z^T V^{-1} \\ M_{t+k}^{\text{KK}} \end{pmatrix} - \tilde{\mathbf{F}}_{t+l+t_0:k+l+t} \begin{pmatrix} A_{t+l+k} Z^T V^{-1} \\ M_{t+l+k}^{\text{KK}} \end{pmatrix} \right) \varepsilon_{t+k} \end{aligned}$$

Nun läßt sich auch $\tilde{\mathbf{F}}_t$ schreiben als $\tilde{\mathbf{F}}_t = \tilde{\mathbf{F}}(1 + \rho^t \tilde{L})$ und somit $\tilde{\mathbf{F}}_{t+t_0:t}$ als $\tilde{\mathbf{F}}^{t_0} (1 + \rho^t \tilde{L}_2)$ mit $L_2 = O(t^0)$ und einem $\tilde{\mathbf{F}}$ mit $\text{rk} \tilde{\mathbf{F}} = p$ und $|\tilde{\mathbf{F}}| < 1$. Daher ist mit $L_3 = O(t^0)$

$$\tilde{\mathbf{F}}_{t+t_0:k+t} - \tilde{\mathbf{F}}_{t+l+t_0:k+l+t} = \tilde{\mathbf{F}}^{t_0-k} (\rho^t L_3)$$

Ähnlich sieht man

$$|\underline{\Delta\beta}_{t+t_0}^{(1)} - \underline{\Delta\beta}_{t+t_0}^{(2)}| \leq 2\varepsilon + \sum_{k=2}^{t_0} |\tilde{\mathbf{F}}|^{t_0-k} \rho^t (|L_3| |v_{t+1+k}| + |L_4| |\varepsilon_{t+k}|) = O(\rho^t)$$

und damit wie behauptet (27). Die restlichen Aussagen des Theorems folgen wie im rLS-Fall. ////

7.6 asymptotische Stabilität und Stationarität des mIC-Filters

Theorem 7.6.1 *Sei ein zeitinvariantes Zustandsraummodell gemäß (1) und (2) gegeben, also $F_t = F$ und $Z_t = Z$ für alle t . Weiter sei die Verteilungsannahme (V2) mit $Q_t = Q$ und $V_t = V$ für alle t erfüllt. Sei (F, Z^T) vollständig ermittelbar. Die Fehlerterme ε_t und die Innovationen v_t mögen strikt stationär sein und (im idealen Modell) (V2) genügen, und gelte zusätzlich (22) und für die Varianten mit Ausnahme der separat-IO-gestützten Variante des rIC-Filters gelte $|F| < 1$.*

Dann gilt für alle eingeführten mIC-Varianten zu festem m , d.h. blockweise /

total gestutzt, separat / gemeinsam gestutzt, bzw. un-/ informations-/ selbststandardisiert:

Kalibriert man diese gemäß (26), so gibt es ein $\delta_0 > 0$, so daß für alle $0 < \delta < \delta_0$ der Vektor der Fehler $\Delta\beta_{t-i|t}$, $i = 0, \dots, m$ strikt asymptotisch stationär und asymptotisch stabil ist. Gilt weiter $|F| < 1$, so sind dann auch der Vektor der Vorhersage, Filter und Glätter $\Delta\beta_{t-i|t}$, $i = 0, \dots, m$ bzw. $i = -1, \dots, m-1$ strikt asymptotisch stationär.

Sind die treibenden Prozesse v_t, ε_t zusätzlich ergodisch, so gilt dies auch für die Prozesse der $[(\Delta\beta_{t-i|t})_{i=0, \dots, m}]_t$, und falls $|F| < 1$ auch für die Prozesse der $[(\beta_{t-i|t})_{i=-1, 0, \dots, m-1}]_t$.

Wie im rIC-Fall beweisen wir Theorem 7.6.1 nur einmal; ein Beweis analog zu dem für den rLS-Fall, basierend auf den Techniken aus Abschnitt 7.3 unter den dort gemachten Zusatzvoraussetzungen läßt sich ebenfalls gewinnen, wir verzichten aber an dieser Stelle darauf.

BEWEIS: Wir bemerken, daß sich die Filtergleichung jeder der Varianten schreiben läßt als

$$\beta_{t:t-m|t}^{\text{mIC}} = \beta_{t:t-m|t-1}^{\text{mIC}} + A_t \Lambda_{1,t} w_{1,t} + A_t \Lambda_{2,t} w_{2,t}$$

mit

$$\Lambda_{1,t} = Z^\tau V^{-1} Z (\Delta\beta_{t:t-m|t-1}^{\text{mIC}} + \varepsilon_{t:t-m}),$$

$$\Lambda_{2,t} = \begin{pmatrix} Q^{-1} v_t \\ \left\{ \begin{array}{l} F^\tau Q^{-1} (\Delta\beta_{t:t-m+2|t-1}^{\text{mIC}} - F \Delta\beta_{t-1:t-m+1|t-1}^{\text{mIC}} - v_{t:t-m+2}) \\ -Q^{-1} (\Delta\beta_{t-1:t-m+1|t-1}^{\text{mIC}} - F \Delta\beta_{t-2:t-m|t-1}^{\text{mIC}} - v_{t-1:t-m+1}) \end{array} \right\} \\ \left\{ \begin{array}{l} F^\tau Q^{-1} (\Delta\beta_{t-m+1|t-1}^{\text{mIC}} - F \Delta\beta_{t-m|t-1}^{\text{mIC}} - v_{t-m+1}) \\ -\Sigma_{t-m|t-m-1}^{\text{KK}} (\Delta\beta_{t-m|t-1}^{\text{mIC}} - \Delta\beta_{t-m}^{\text{KK}}) \end{array} \right\} \end{pmatrix}$$

wobei $\Delta\beta_{t-m}^{\text{KK}}$ und $\Sigma_{t-m|t-m-1}^{\text{KK}}$ die entsprechenden Größen aus dem klassischen Kalmanfilter sind, und für $i = 1, 2$ $w_{i,t}$ die jeweils passende⁹ Gewichtsfunktion. Wir interpretieren dabei die jeweils außerhalb der Klammer stehende $p \times p$ Matrix als Faktor, der mit jedem der $p \times p$ Blöcke in der Klammer im Sinne eines verallgemeinerten Distributivgesetzes multipliziert wird. Dementsprechend ergibt sich

$$\begin{pmatrix} \Delta\beta_{t+1:t-m+1|t}^{\text{mIC}} \\ \Delta\beta_{t-m+1}^{\text{KK}} \end{pmatrix} = A \tilde{F}_t \begin{pmatrix} \Delta\beta_{t:t-m|t-1}^{\text{mIC}} \\ \Delta\beta_{t-m}^{\text{KK}} \end{pmatrix} + A \eta_t \quad (7.32)$$

mit

$$\tilde{F}_t = \text{diag}[F Z^\tau V^{-1} Z w_{1,t,t}, \quad (7.33)$$

$$\begin{aligned} & \{Z^\tau V^{-1} Z w_{1,t,t-i} + (Q + F^\tau Q F) w_{2,t,t-i}\}_{i=0, \dots, m-1}, \\ & Z^\tau V^{-1} Z w_{1,t,t-m} + (F^\tau Q F - \Sigma_{t-m|t-m-1}^{\text{KK}}) w_{2,t,t-m}, F(\mathbb{I}_p - M_t^{\text{KK}} Z) \\ & + \text{ndiag}_1[0, \{Q^{-1} F w_{2,t,t-i}\}_{i=1, \dots, m-1}, \Sigma_{t-m|t-m-1}^{\text{KK}} w_{2,t,t-m}] \\ & + \text{ndiag}_1[\{Q^{-1} F w_{2,t,t-i}\}_{i=1, \dots, m}, 0]^\tau \end{aligned} \quad (7.34)$$

⁹block-/ total-, , separat / gemeinsam gestutzt, un-/ informations-/ selbststandardisiert

und

$$\begin{aligned} \eta_t = & Z^\tau V^{-1} \varepsilon_{t:t-m} w_{1,t:t-m} + Q^{-1} (v_{t:t-m+1}^\tau w_{2,t:t-m+1}, 0^\tau)^\tau - \\ & - F^\tau Q^{-1} (0^\tau, v_{t:t-m+1}^\tau w_{2,t:t-m+1}^\tau)^\tau \end{aligned} \quad (7.35)$$

Setzen wir alle Clippingschwellen b auf ∞ , so erhalten wir den klassischen Kalman-Filter und -Glätter der Ordnung m . Hierfür schreiben wir $\tilde{\mathbf{F}}_t^0$. Damit erhalten wir

$$\begin{pmatrix} \Delta \beta_{t+1:t-m+1|t}^{\text{mIC}} \\ \Delta \beta_{t-m+1}^{\text{KK}} \end{pmatrix} = \mathcal{I}^{-1} \tilde{\mathbf{F}}_t^0 \begin{pmatrix} \Delta \beta_{t:t-m|t-1}^{\text{mIC}} \\ \Delta \beta_{t-m}^{\text{KK}} \end{pmatrix} + A \eta_t - \quad (7.36)$$

$$-\mathcal{I}^{-1} (\tilde{\mathbf{F}}_t^0 - \tilde{\mathbf{F}}_t) \begin{pmatrix} \Delta \beta_{t:t-m|t-1}^{\text{mIC}} \\ \Delta \beta_{t-m}^{\text{KK}} \end{pmatrix} + (A - \mathcal{I}^{-1}) \tilde{\mathbf{F}}_t \begin{pmatrix} \Delta \beta_{t:t-m|t-1}^{\text{mIC}} \\ \Delta \beta_{t-m}^{\text{KK}} \end{pmatrix} \quad (7.37)$$

Die rechte Seite in (35) ist nach Korollar 7.2.19 (asymptotisch) schwach stationär, und somit muß $\rho(\mathcal{I}^{-1} \tilde{\mathbf{F}}_t^0) < 1$ sein. Ähnlich wie im rIC-Fall können wir durch Wahl von δ hinreichend klein erreichen, daß $\rho(\mathcal{I}^{-1} (\tilde{\mathbf{F}}_t^0 - \tilde{\mathbf{F}}_t))$ und $\rho((A - \mathcal{I}^{-1}) \tilde{\mathbf{F}}_t)$ so klein sind, daß auch $\rho(A \tilde{\mathbf{F}}_t) < 1$. Dies führen wir hier nicht im Detail aus.

Die weiteren Aussagen des Theorems folgen mutatis mutandis wie in Theorem 7.5.1. ////

Kapitel 8

optimal–robuste Filter

8.1 Problemstellung

Während wir in den vorangegangenen Kapiteln die Ansätze zur Robustifizierung des Kalman–Filters durch Übertragung von Methoden gewonnen haben, die sich in anderen Kontexten als optimal herausgestellt haben — den robust–optimalen Regressions–IC’s für den rIC– bzw. mIC–Filter und die Huber’sche Stutzfunktion für eine ad–hoc Robustifizierung des optimalen, linearen Verfahrens beim rLS, wollen wir in diesem Abschnitt robuste Optimierungsprobleme im Kontext des Zustandsraummodells stellen und sie, soweit möglich, auch darin lösen.

8.1.1 verfolgte Ansätze

Im folgenden werden wir dazu zwei Ansätze verfolgen: den Minimax–Ansatz gemäß Huber (1964) und den “Lemma 5”–Ansatz, benannt nach einem Lemma in Hampel (1968).

Ersterer versucht das Risiko R eines Verfahrens f auf Umgebung \mathcal{U} einer idealen Verteilung zu minimieren, salopp also \hat{f} als Lösung von $R = \min_f \max_{\mathcal{U}}!$ zu finden; im Spezialfall des univariaten Lokationsmodells¹

$$\mathcal{P} = \{P_\theta = F(\cdot - \theta) \mid \theta \in \mathbb{R}\},$$

mit wahren, unbekanntem Parameter $\theta_0 = 0$ und $\mathcal{U} = \{(1-r)F + rH, H \in W \subset \mathcal{M}_1(\mathbb{B})\}$ zu bekannten $r > 0$ und W lautet diese Aufgabe: Finde zu unabhängig identisch P_θ –verteilten Beobachtungen X_1, \dots, X_n ein $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n)$, so daß

$$\sup_{\mathcal{U}}[\text{as.}] \text{Var}[\hat{\theta}_n] = \min_{\theta_n} \sup_{\mathcal{U}}[\text{as.}] \text{Var}[\theta_n]$$

Diese Aufgabe ist für F symmetrisch mit W als der Menge aller symmetrischen Verteilungen von Huber (1964) gelöst worden und ergibt im Fall $F = \mathcal{N}(0, 1)$ einen Schätzer $\hat{\theta}_n$ mit Influenzkurve $\psi(x) = a(-\tilde{b} \vee x \wedge \tilde{b}) = ax \min\{1, b/|ax|\}$ mit einer Konstanten a , die die Fisherkonsistenz sichert und $\tilde{b} = b/a$ einer Clipping–Schwelle.

¹ F mit endlicher Lokations–Fisher–Information \mathcal{I}_F gemäß (4), Abschnitt B.2

Der “Lemma 5”-Ansatz dagegen legt der klassischen (nicht-robusten), im idealen Modell auszuwertenden Zielfunktion eine Nebenbedingung auf, die für die Robustheit sorgt. Zum Beispiel kann die Varianz im idealen Modell als Zielfunktion genommen werden, wobei nur Schätzer zugelassen werden, die einer Schranke an den Bias in einer Umgebung genügen. Dies ist für asymptotisch lineare Schätzer äquivalent dazu (vgl. Rieder (1994)), die L_2 -Norm der Influenzkurve unter einer L_∞ -Normschranke zu minimieren, was im Spezialfall des univariaten Lokationsmodells auf dieselbe Influenzkurve führt wie im Minimax-Ansatz.

8.1.2 Problemstellung im vereinfachten Zustandsraummodell

Wir wollen die Rekursivität des Modells beibehalten und formulieren daher das Problem um in eine Folge von Problemen zu verschiedenen Zeitpunkten, die dann jeweils unter der Prämisse zu lösen sind, bisher das optimale Verfahren verwendet zu haben.

Um in einheitlichen Dimensionen arbeiten zu können vereinfachen wir das Zustandsraummodell zunächst wie folgt: X sei eine \mathbb{R}^p -wertige Zufallsvariable, die im idealen Modell der Verteilung P^X mit endlichem zweiten Moment folgt. Diese kann man allerdings nicht beobachten, sondern nur eine Variable Y ,

$$Y := X + \varepsilon, \quad (8.1)$$

wobei ε eine von X stochastisch unabhängige, \mathbb{R}^p -wertige Zufallsvariable ist, die im idealen Modell der Verteilung P^ε gehorcht.

Aufgabe des Statistikers ist es nun, aufgrund von Y eine möglichst genaue Aussage über X zu treffen, genauer ist mit der quadratischen Verlustfunktion

$$E_{\text{id}}[|X - f(Y)|^2] = \min_f! \quad (8.2)$$

zu optimieren.

Im idealen Modell wird dies bekanntlich durch den bedingten Erwartungswert $\check{f}(y) = E_{\text{id}}[X|Y]$ erreicht, der aber nicht robust gegenüber Ausreißern in X , ε oder Y ist.

Während nun die Minimax-Formulierung des Problems längs geeigneter Umgebungen klar ist, müssen wir für den “Lemma 5”-Ansatz noch die robuste Nebenbedingung spezifizieren: Wir lassen nur solche $f \in L_2(P^Y)$ zu, für die gilt²

$$|E_{\text{real}}[X - f(Y)]| \leq b \quad (8.3)$$

$$E_{\text{id}}[f(Y)] = E_{\text{id}}[X] \quad (8.4)$$

8.1.3 SO- und AO-Ausreißer

Im folgenden wollen wir uns auf transiente Ausreißer beschränken, also keine IO-Ausreißer betrachten.

²Die Unverzerrtheitsbedingung (4) ist der Preis dafür, daß wir ansonsten die additive Struktur des Modells (8.1) in diesem Ansatz gar nicht berücksichtigen werden.

SO-Ausreißer

Zusätzlich zu den bisher eingeführten AO-Ausreißern [in ε] wollen wir nun auch substitutive Ausreißer (SO) betrachten, bei denen die Verteilung P^Y von Y direkt kontaminiert wird, indem mit einer gewissen Wahrscheinlichkeit die “richtige” Beobachtung Y durch eine Beobachtung \tilde{Y} ersetzt wird, die, unabhängig von X , ε , einer unbekanntem Verteilung \tilde{P}^Y folgt.

Formal kann man dies so modellieren:

$$Y := X + \varepsilon, \quad (8.5)$$

$$\hat{Y} := (1 - U)Y + U\tilde{Y}, \quad (8.6)$$

$X \sim P^X$, $\varepsilon \sim P^\varepsilon$, $\tilde{Y} \sim \tilde{P}^Y$, und $U \sim \text{Bin}(1, r)$, $X, \varepsilon, \tilde{Y}, U$ stochastisch unabhängig, r , P^ε und P^X bekannt, \tilde{P}^Y unbekannt.

Dabei kann man \hat{Y} auch als Zufallsvariable mit Verteilung \hat{P}^Y aus einer Konvexkombinationsumgebung mit Radius r um $P^Y := P^X * P^\varepsilon$ auffassen, also $\hat{P}^Y \in U_r(P^X * P^\varepsilon)$ mit

$$U_r(P^X * P^\varepsilon) := \{\hat{P}^Y = (1 - r)(P^X * P^\varepsilon) + r\tilde{P}^Y \in \mathcal{M}_1(\mathbb{B}^p)\} \quad (8.7)$$

Genauer wollen wir aber die Menge der Bildverteilungen von (X, \hat{Y}) betrachten, also

$$\mathcal{U}_r := \{\mathcal{L}(X, \hat{Y}), X \sim P^\varepsilon, \hat{Y} \text{ gemäß (5) und (6)}\} \quad (8.8)$$

Bemerkung 8.1.1 Ohne den Begriff SO dafür zu verwenden, benützen Birnwal und Shen (1993) und McGarty (1975) diese Ausreißerdefinition.

AO-Ausreißer

Für AO's sei folgende Konvex-Kontaminationsumgebung um P^ε mit Radius r mit kontaminierenden Verteilungen aus einer bekannten konvexen Teilmenge $W \subset \mathcal{M}_1(\mathbb{B}^p)$ gegeben:

$$U_r(P^\varepsilon, W) := \{\hat{P}^\varepsilon = (1 - r)P^\varepsilon + r\tilde{P}^\varepsilon, \tilde{P}^\varepsilon \in W\} \quad (8.9)$$

Die Menge aller Maße, deren erste Randverteilung P^X ist und deren zweite sich als Faltung aus P^X und einem $\hat{P}^\varepsilon \in U_r(P^\varepsilon, W)$ darstellen läßt, wollen wir mit

$$\mathcal{V}_r(W) := \{\mathcal{L}(X, X + \hat{\varepsilon}), X \sim P^X, \hat{\varepsilon} \sim \hat{P}^\varepsilon \in U_r(P^\varepsilon, W)\} \quad (8.10)$$

bezeichnen. Formuliert mit Zufallsvariablen sieht dies so aus:

$$\hat{\varepsilon} := (1 - U)\varepsilon + U\tilde{\varepsilon} \quad (8.11)$$

$$\hat{Y} := X + \hat{\varepsilon}, \quad (8.12)$$

$X \sim P^X$, $\varepsilon \sim P^\varepsilon$, $\tilde{\varepsilon} \sim \tilde{P}^\varepsilon \in W$, und $U \sim \text{Bin}(1, r)$, $X, \varepsilon, \tilde{\varepsilon}, U$ stochastisch unabhängig, r , W , P^ε und P^X bekannt, \tilde{P}^ε unbekannt.

Bemerkung 8.1.2 (a) Um ein Beispiel für SO-Kontamination im Kontext des Satelliten-Steuerungsproblems von Kapitel 1 zu geben, stelle man sich vor, durch einen Ausfall der Meßgeräte sei das Beobachtungssignal für eine gewisse Zeit unterbrochen und man empfangt stattdessen ein Rauschen als Signal, das nichts mit der Position des Satelliten zu tun hat.

(b) Erlaubt man als AO-Kontaminationsverteilung eine Abhängigkeit vom “Regressor” X , so fallen AO und SO-Konzept zusammen, da man dann für jede SO-Kontamination $\tilde{P}^Y(dy)$ eine AO-kontaminierte Realisation des Markov-Kerns $\tilde{P}^{\varepsilon|X=x}(d\varepsilon) = \tilde{P}^{Y|X=x}(dy - x)$ wählen kann. Umgekehrt führt natürlich jede AO-Kontamination $\tilde{P}^{\varepsilon}(d\varepsilon)$ zu einer SO-Kontamination, indem wir als kontaminierendes Maß $\tilde{P}^Y(dy) = (\tilde{P}^{\varepsilon} * P^X)(dy)$ wählen.

(c) Ist im AO-Fall die Kontamination aber unabhängig von X zu wählen, so ist die Menge der erreichbaren Kontaminationen echt kleiner als die im SO-Fall, und man kann im Prinzip effizientere, robuste Verfahren zum gleichen Kontaminationsradius finden. In diesem Sinn handelt es sich bei SO-Ausreißern in diesem Einschnittszenario um die “gefährlicheren” als die AO-Ausreißer.

(d) Bei Betrachtung von Prozeßverteilungen geht die Möglichkeit, AO’s durch SO’s zu reproduzieren, nach dieser Definition verloren, sobald wir nicht nur die eindimensionalen Randverteilungen von Prozessen betrachten, wie H. Rieder dem Autor gegenüber mündlich erläutert hat.

8.1.4 Abgrenzung zu anderen Ansätzen

der ACM-(type-)Filter von Martin / Masreliez

Mit ihrem ACM-(type-)Filter haben Masreliez und Martin (1977) und Martin (1979) ein erstes Minimax-Verfahren in unserem Setup publiziert. Allerdings gelingt es den Autoren nicht, einen Sattelpunkt für Konvexkontaminationsumgebungen für AO’s oder IO’s herzuleiten, sondern nur für sogenannte p -Punkt-Umgebungen, im eindimensionalen für $0 < p < 1$ definiert als

$$\mathcal{U}_p = \{F \mid F(-u_{p/2} - 0) = F(-u_{p/2}) = \Phi(-u_{p/2}) = p/2, F \text{ symm.}\}.$$

Für die Konvexkontaminationsumgebungen schließen sie per Analogie zum Resultat von Huber (1964) im Lokationsmodell.

Die Lösung in Masreliez und Martin (1977) im Mehrdimensionalen stützt sich auf die Tatsache, daß sich je zwei beliebige Kovarianzmatrizen gleicher Dimension simultan diagonalisieren lassen³, d.h. Umgebungsbegriff und Robustifizierung hängen von dieser technisch zu wählenden Transformation ab, was aus Sicht der Robustheit willkürlich erscheint.

Auch die Verallgemeinerungen im Mehrdimensionalen zum ACM-type-Filter sind aus diesem Licht betrachtet unbefriedigend.

Die Autoren arbeiten unter der Annahme der Normalität von $\mathcal{L}(\beta_t|y_{1:t-1})$, die aber, wie wir in Abschnitt 6.3 gezeigt haben, nicht zutreffen kann. Dennoch hat der ACM-Filter eine weite Verbreitung gefunden und ist unter anderem in `Splus` verfügbar.

der Filter von Schick / Mitter

Schick (1989) und Schick und Mitter (1994) liefern im Setup der AO-Kontamination Argumente für eine Nähe von $\mathcal{L}(\beta_t|y_{1:t-1})$ zu einer Normalverteilung:

Indem wir zu jedem Zeitpunkt eine kontaminierende Verteilung $\tilde{P}_t^{\varepsilon}$ festlegen, und diese durch einen Mechanismus U_t realisieren, können wir die Vergangenheit gemäß der Realisationen $u_{1:t-1}$ partitionieren. Die Idee ist nun,

³vgl. Golub und Van Loan (1983), Theorem 8.6-1.

daß in einem Fenster der festen Länge ω mit sehr großer Wahrscheinlichkeit nur höchstens ein Ausreißer auftritt, daß aber Ausreißer außerhalb dieses Fensters wegen der exponentiell asymptotischen Stabilität des Systems, vgl. Abschnitt 7.2, auch (fast) keinen Einfluß mehr auf den Filter haben.

Was die Wahl der Kontaminationen \tilde{P}_t^ε anlangt, so gelingt es den Autoren nicht, diese ungünstigst zu ermitteln. Sie schlagen stattdessen vor, diese konservativ zu wählen, was, wie sich herausstellt, auf den Übergang zu SO-Kontaminationsumgebungen herausläuft. Allerdings verwenden sie als ungünstigste Verteilung im Analogieschluß die von Huber (1964), die sich aber in Theorem 8.3.1 nicht als die Sattelpunktsverteilung herausstellt.

Weiterhin müssen die Autoren zum Zeitpunkt t ein Bündel von t verschiedenen Filtern berechnen, was der Verwendung beim “online”-Filtern nicht dienlich ist. Außerdem ist die Tatsache, daß weit in der Vergangenheit zurückliegende Beobachtungen (fast) keinen Einfluß mehr auf den Filter haben, nur im Sinne eines $O_P((\omega \wedge t)^2 r_{\text{AO}}^2)$ zu verstehen, so daß auch hier ein besseres Resultat wünschenswert ist.

der Filter von Birmiwal / Shen / Papantoni-Kazakos

Indem Birmiwal und Shen (1993) und Birmiwal und Papantoni-Kazakos (1994) — ohne diese so zu nennen — SO-Umgebungen verwenden, sind sie in der Lage, ein exaktes Minimax-Resultat herzuleiten, wenn auch nur eindimensional und für absolutstetige Kontaminationen.

Leider ist auch ihr Verfahren nicht “online”-tauglich, da sie die Filtervergangenheit gemäß der gesamten U_t -Vergangenheit partitionieren, also im Zeitpunkt t ein Bündel von 2^t Filtern betrachten müssen. Auf diesen Ansatz werden wir in den Abschnitten 8.3 und 8.4 noch detaillierter zu sprechen kommen.

8.2 “Lemma 5”-Ansatz bei SO’s

8.2.1 Lösung im reduzierten Modell

Zunächst ist wegen der Y -Meßbarkeit von f klar, daß wir (2) ersetzen können durch

$$E_{\text{id}}[|E_{\text{id}}[X|Y] - f(Y)|^2] = \min_f! \quad (8.13)$$

Offenbar können wir dafür sorgen, daß (3) sicher erfüllt ist, indem wir anstelle von (3)

$$\sup_y |f(y) - E_{\text{id}}[X]| \leq b \quad (8.14)$$

fordern, denn es gilt

$$|E_{\text{real}}[X - f]| = |E_{\text{id}}[X] - E_{\text{real}}[f]| \leq \sup_y |f(y) - E_{\text{id}}[X]| \leq b$$

Im allgemeinen allerdings wird diese Einschränkung für gegebenes b zu stark sein, da ja auch unkontaminierte Bestandteile P^Y in \hat{P}^Y stecken. Immerhin werden wir sehen, daß (3) unter allen im idealen Modell unverzerrten Verfahren, also unter (4) genau dann mit Schranke b gilt, wenn (14) mit Schranke b/r gilt:

Schreiben wir $E_{\text{real}} = (1-r)E_{\text{id}} + rE_{\tilde{P}^Y}$, so gilt

$$\begin{aligned} & |E_{\text{real}}[X - f(Y)]| = |E[X] - E_{\text{real}}[f(Y)]| \leq b \\ \iff & |E[X] - \{(1-r)E_{\text{id}}[f(Y)] + rE_{\tilde{P}^Y}[f(Y)]\}| \leq b \\ \iff & |E[X] - E_{\tilde{P}^Y}[f(Y)]| \leq b/r \end{aligned} \quad (8.15)$$

Durch Wahl von Dirac-Maßen \tilde{P}^Y kann man leicht sehen — (c.f. Rieder (1994), pp. 172-77), daß dies wiederum äquivalent ist zu

$$\sup_y |f(y) - E[X]| \leq b/r \quad (8.16)$$

Hiermit können wir nun das Problem lösen: Da in der Nebenbedingung (16) jede in Y meßbare Funktion f zugelassen ist, können wir (13) minimieren, indem wir den Integranden punktweise in Y minimieren.

Also folgt mit Lemma C.2.2⁴, weil sowohl das zulässige Gebiet als auch die Zielfunktion konvex in Y sind, daß

$$\hat{f}^{\text{so}} := (E_{\text{id}}[X|Y] - E[X]) \min\left\{1, \frac{b/r}{|E_{\text{id}}[X|Y] - E[X]|}\right\} + E[X], \quad (8.17)$$

und als Wert der Zielfunktion ergibt sich

$$\begin{aligned} \text{MSE}(\hat{f}^{\text{so}}) &= E_{\text{id}}[|X - E_{\text{id}}[X|Y]|^2] + E_{\text{id}}[|E_{\text{id}}[X|Y] - \hat{f}^{\text{so}}|^2] = \\ &= E_{\text{id}}[|X - E_{\text{id}}[X|Y]|^2] + E_{\text{id}}[(|E_{\text{id}}[X|Y] - E[X]| - b/r)_+^2]. \end{aligned} \quad (8.18)$$

Im folgenden arbeiten wir mit (4) und setzen

$$D(Y) := E_{\text{id}}[X|Y] - E_{\text{id}}[X] \quad (8.19)$$

8.2.2 Lösung im ursprünglichen Modell – der L5-Filter

Identifizieren wir $\beta_t \rightsquigarrow X$ durch $Y_{1:t} \rightsquigarrow Y$, $b_t \rightsquigarrow br$, so erhalten wir einen robustifizierten Filter, den wir *L5-Filter* nennen wollen. Im Gaußschen Zustandsraummodell mit $E\beta_0 = a_0$ lautet der L5-Filter⁵

$$\beta_{t|t}^{\text{L5}} = F^t a_0 + (\beta_{t|t}^{\text{KK}} - F^t a_0) \min\left\{1, \frac{b_t}{|\beta_{t|t}^{\text{KK}} - F^t a_0|}\right\} \quad (8.20)$$

mit $\beta_{t|t}^{\text{KK}}$ dem klassischen Kalman-Filter. An dieser Stelle haben wir so getan, als kämen wir aus einer Gaußschen Vergangenheit, d.h. $\Delta\beta_t^{\text{L5}} \sim \mathcal{N}_p(\mu, S)$ mit gewissen $\mu \in \mathbb{R}^p$ und $0 \preceq S = S^T \in \mathbb{R}^{p \times p}$. Dem ist aber nur für $t = 1$ so, und so müssen wir (20) als Heuristik verstehen.

Verwendet man dennoch den in (20) definierten Filter, so sieht man immerhin

Proposition 8.2.1 *Gilt die Symmetrievoraussetzung (3) aus Proposition 6.1.2, so ist der L5-Filter aus (20) unverzerrt.*

⁴wie unmittelbar durch die Umformung von (13) auf $E_{\text{id}}[|(E_{\text{id}}[X|Y] - E[X]) - (f(Y) - E[X])|^2]$ deutlich wird;

⁵im zeitinvarianten Fall, die Verallgemeinerung auf den zeitvariablen Fall ist klar;

BEWEIS: Für $t = 1$ ist $\Delta\beta_1^{L5} = \Delta\beta_1^{KK}$ und, da $\Delta\beta_1^{KK}$ linear in den symmetrisch verteilten Beobachtungen ist, gilt wie in Proposition 6.1.2 $\mathcal{L}_{\text{id}}(\Delta\beta_1^{L5}) = \mathcal{L}_{\text{id}}(-\Delta\beta_1^{L5})$; wegen der Symmetrie von ε_1 gilt damit nach Lemma 6.1.1 auch für $\Delta\beta_{1|1}^{L5} = \beta_1 - \beta_{1|1}^{L5}$ die Symmetrie $\mathcal{L}_{\text{id}}(\Delta\beta_{1|1}^{L5}) = \mathcal{L}_{\text{id}}(-\Delta\beta_{1|1}^{L5})$. Daher ist insbesondere $E[\beta_{1|1}^{L5}] = E[\beta_1]$. Die Multiplikation mit F und anschließende Faltung mit dem symmetrisch verteilten v_2 erhält die Symmetrie, so daß auch $\mathcal{L}_{\text{id}}(\Delta\beta_2^{L5}) = \mathcal{L}_{\text{id}}(-\Delta\beta_2^{L5})$. Der Induktionsschluß erfolgt mit denselben Argumenten. ////

8.2.3 Wahl von b

Wie schon beim rLS-Filter in Abschnitt 2.2.4 schlagen wir vor, das Verhältnis der MSE's als Kriterium heranzuziehen, also im Sinne eines Anscombe-Kriteriums (vgl. Anscombe (1960)) ein gewisses Verhältnis $\delta > 0$ vorzugeben, so daß

$$E_{\text{id}}[|\hat{f} - X|^2] \stackrel{!}{=} (1 + \delta) E_{\text{id}}[|X - \beta_{t|t}^{KK}|^2] = (1 + \delta) \text{tr} \Sigma_{t|t}^{KK}, \quad (8.21)$$

was sich auch schreiben läßt als

$$E_{\text{id}}[(|\beta_{t|t}^{KK} - E_{\text{id}}[X]|^2 - b_t^2)_+] \stackrel{!}{=} \delta \text{tr} \Sigma_{t|t}^{KK} \quad (8.22)$$

8.2.4 Beispiele zum L5-Filter

Wir beschränken uns hier auf ein zweidimensionales Beispiel aus einem stationären Zustandsraum. Im eindimensionalen würden alle Filterwerte vom Betrag größer b_t auf $\pm b_t$ gestutzt, was wegen der Stationarität ($|F| < 1$) in einer Graphik auf zwei Schwellen $\pm b$ hinausläuft, über die der L5-Filter nicht hinwegkommt. Im mehrdimensionalen gibt es noch Kompensationen zwischen den Dimensionen.

Beispiel 8.2.2 Wir unterstellen (V4) und setzen $p = q = 2$,

$$\begin{aligned} F &= \begin{pmatrix} 0.5 & 0.3 \\ 0.6 & 0.5 \end{pmatrix} & Z &= \begin{pmatrix} 0.46 & 0.25 \\ 0.55 & 0.46 \end{pmatrix} \\ Q &= \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} & V &= \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} & a_0 &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{aligned} \quad (8.23)$$

Als Kontamination verwenden wir eine AO-Kontamination von

$$\hat{P}^\varepsilon = 0.9\mathcal{N}_2(0, V) + 0.1\mathcal{N}_2(0, \begin{pmatrix} 100 & 80 \\ 80 & 100 \end{pmatrix}).$$

Der L5-Filter ist auf ein δ von 20% kalibriert, was einem b von 1.27 entspricht. Unter diesen Annahmen wurde im idealen respektive im kontaminierten Modell Abbildung 8.1 erzeugt.

8.2.5 Bewertung — L5-Filter vs. rLS-Filter

Der L5-Filter leistet eine einfache, unverzerrte Robustifizierung des Kalman-Filters gegen AO's/SO's.

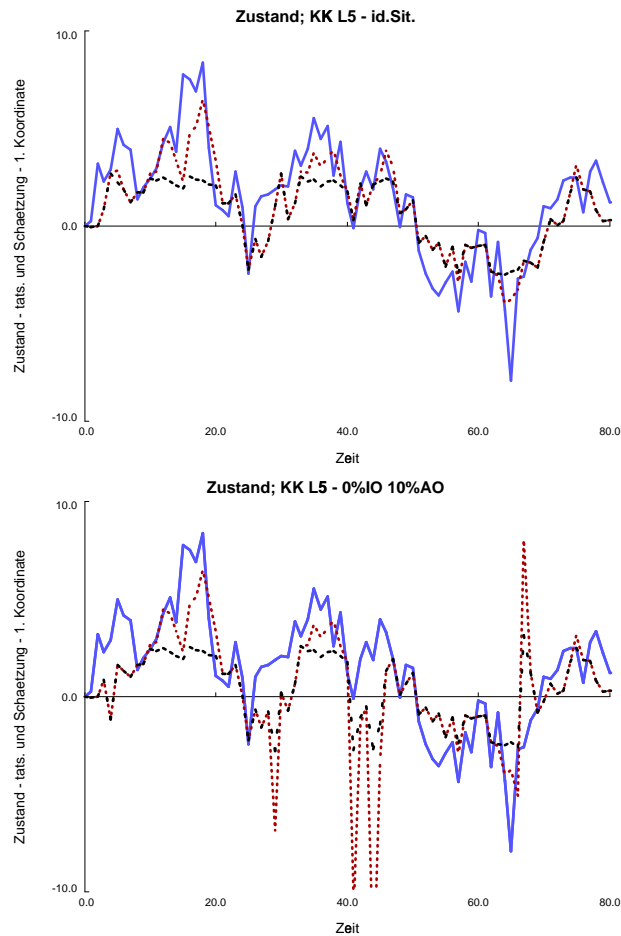


Abbildung 8.1: der L5-Filter in Beispiel 8.2.2

durchgezogen in grau β_t , gepunktet $\beta_{t|t}^{KK}$, gestrichelt in schwarz $\beta_{t|t}^{L5}$

Bei dieser Definition machen wir allerdings keinerlei Gebrauch von der rekursiven Modellstruktur, die aus der Zustandsgleichung $\beta_t = F_t \beta_{t-1} + v_t$ hervorgeht. Insbesondere nimmt man im Gaußschen Fall keine Rücksicht auf eventuelle Ausreißer bereits in der σ -Algebra $\sigma(Y_1, \dots, Y_t)$. Dies äußert sich unter anderem darin, daß im zeitinvarianten Fall $|\beta_{t|t}|$ beschränkt in t ist, was dem Fall $\rho(F) > 1$ sicher nicht angemessen ist. Insofern handelt es sich um ein nur begrenzt einsetzbares Verfahren, und insgesamt erweist sich der rLS-Ansatz als geeigneter, indem er auf die Innovationsterme $\Delta\beta$ bzw. Δy anstelle von β und y zurückgreift.

Argumentieren wir wieder heuristisch wie in (20), das heißt unterstellen wir eine Gaußsche Vergangenheit des Filters, also hier von $\Delta\beta_t$, so erhalten wir den rLS-Filter als optimal-robuste Lösung im "Lemma 5"-Ansatz. Da die Annahme der Normalität für $t = 1$ zutrifft, handelt es sich in diesem Zeitpunkt tatsächlich um das Optimum — vgl. Proposition 2.3.1.

8.2.6 Verfügbarkeit

ISP-Routinen zur Bestimmung des L5-Filters sowie zu dessen Kalibrierung können bezogen werden unter

<http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/diss/ispmacros/L5>

8.3 Minimaxproblem für SO

In der Arbeit von Birmiwal und Shen (1993) wird analog zu Huber (1964) ein Sattelpunktsresultat im Kontext des Filterproblems — zunächst allgemein und dann im Zustandsraum — unter r -Konvexkombinationen als SO-Kontamination hergeleitet.

Um hier die Essenz des Ansatzes herauszuarbeiten, wollen wir zunächst das Problem im reduzierten Modell (1) lösen.

Zusätzlich zu Birmiwal und Shen geben wir hier gleich die Lösung für $p \geq 1$ an und füllen einige “mathematische Lücken” im dort angegebenen Beweis.

8.3.1 Problemstellung

Ziel ist es, eine Minimallösung $f_0(\hat{Y})$ für die Schätzung von X aufgrund von \hat{Y} zu finden, für die gilt

$$\sup_{\mathcal{U}_r} E_P[|X - f_0(\hat{Y})|^2] = \inf_f \sup_{\mathcal{U}_r} E_P[|X - f(\hat{Y})|^2] \quad (8.24)$$

mit \mathcal{U}_r aus Formel (8).

8.3.2 Lösung

Um eine Lösung finden zu können, müssen wir die technische Voraussetzung machen, daß die Maße $P^\varepsilon(\cdot - t)$ für $t \in \text{supp}(P^X)$ eine dominierte Familie bilden. Dies ist insbesondere der Fall, falls P^ε Lebesgue-stetig ist oder P^X nur Masse auf diskret viele Punkte wirft. Dann aber gilt

Theorem 8.3.1 *Seien $X \sim P^X$, $\varepsilon \sim P^\varepsilon$, stochastisch unabhängige \mathbb{R}^p -wertige Zufallsgrößen, wobei*

$$\exists K \in \mathcal{M}^\sigma(\mathbb{B}^p) : \forall t \in \text{supp}(P^X) \quad P^\varepsilon(\cdot - t) \ll K, \quad (8.25)$$

also die Maße $P^\varepsilon(\cdot - t)$ für $t \in \text{supp}(P^X)$ eine dominierte Familie bilden.

*Weiter sei das Paar (\hat{P}_0, f_0) , $P_0 \in \mathcal{U}_r$ mit $\hat{P}_0^Y = (1-r)P^Y + r\tilde{P}_0^Y$ als zweiter Randverteilung, $P^Y = P^X * P^\varepsilon$ und f_0 eine \hat{Y} -meßbare Funktion, definiert als*

$$\tilde{P}_0^Y(dy) := \frac{1-r}{r} \left(\tilde{\rho} \left| E_{\text{id}}[X|Y=y] - E_{\text{id}}[X] \right| - 1 \right)_+ P^Y(dy) \quad (8.26)$$

und

$$f_0 := E_{\text{id}}[X] + (E_{\text{id}}[X|Y] - E_{\text{id}}[X]) \min\left\{1, \frac{1/\tilde{\rho}}{|E_{\text{id}}[X|Y] - E_{\text{id}}[X]|}\right\}, \quad (8.27)$$

wobei $\tilde{p} > 0$ eine reelle Zahl ist, die gewährleistet, daß \tilde{P}_0^Y ein Wahrscheinlichkeitsmaß ist.

Dann bildet dieses Paar einen Sattelpunkt für Problem (24).

BEWEIS ZU THEOREM 8.3.1: Wir betrachten zunächst das Problem

$$E_P[|X - f(\hat{Y})|^2] = \max_{\mathcal{U}_r} \min_f! \quad (8.28)$$

und weisen dann nach, daß

$$\max_{\mathcal{U}_r} \min_f E_P[|X - f(\hat{Y})|^2] = \min_f \max_{\mathcal{U}_r} E_P[|X - f(\hat{Y})|^2] \quad (8.29)$$

Offenbar ist die Lösung von $E_P[|X - f(\hat{Y})|^2] = \min_f!$ für festes P gerade $E_P[X|\hat{Y}]$ mit

$$\text{MSE}_P(E_P[X|\hat{Y}]) = E_{\text{id}}[|X|^2] - E_P[|E_P[X|\hat{Y}]|^2], \quad (8.30)$$

so daß wir für die Lösung von (28) das $\hat{P} \in \mathcal{U}_r$ finden müssen, das

$$\hat{P}_0 = \underset{\mathcal{U}_r}{\text{argmin}}(E_P[|E_P[X|\hat{Y}]|^2]) \quad (8.31)$$

löst.

[Beweis wird auf Seite 122 fortgesetzt]

Bemerkung 8.3.2 (a) Bei Birnival und Shen wird ein Minimax-Satz von Ferguson (1967) als Begründung für den Schluß (29) von (28) auf (24) angeführt; dieser ist aber nur im endlich-dimensionalen gültig. Wir werden im Zuge des Beweises ein direktes Argument liefern.

(b) Wir könnten für den Schluß (29) von (28) auf (24) aber auch stattdessen folgenden Minimax-Satz⁶ anwenden:

MINIMAX-THEOREM: Seien X, Y Teilmengen zweier Vektorräume, X kompakt, konvex und Y konvex. Sei $f : X \times Y \rightarrow [0; \infty]$ eine Funktion, so daß

- (i) $f(\cdot, y)$ konkav für alle $y \in Y$,
- (ii) $f(x, \cdot)$ konvex für alle $x \in X$ und
- (iii) $f(\cdot, y)$ oberhalbstetig für alle $y \in Y$,

Dann gilt

$$\inf_y \sup_x f(x, y) = \sup_x \inf_y f(x, y)$$

Wir identifizieren die Zielfunktion $f(x, y) \rightsquigarrow E_P[|f - X|^2]$ und ihre Argumente $x \in X \rightsquigarrow P \in \mathcal{U}_r \subset \mathcal{M}_b(\mathbb{B}^{2p})$, $y \in Y \rightsquigarrow f \in \{g | g : \mathbb{R}^p \rightarrow \mathbb{R}^p \text{ meßbar in } Y\}$. Dazu fassen wir die Maße P als Elemente von $\mathcal{M}_b(\mathbb{B}^{2p})$ auf der Kompaktifizierung $\bar{\mathbb{R}}^{2p}$ des \mathbb{R}^{2p} auf versehen mit der vagen Topologie. In der Y -Komponente identifizieren wir P^Y -fast sicher übereinstimmende Funktionen, P^Y jeweils die Randverteilung zu den Koordinaten $p + 1, \dots, 2p$. Für die Oberhalbstetigkeit in P verweisen wir auf Proposition 4.2 aus Franke und Poor (1984): Wir identifizieren in deren Notation $Y \rightsquigarrow X$ und $X \rightsquigarrow Y$, so daß deren Bedingung (i) bei uns lautet, $\int |X|^2 dP = \text{const}$ für alle $P \in \mathcal{U}_r$ und daher erfüllt ist, weil wir keine Kontamination in X betrachten. In P ist die Zielfunktion linear, also konkav. Die Konvexität in f ist als Mittelung des quadratischen Integranden klar.

⁶Beweis: Kneser (1952) — zitiert mit der Vertauschung konvex-konkav aus Millar (1983), Thm. 1.3

(c) Birimiwal und Shen zeigen den Satz für den Fall $p = 1$ und nehmen dabei an, P^X , P^ε und \tilde{P}^Y seien Lebesgue-stetig. In unserem Setup verlangen wir (25) nur im idealen Modell und lassen beliebige Kontaminationen \tilde{P}^Y zu, weiterhin ist die Dimension p von X beliebig.

(d) Birimiwal und Shen gehen zunächst von der Existenz eines Lagrangemultiplikators aus, der die Bedingung $\tilde{P}_0^Y(\mathbb{R}) = 1$ auffängt, und schließen nach Feststellung der Lösung mit der Bemerkung, daß die Existenz und Eindeutigkeit des Lagrangemultiplikators klar sei. Wir werden die Frage der Existenz vorneweg klären.

(e) Bei den Suprema/Infima in (24) und (31) scheint man sich zunächst auf Suprema/Infima über einer abzählbaren Teilmenge U' von \mathcal{U}_r beschränken zu müssen, um die Existenz eines dominierenden Maßes K zu bekommen. Später werden wir anhand der Lösung einsehen, daß diese Einschränkung nicht nötig ist, weil die Lösung nicht von K abhängt, und daher das Supremum tatsächlich über die "volle Umgebung" \mathcal{U}_r gebildet werden kann.

(f) Wie im Lokationsproblem fallen die "Lemma 5"-Lösung \hat{f}^{so} zu Problem (2), (3) und die Minimallösung f_0 zu Problem (24) zusammen.

(g) In Birimiwal und Papantoni-Kazakos (1994) wird Theorem 8.3.1 im Spezialfall des reduzierten Modells (1) noch einmal bewiesen, die Punkte (a), (c) und (d) gelten aber auch für diese Arbeit.

FORTSETZUNG BEWEIS ZU THEOREM 8.3.1: Sei K ein σ -finites Maß auf \mathbb{B}^p , das P^X , $P^\varepsilon(\cdot - t)$ für alle $t \in \text{supp}(P^X)$, sowie eine Menge $U' \subset \mathcal{M}_1(\mathbb{B}^p)$ von kontaminierenden Maßen \tilde{P}^Y dominiert, die bei der Definition der $P \in \mathcal{U}_r$ auftreten, und p^X , $p^\varepsilon(\cdot - t)$, \tilde{p}^Y die entsprechenden K -Dichten. Dann gilt für $\hat{Y} \sim \hat{P}^Y = (1-r)P^Y + r\tilde{P}^Y$

$$p^Y(y) := \int_{\mathbb{R}^p} p^\varepsilon(y-x)p^X(x)K(dx), \quad (8.32)$$

so daß für alle $A \in \mathbb{B}^p$

$$\begin{aligned} \Pr(\hat{Y} \in A) &= \int_{\mathbb{R}^p \times A} r\tilde{p}^Y(y)p^X(x) + (1-r)p^\varepsilon(y-x)p^X(x)K(dy)K(dx) \\ &= \int_A r\tilde{p}^Y(y) + (1-r)p^Y(y)K(dy) \end{aligned} \quad (8.33)$$

und für alle $B \in \mathbb{B}^{2p}$

$$\Pr((X, \hat{Y}) \in B) = \int_B r\tilde{p}^Y(y)p^X(x) + (1-r)p^\varepsilon(y-x)p^X(x)K(dy)K(dx) \quad (8.34)$$

Daher gilt für alle $A \in \mathbb{B}^p$ $\hat{P}^Y(dy)$ -fast-sicher

$$\Pr(X \in A | \hat{Y} = y) = \frac{\int_{x \in A} r\tilde{p}^Y(y)p^X(x) + (1-r)p^\varepsilon(y-x)p^X(x)K(dx)}{r\tilde{p}^Y(y) + (1-r)p^Y(y)} \quad (8.35)$$

und somit $\hat{P}^Y(dy)$ -fast-sicher

$$\begin{aligned} E_P[X | \hat{Y} = y] &= \frac{\int x[r\tilde{p}^Y(y)p^X(x) + (1-r)p^\varepsilon(y-x)p^X(x)]K(dx)}{r\tilde{p}^Y(y) + (1-r)p^Y(y)} \\ &= \frac{r E_{\text{id}}[X]\tilde{p}^Y(y) + (1-r) E_{\text{id}}[X | \hat{Y} = y]p^Y(y)}{r\tilde{p}^Y(y) + (1-r)p^Y(y)}, \end{aligned} \quad (8.36)$$

sowie

$$\mathbb{E}_P[|\mathbb{E}_P[X|\hat{Y}]|^2] = \int \frac{|r\mathbb{E}_{\text{id}}[X]\tilde{p}^Y(y) + (1-r)\mathbb{E}_{\text{id}}[X|\hat{Y}=y]p^Y(y)|^2}{r\tilde{p}^Y(y) + (1-r)p^Y(y)} K(dy) \quad (8.37)$$

Setzt man im Integranden $a := r\mathbb{E}_{\text{id}}[X]$, $b(y) := (1-r)\mathbb{E}_{\text{id}}[X|\hat{Y}=y]p^Y(y)$, $c := r$, $d(y) := (1-r)p^Y(y)$, so lautet der Integrand

$$I(\tilde{p}^Y, y) := \frac{|a\tilde{p}^Y(y) + b(y)|^2}{c\tilde{p}^Y(y) + d(y)} \quad (8.38)$$

und man hat zu lösen

$$\hat{h} \stackrel{!}{=} \operatorname{argmin}_{h \in \mathcal{H}} \left\{ \int I(h, y) K(dy) \right\}, \quad (8.39)$$

$$\mathcal{H} := \{h \in L_1(K) : h \geq 0 [K], \int h dK = 1\}. \quad (8.40)$$

Dabei ist $\int I(h, y)$ konvex in h mit $h \geq 0 [K]$ für alle y , denn

$$\frac{\partial}{\partial h}(I(h, y)) = \frac{2a^\tau(ah + b)}{ch + d} - \frac{c|ah + b|^2}{(ch + d)^2}, \quad (8.41)$$

$$\frac{\partial^2}{\partial h^2}(I(h, y)) = \frac{2}{ch + d} \left| a - \frac{c(ah + b)}{ch + d} \right|^2 \geq 0. \quad (8.42)$$

Also ist (39) ein konvexes Optimierungsproblem, auf das wir Theorem B.2.3 in Rieder (1994) anwenden können:

Mit den Identifikationen $X \rightsquigarrow L_1(K)$, $Y \rightsquigarrow \mathbb{R}$, $A \rightsquigarrow \{h \in L_1(K) : h \geq 0 [K]\}$, $f \rightsquigarrow \int I dK$, $H \rightsquigarrow h \mapsto \int h(x) K(dx)$ und $y_0 \rightsquigarrow 1$ folgt:

f ist als Mittel der nach (42) auf A konvexen $I(h, y)$ tatsächlich konvex. Die Funktion $p^Y(y)$ aus (32) ist zulässig, d.h. $p^Y(y) \in A$, $H(p^Y(y)) = y_0$, mit Funktionswert $\mathbb{E}_{\text{id}}[|\mathbb{E}_{\text{id}}[X|\hat{Y}]|^2] < \infty$, und $f \geq 0$; also ist das Infimum $\inf f$ über alle zulässigen x endlich.

Mit Polynomdivision erhalten wir

$$\begin{aligned} I(h, y) &= h|a|^2/c + 2a^\tau b/c - |a|^2 d/c^2 + \frac{|b - ad/c|^2}{ch + d} \leq \\ &\leq h|a|^2/c + 2a^\tau b/c - a^2 d/c^2 + \frac{|bc - ad|^2}{c^2 d} \end{aligned} \quad (8.43)$$

und daher

$$\begin{aligned} &|\int I(h, y) K(dy)| \leq \\ &\leq \|h\|_{L_1(K)} |a|^2/c + 2|a|^2/c + |a|^2/c^2 + \int \frac{|bc - ad|^2}{c^2 d} K(dy) = \\ &= |\mathbb{E}_{\text{id}}[X]|^2 (\|h\|_{L_1(K)} + 2/r + 1/r^2) + (1-r) \mathbb{E}_{\text{id}} \left[|\mathbb{E}_{\text{id}}[X|\hat{Y}] - \mathbb{E}_{\text{id}}[X]|^2 \right]. \end{aligned}$$

Es folgt, daß für

$$V := \{h \in L_1(K) : \|h\|_{L_1(K)} \leq 2\}$$

die Slater-Bedingung für unendlich-dimensionale Probleme gilt:

$$1 \in \{x \in \mathbb{R} : \exists h \in V : x = \int h dK\}^\circ \neq \emptyset, \quad \sup\{\int I dK \mid h \in V\} < \infty.$$

mit A° dem offenen Kern von A . Damit ist Problem (39) wohlgestellt (“well posed”). Es gibt also einen Lagrangemultiplikator $\rho \in \mathbb{R}$, so daß das optimale $\hat{h} = \hat{p}_0^Y$ auch die Lagrangefunktion

$$L(h, \rho) := \int I(h, y) + \rho h K(dy)$$

unter allen $h \in \mathcal{H}$ minimiert.

Dieses Problem wollen wir nun lösen, indem wir punktweise in y den Integranden $I(h, y) + \rho h$ in $h \in \mathbb{R}_{\geq 0}$ minimieren. Dazu verwenden wir Lemma C.2.2 und nehmen im folgenden an, daß $\rho > -|a|^2/c$. Dann ist $I(h, y)$ stetig in h auf \mathbb{R} , und aus (43) folgt $\lim_{h \rightarrow \pm\infty} I(h, y) + \rho h = \pm\infty$, so daß also $I(h, y)$ in h auf $\bar{\mathbb{R}}$ stetig ist. Seien

$$h_{1,2} := \frac{1}{c} \left(-d \pm \frac{|ad - bc|}{\sqrt{|a|^2 + \rho c}} \right).$$

Dann sind $h_{1,2}$ die einzigen Nullstellen von $\frac{\partial}{\partial h}(I(h, y) + \rho h)$. Außerdem ist

$$\frac{\partial^2}{\partial h^2}(I(h, y) + \rho h)|_{h=h_{1,2}} = \pm \frac{2\sqrt{|a|^2 + \rho c}}{|ad - bc|} \left| a - \frac{c(ah_{1,2} + b)}{ch_{1,2} + d} \right|^2,$$

so daß h_1 das absolute Minimum von $I(h, y) + \rho h$ in $B_y = [\min\{h_1, 0\}; \infty]$ ist. Setzen wir $A_y := [0, \infty]$, dann ist Lemma C.2.2 in Kraft mit $A \rightsquigarrow A_y$ und $B \rightsquigarrow B_y$, und es gilt punktweise in y , daß

$$\operatorname{argmin}_{h \geq 0 [K]}(I(h, y) + \rho h) = \left(\operatorname{argmin}_{h \in L_1(K)}(I(h, y) + \rho h) \right)_+.$$

Die Lagrangefunktion $L(h, \rho)$ wird daher punktweise in $h \geq 0 [K]$ minimiert durch

$$\hat{h} = \frac{1}{c} \left(\frac{|da - cb|}{\sqrt{|a|^2 + \rho c}} - d \right)_+ \quad (8.44)$$

oder nach Rücksubstitution

$$\begin{aligned} \hat{p}_0^Y &= \frac{1-r}{r} p^Y(y) \left(\tilde{\rho} \left| \mathbb{E}_{\text{id}}[X|\hat{Y} = y] - \mathbb{E}_{\text{id}}[X] \right| - 1 \right)_+ = \\ &= \frac{1-r}{r} p^Y(y) (\tilde{\rho} |D(y)| - 1)_+ \end{aligned} \quad (8.45)$$

wobei ρ , $\tilde{\rho} = r/\sqrt{|a|^2 + \rho r}$ jeweils für die Normierung als Wahrscheinlichkeitsmaß sorgen. Insbesondere ist tatsächlich $\rho = r/\tilde{\rho}^2 - |a|^2/c > -|a|^2/c$.

In Verbindung mit dem dominierenden Maß K erhalten wir

$$\tilde{P}_0^Y(dy) = \frac{1-r}{r} P^Y(dy) (\tilde{\rho} |D(y)| - 1)_+ \quad (8.46)$$

und damit

$$\hat{P}_0^Y(dy) = \frac{1-r}{r} P^Y(dy) \max(\tilde{\rho} |D(y)|, 1) \quad (8.47)$$

Dies führt auf den Schätzer

$$f_0(y) := \mathbb{E}_{\hat{P}}[X|\hat{Y} = y] \stackrel{[P^Y]}{=} \mathbb{E}_{\text{id}}[X] + D(y) \min\left\{1, \frac{1/\tilde{\rho}}{|D(y)|}\right\}, \quad (8.48)$$

denn

$$\begin{aligned}
f_0(y) &\stackrel{[K]}{=} \frac{a\hat{h} + b}{c\hat{h} + d} = \frac{(1-r)p^Y(y) \left(\mathbb{E}_{\text{id}}[X](\tilde{\rho}|D(y)| - 1)_+ + \mathbb{E}_{\text{id}}[X|\hat{Y}] \right)}{(1-r)p^Y(y) \left((\tilde{\rho}|D(y)| - 1)_+ + 1 \right)} = \\
&\stackrel{[P^Y]}{=} \frac{\mathbb{E}_{\text{id}}[X] \{ (\tilde{\rho}|D(y)| - 1)_+ + 1 \} + \mathbb{E}_{\text{id}}[X|\hat{Y}] - \mathbb{E}_{\text{id}}[X]}{(\tilde{\rho}|D(y)| - 1)_+ + 1} = \\
&= \mathbb{E}_{\text{id}}[X] + \frac{D(y)}{(\tilde{\rho}|D(y)| - 1)_+ + 1} = \mathbb{E}_{\text{id}}[X] + D(y) \min\left\{1, \frac{1/\tilde{\rho}}{|D(y)|}\right\}
\end{aligned}$$

Für den MSE gilt

$$\text{MSE}_{\hat{P}_0}[f_0] = \mathbb{E}_{\hat{P}_0} |X - \mathbb{E}_{\hat{P}_0}[X|\hat{Y}]|^2 = \text{tr Cov}_{\text{id}} X - \text{tr Cov}_{\hat{P}_0}(\mathbb{E}_{\hat{P}_0}[X|\hat{Y}]),$$

so daß

$$\begin{aligned}
&\text{tr Cov}_{\text{id}}(X) - \text{MSE}_{\hat{P}_0}[f_0] = \mathbb{E}_{\hat{P}_0} [|\mathbb{E}_{\hat{P}_0}[X|\hat{Y}] - \mathbb{E}_{\text{id}}[X]|^2] = \\
&= (1-r) \int \min\{|D(y)|^2, \frac{1}{\tilde{\rho}^2}\} p^Y(y) \left((\tilde{\rho}|D(y)| - 1)_+ + 1 \right) K(dy) = \\
&= (1-r) \mathbb{E}_{\text{id}} \left[|D(\hat{Y})| \min\{|D(\hat{Y})|, 1/\tilde{\rho}\} \right]
\end{aligned}$$

also

$$\text{MSE}_{\hat{P}_0}[f_0] = \text{tr Cov}_{\text{id}} X - (1-r) \mathbb{E}_{\text{id}} \left[|D(\hat{Y})| \min\{|D(\hat{Y})|, 1/\tilde{\rho}\} \right] \quad (8.49)$$

Offenbar hängen weder f_0 in (48), \tilde{P}_0^Y in (46) noch $\text{MSE}_{\hat{P}_0}[f_0]$ in (49) von der Wahl des dominierenden Maßes K aus (25) und daher von der speziellen Wahl einer abzählbaren Teilmenge $U' \subset \mathcal{U}_r$ ab, längs der optimiert werden soll. Weiter gilt für beliebiges $\tilde{P} \in \mathcal{U}_r$ mit kontaminierendem Maß \tilde{P}^Y

$$\begin{aligned}
\text{MSE}_{\tilde{P}}(f_0) &:= \mathbb{E}_{\tilde{P}}[|f_0 - X|^2] = (1-r) \mathbb{E}_{\text{id}}[|f_0 - X|^2] + r \mathbb{E}_{\text{cont}}[|f_0 - X|^2] = \\
&= \text{const} + r \mathbb{E}_{\text{cont}}[|f_0 - X|^2] = \text{const} + r \mathbb{E}_{P \times \otimes \tilde{P}^Y}[|f_0 - X|^2] = \\
&= \text{const} + r(-2 \mathbb{E}_{\text{id}}[X]^T \mathbb{E}_{\tilde{P}^Y}[f_0] + \mathbb{E}_{\tilde{P}^Y}[|f_0|^2]) = \\
&= \text{const} + r |\mathbb{E}_{\tilde{P}^Y}[f_0] - \mathbb{E}_{\text{id}} \mathbb{E}[X]|^2 \\
&= \text{const} + r \mathbb{E}_{\tilde{P}^Y} \left[\min\{|D(\hat{Y})|^2, \frac{1}{\tilde{\rho}^2}\} \right].
\end{aligned}$$

Also wird das Maximum in \tilde{P}^Y genau dann erreicht, falls \tilde{P}^Y seine gesamte Masse auf $\{|D(\hat{Y})| \geq \frac{1}{\tilde{\rho}}\}$ wirft, was ja auch \tilde{P}_0^Y tut, so daß gilt

$$\begin{aligned}
\sup_{P \in \mathcal{U}_r} \mathbb{E}_P[|f_0 - X|^2] &= \text{tr Cov}_{\text{id}} X - (1-r) \mathbb{E}_{\text{id}} \left[|D(\hat{Y})| \min\{|D(\hat{Y})|, 1/\tilde{\rho}\} \right] = \\
&= \text{MSE}_{\tilde{P}}[f_0]
\end{aligned} \quad (8.50)$$

Damit ist tatsächlich

$$\sup_{\mathcal{U}_r} \mathbb{E}_P[|f_0 - X|^2] \leq \max_{\mathcal{U}_r} \min_f \mathbb{E}_P[|X - f(\hat{Y})|^2]$$

und es gilt (29). ////

Abschließend stellen wir in Abbildung 8.2 für den Fall $\varepsilon, X \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}(0, 1)$, $r = 0.1$ die Dichten p^Y , \hat{p}_0^Y und \tilde{p}_0^Y graphisch dar.

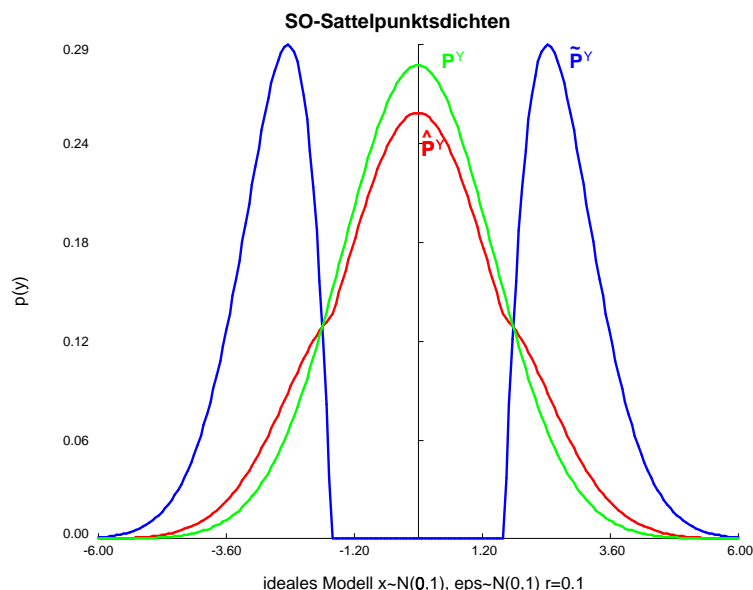


Abbildung 8.2: die Dichten p^Y , \hat{p}_0^Y und \tilde{p}_0^Y

Im Fall $\varepsilon, X \stackrel{\text{i.i.v.}}{\sim} \mathcal{N}(0, 1)$ stellen wir für Radius $r = 0.1$ die ideale Dichte von $Y = X + \varepsilon$, p^Y , die Dichte der ungünstigsten SO-Kontamination \tilde{p}_0^Y , sowie die entstehende reale Dichte $\hat{p}_0^Y = (1 - r)p^Y + r\tilde{p}_0^Y$ dar. Die Verteilung \hat{P}^Y wirft überraschend wenig Masse auf die Flanken.

8.4 Lösung im ursprünglichen Modell

8.4.1 das ursprüngliche Problem im reduzierten Modell

Will man nun diese Lösung im ursprünglichen Zustandsraummodell verwenden, das heißt

$$E_{\text{id}}[|\Delta\beta_t - f_t|^2] = \min_{f_t} \quad (8.51)$$

für

$$f_t = f_t(\hat{Y}_1, \dots, \hat{Y}_t) \in L_2(P^{\Delta\beta_t}) \quad (8.52)$$

unter SO-Kontamination lösen, so wird man sich, um das Problem handhabbar zu halten, auf rekursiv definierte Filter beschränken müssen, die jeweils in $\Delta y_t = Z\Delta\beta_t + \varepsilon_t$ meßbar sind, wobei $\Delta\beta_t$ der Prognosefehler $\beta_t - F\Delta\beta_{t-1|t-1}$ und $\Delta\beta_{t-1|t-1}$ der mit dem bisherigen Verfahren ermittelte Filterwert zum Zeitpunkt $t - 1$ ist.

Dann aber können wir in Theorem 8.3.1 die Größen X mit $\Delta\beta_t$ bzw. Y mit Δy_t identifizieren, wobei wir die Terme $y - x$, die in die Dichten p^ε einzusetzen sind, durch $\Delta y_t - Z\Delta\beta_t$ ersetzen.

Proposition 8.4.1 *Wenn man für den Aufbau der $\Delta\beta_t$ aus der Vergangenheit heraus das Verfahren*

$$f_t(Y_1, \dots, Y_t) = f(\Delta y_t) = E_{\text{id}}[\Delta\beta_t] + D(\Delta y_t) \min\left\{1, \frac{1/\tilde{\rho}}{|D(\Delta y_t)|}\right\} \quad (8.53)$$

verwendet, gilt:

(a) Ist die Symmetrievoraussetzung (3) aus Proposition 6.1.2 in Kraft, so ist der Filter, definiert durch (53), im idealen Modell biasfrei, denn es gilt

$$E_{\text{id}}[\Delta\beta_t^{\text{so}} - f_t] = 0.$$

(b) Falls $\mathcal{L}(v_t)$, $\mathcal{L}(\varepsilon_t)$ im Zustandsraummodell Lebesgue–stetig sind, so sind auch $\mathcal{L}(\Delta\beta_t^{\text{so}})$ und $\mathcal{L}(\Delta y_t^{\text{so}})$ Lebesgue–stetig. Dann ist die Dominierteitsvoraussetzung (25) in Theorem 8.3.1 mit $K = \lambda^p$ erfüllt.

(a) und (b) gelten insbesondere unter den Normalverteilungsvoraussetzungen (V3) bzw (V4).

BEWEIS: (a): wie Proposition 6.1.2 zeigen wir, daß $\mathcal{L}_{\text{id}}(\Delta\beta_1^{\text{so}}) = \mathcal{L}_{\text{id}}(-\Delta\beta_1^{\text{so}})$ für $t = 1$ gilt; wie für jedes t gilt $\mathcal{L}_{\text{id}}(\varepsilon_1) = \mathcal{L}_{\text{id}}(-\varepsilon_1)$. Damit ist nach Lemma 6.1.1 $\mathcal{L}_{\text{id}}(E[\Delta\beta_1^{\text{so}} | \Delta y_1^{\text{so}}]) = \mathcal{L}_{\text{id}}(-E[\Delta\beta_1^{\text{so}} | \Delta y_1^{\text{so}}])$, und da das Clipping symmetrisch um den idealen Erwartungswert erfolgt, gilt auch für $\Delta\beta_{1|1}^{\text{so}} = \beta_1 - \beta_{1|1}^{\text{so}}$ die Symmetrie $\mathcal{L}_{\text{id}}(\Delta\beta_{1|1}^{\text{so}}) = \mathcal{L}_{\text{id}}(-\Delta\beta_{1|1}^{\text{so}})$. Daher ist insbesondere $E[\beta_{1|1}^{\text{so}}] = E[\beta_1]$. Die Multiplikation mit F und anschließende Faltung mit dem symmetrisch verteilten v_2 erhält die Symmetrie, so daß auch $\mathcal{L}_{\text{id}}(\Delta\beta_2^{\text{so}}) = \mathcal{L}_{\text{id}}(-\Delta\beta_2^{\text{so}})$. Der Induktionsschluß erfolgt mit denselben Argumenten.

(b) folgt, weil $\mathcal{L}_{\text{id}}(\Delta\beta_t)$ eine Faltung einer Transformation von $\Delta\beta_{t-1}$ mit v_t und $\mathcal{L}_{\text{id}}(\Delta y_t)$ eine Faltung von $Z\Delta\beta_t$ mit ε_t ist. Für Lebesgue–stetige Verteilungen von ε gilt die Dominierteitsvoraussetzung (25) wegen der Translationsinvarianz des Lebesgue–Maßes. ////

Im Gaußschen Zustandsraummodell stehen wir damit vor der Schwierigkeit,

$$D(\Delta y_t) = E_{\text{id}}[\Delta\beta_t | \Delta y_t] \tag{8.54}$$

exakt zu berechnen.

8.4.2 sequentielles vs. rekursives Spiel

Betrachten wir nun für $t > 0$ folgendes sequentielles Optimierungsproblem: Gesucht sei die Lösung \hat{f} von Problem (Pb)_t

$$(Pb)_t :: \sup_{\mathcal{U}_{r,t}} E_P[|\Delta\beta_t - \hat{f}(\Delta\hat{y}_t)|^2] \stackrel{!}{=} \inf_f \sup_{\mathcal{U}_{r,t}} E_P[|\Delta\beta_t - f(\Delta\hat{y}_t)|^2] \tag{8.55}$$

Dabei ist $\mathcal{U}_{r,t}$ die Umgebung, die sich mit $\Delta\beta_t$ aus einer robustifizierten Vergangenheit des Filters ergibt, und $\Delta\hat{y}_t = (1 - U_t)\Delta y_t + U_t\Delta\tilde{y}_t$ geht aus der SO-kontaminierten Beobachtung zum Zeitpunkt t hervor. Hierbei sind nun zwei Ansätze denkbar:

ein sequentielles Spiel

Wir partitionieren die Vergangenheit des Problems nach Realisationen $U_{1:t-1} = u_{1:t-1}$ und lösen für jeden Pfad $u_{1:t-1} \in \{0, 1\}^{t-1}$ das entsprechende Optimierungsproblem. Dieser Ansatz wird von Birmiwal und Shen (1993) verfolgt. Hierzu müssen — vgl. Birmiwal und Shen (1993) — die kontaminierenden Variablen $\Delta\tilde{y}_t$ jeweils bedingt nach $\sigma(\hat{y}_{1:t-1})$ unabhängig von $(\Delta\beta_t, \varepsilon_t)$ sein, und für alle

Realisationen $U_t = 1$ muß die realisierte Kontamination $\Delta\hat{y}_t = \tilde{y}_t$ für alle τ bedingt nach $\sigma(\hat{y}_{1:\tau})$ unabhängig von $(\Delta\beta_\tau, \varepsilon_\tau)$ sein.

Unter dieser Prämisse können wir nun — zumindest theoretisch — für die 2^{t-1} Kontaminationsvergangenheiten die Sattelpunktlösungen $(f_{0,t,i}, \tilde{P}_{0,t,i})$, $i = 1, \dots, 2^{t-1}$ ermitteln. Schreibt man i in Binärdarstellung als $i = \sum_{k=0}^{t-2} i_k 2^k$, so erhalten wir

$$f_{0,t} := \sum_{i=1}^{2^{t-1}} \prod_{k=0}^{t-2} r^{i_k} (1-r)^{1-i_k} f_{0,t,i}.$$

Für die genauere Berechnung der $f_{0,t,i}$ siehe Birmiwal und Shen (1993).

Der Aufwand allerdings, 2^{t-1} Probleme zu lösen, ist nicht vertretbar, so daß Näherungsvarianten angebracht sind.

Die allereinfachste besteht darin, wie Masreliez und Martin (1977) anzunehmen, daß weiterhin $\Delta\beta_t$ normalverteilt ist, was allerdings aus theoretischen Gründen (vgl. Abschnitt 6.3.1 und Theorem 1.3.1) nicht korrekt sein kann; dennoch erweist sich dieser Ansatz in Beispielen als Näherung praktikabel, wie in Abschnitt 2.6 gezeigt worden ist; siehe dazu auch den nächsten Abschnitt.

Eine weitere Möglichkeit besteht darin, als erste Näherung nur Prozeßvergangenheiten $(U_s)_{s < t}$ zu betrachten, bei denen $\sum_{s=t-w}^w U_s \leq 1$ ist, also maximal ein Ausreißer in den letzten w Beobachtungen aufgetreten ist. Dies entspricht der Herangehensweise von Schick (1989), Schick und Mitter (1994). Diese zeigen auch, daß bei asymptotisch exponentieller Stabilität des Kalmanfilters Ausreißer in einer weiter zurückliegenden Vergangenheit vernachlässigt werden können.

ein rekursives Spiel

Der Ansatz, den wir hier verfolgen wollen, besteht darin, in der Stelle t die genaue Prozeßvergangenheit zu vergessen, und uns auf die Untersuchung der “eindimensionalen”⁷ Randverteilung $\mathcal{L}(\Delta\beta_t)$ zu beschränken. Hierfür müssen wir im Unterschied zu Birmiwal und Shen (1993) nicht 2^t sondern “nur” t Probleme lösen.

Dazu nehmen wir an, der Kontaminationsmechanismus $U_t \stackrel{\text{u.i.v.}}{\sim} \text{Bin}(1, r)$ mit einem festen $r \in (0; 1)$ und unabhängig von den restlichen Zufallsprozessen.

Die Verteilung von $\Delta\tilde{Y}_t$ ist $\tilde{P}_{0,t}^Y$, die zur ungünstigsten Verteilung in $\mathcal{U}_{r,t}$ gehört, und $P_t^X = \mathcal{L}(\Delta\beta_t)$ und $\Delta\beta_t$ sind über die jeweiligen Sattelpunktlösungen $f_{0,s}(\Delta y_s)$ für die Probleme $(\text{Pb})_s$, $1 \leq s < t$ definiert.

Dann können Statistiker und Natur im Spiel jeweils nach jedem Schritt ungünstigste Verteilung und besten Schätzer neu wählen, wissend, daß jeder der beiden Spieler eine Minimaxstrategie verfolgt.

Versteht man unter “id” in Theorem 8.3.1 immer die Verteilung von $\Delta\beta_t, \varepsilon_t$ aufgebaut mit den Lösungen für die Probleme $(\text{Pb})_s$, $1 \leq s < t$, so kann man mit den in Abschnitt 8.4.1 angegebenen Identifikationen die Lösungen $(f_{0,t}, \tilde{P}_{0,t})$ jeweils induktiv weiter aus Theorem 8.3.1 ablesen.

Selbst dies ist aber angesichts der unter Umständen unhandlichen, “eindimensionalen” Randverteilung von $\Delta\beta_t$ noch nicht befriedigend, und so versuchen wir im folgenden, eine leicht weiterzuverwendende Näherung für $\mathcal{L}(\Delta\beta_t)$ zu finden.

⁷genauer die “ein-Block-dimensionale” Randverteilung, also die p -variate Verteilung der Koordinaten von $\Delta\beta_t$

8.4.3 Optimalität des rLS-Filters auch für $t > 1$

Aus Theorem 8.3.1 bzw. (17) lesen wir ab, daß der rLS-Filter dann und genau dann SO-optimal im Minimax- bzw. "Lemma 5"-Sinn ist, wenn der ideale bedingte Erwartungswert linear ist. Diese Eigenschaft wiederum wird in Theorem 1.3.1 charakterisiert. Für $\varepsilon_t \sim \mathcal{N}(0, V_t)$ ist dies nur für normalverteiltes $\Delta\beta_t$ der Fall, was wie oben schon erwähnt nicht der Fall ist, sobald $0 < b < \infty$. Im folgenden geben wir aber gewissermaßen ex-post eine Rechtfertigung für das Vorgehen von Schick (1989) bzw. Masreliez und Martin (1977); da, wie wir gesehen haben, der exakt-optimale, robuste, sequentielle Filter von Birmiwal und Shen (1993) für praktische Anwendungen unpraktikabel ist, schlagen wir vor, folgende Näherung durchzuführen:

ein erweitertes Kontaminationsmodell

Wir erschweren die Schätzaufgabe gegenüber der ursprünglichen SO-Situation aus Abschnitt 8.3.1 noch zusätzlich dadurch, daß auch die Verteilung von X kontaminiert werden darf; genauer schlagen wir folgendes Gross-Error-Modell zum Radius r vor:

$$\mathcal{L}(\hat{X}, \hat{Y}) = (1 - r)\mathcal{L}(X, X + \varepsilon) + r\mathcal{L}(\tilde{X}, \tilde{Y}) \quad (8.56)$$

mit \tilde{Y} , \tilde{X} , X , ε alle untereinander stochastisch unabhängig, $\tilde{Y} \sim \tilde{P}^Y$, \tilde{P}^Y beliebig wie bisher, $\tilde{X} \sim \tilde{P}^X$ mit \tilde{P}^X beliebig, nur daß $E[|\tilde{X}|^2] \leq G$ für ein $G > 0$ und $E[\tilde{X}] = E_{\text{id}}[X]$. Damit definieren wir

$$\tilde{\mathcal{U}}_r := \{\mathcal{L}(\hat{X}, \hat{Y}) \in \mathcal{M}_1(\mathbb{B}^{2p}) \text{ wie in (56)}\}. \quad (8.57)$$

Diese letzte Bedingung widerspricht zwar dem Robustheitsgedanken, ist hier aber auch nicht aus diesem Aspekt zu sehen. Vielmehr werden wir durch diese Erweiterung der Kontaminationskandidaten auch Verteilungen von $\Delta\beta_t$ aus einer rLS-Vergangenheit erfassen: Wir wollen [zumindest numerisch] zeigen, daß $\mathcal{L}(\Delta\hat{y}_t, \Delta\beta_t^{\text{LS}})$ immer in einer r -Umgebung einer Normalverteilung bleibt, die wir dann als Zentrum der Umgebung annehmen. Ausgehend von dieser Normalverteilung wird sich dann im folgenden Aufdatierungsschritt der rLS-Filter als optimal erweisen. In diesem erweiterten Kontaminationsmodell gilt:

Theorem 8.4.2 *Seien $X \sim P^X$, $\varepsilon \sim P^\varepsilon$, stochastisch unabhängige \mathbb{R}^p -wertige Zufallsgrößen, wobei die Maße $P^\varepsilon(\cdot - t)$ wie in (25) für $t \in \text{supp}(P^X)$ dominiert sind. Sei $\hat{P}'_0 = \mathcal{L}(\hat{Y}, \hat{X})$ mit*

$$\mathcal{L}(\hat{Y}) = (1 - r)(P^X * P^\varepsilon) + r\tilde{P}_0^Y, \quad \mathcal{L}(\hat{X}) = (1 - r)P^X + r\tilde{P}^X, \quad (8.58)$$

\tilde{P}_0^Y wie in (26), und \tilde{P}^X irgendein Maß auf $\mathcal{M}_1(\mathbb{B}^p)$, so daß $E[|\tilde{X}|^2] = G$ und $E[\tilde{X}] = E_{\text{id}}[X]$. Sei weiter f_0 wie in (27); dann bildet (\hat{P}'_0, f_0) einen Sattelpunkt für Problem

$$E_{\hat{P}'_0}[|X - f_0(\hat{Y})|^2] \stackrel{!}{=} \inf_f \sup_{\mathcal{U}_r} E_P[|X - f(\hat{Y})|^2] \quad (8.59)$$

BEWEIS: Zunächst lösen wir das Problem unter der Gleichheits-Nebenbedingung

$$E_{\tilde{P}^X}[|X|^2] = g.$$

Betrachten wir wieder (30) im Beweis zu Theorem 8.3.1, so sehen wir, daß wir das Problem lösen können, indem wir zunächst (31) für $P \in \tilde{U}_r$ minimieren. Dazu müssen wir nur im weiteren Verlauf des Beweises des Theorems $a = r E_{\tilde{P}_X}[\hat{X}]$ setzen, der Rest kann unverändert übernommen werden. Da aber nach Voraussetzung $E_{\tilde{P}_X}[X] = E_{\text{id}}[X]$, ändert sich nichts an der Lösung (P_0, f_0) von Theorem 8.3.1, nur der Wert des Sattelpunktes, $\text{MSE}_{P'}[f_0]$ ändert sich auf

$$\text{MSE}_{P'}[f_0] = \text{tr Cov}_{\tilde{P}_X}[\hat{X}] - (1 - r) E_{\text{id}} \left[|D(\hat{Y})| \min\{|D(\hat{Y})|, 1/\tilde{\rho}\} \right],$$

mit $\text{tr Cov}_{\tilde{P}_X}[\hat{X}] = (1 - r) \text{tr Cov}_{\text{id}}[X] + r(g - |E_{\text{id}}[X]|)^2$. Maximieren wir jetzt noch den MSE längs $g \leq G$, so erhalten wir, daß $\hat{g} = G$, und damit $E_{\tilde{P}_X}[|X|^2] = G$. ////

Demzufolge ist der rLS-Filter startend von einem normalverteilten X sogar die Minimax-Lösung unter allen $P \in \tilde{U}_r$.

der rLS-Filter in diesem Kontaminationsmodell

Als nächstes wollen wir induktiv nachweisen, daß die Verteilung des aus der rLS-Vergangenheit aufgebauten Filterfehlers — hier kurz $\Delta\beta_t^{\text{rLS}}$ — in jedem Schritt in einer solchen Umgebung \tilde{U}_r um ein normalverteiltes Zentrum $\mathcal{L}(\Delta y_t^{\mathcal{N}}, \Delta\beta_t^{\mathcal{N}})$ liegt.

Zunächst sind die Nebenbedingungen $E_{\tilde{P}_X}[X] = E_{\text{id}}[X]$ und $E_{\tilde{P}_X}[|X|^2] \leq G$ erfüllt; erstere nach Proposition 8.4.1 (b), letztere wegen der Beschränktheit von $f_{0,s}$, $s < t$ und der Endlichkeit der zweiten Momente im idealen Modell.

Es bleibt also die Inklusion zu einem gegebenen Radius r in \tilde{U}_r nachzuweisen. Wegen der Normalität der ε_t genügt es dafür zu zeigen, daß $\mathcal{L}(\Delta\beta_t^{\text{rLS}})$ jeweils in einer entsprechenden r -Umgebung zu einer Verteilung $\mathcal{L}(\Delta\beta_t^{\mathcal{N}}) = \mathcal{N}_p(0, S_t)$ mit noch zu spezifizierender Kovarianz S_t liegt.

Dazu müssen wir zunächst die Eigenschaft charakterisieren, in einer r -Konvexkontaminations-Umgebung um $\mathcal{N}_p(0, S)$ zu liegen: Hinreichend und notwendig — c.f. Rieder (1977), Lemma 4.3 — hierfür ist, daß

$$\int (q(t) - (1 - r)p^{\mathcal{N}(0,S)}(t))_+ \mu(dt) \leq r \quad (8.60)$$

$$\text{bzw. } q(t) \geq (1 - r)p^{\mathcal{N}(0,S)}(t) \quad [\mu] \quad (8.61)$$

mit q der μ -Dichte von Q , dem Maß, dessen Zugehörigkeit zu einer r -Konvexkontaminations-Umgebung um $\mathcal{N}_p(0, S)$ überprüft werden soll, und μ einem Q und $\mathcal{N}_p(0, S)$ dominierenden Maß.

Mithilfe dieser Bedingung sieht man leicht ein, daß, falls $\mathcal{L}(\Delta\beta_{t-1|t-1}^{\text{rLS}})$ in einer r -Umgebung um ein $\mathcal{L}(\Delta\beta_{t-1|t-1}^{\mathcal{N}}) = \mathcal{N}_p(0, S_{t-1|t-1})$ liegt, gleiches auch für $\mathcal{L}(\Delta\beta_t^{\text{rLS}})$ gilt mit $\mathcal{N}_p(0, S_t)$ als Zentrum und $S_t = FS_{t-1|t-1}F^T + Q$, mehr noch: Nach einer Faltung mit der unabhängigen Normalverteilung v_t kann “=” in (60) nicht mehr auftreten.

Im folgenden wollen wir den Fall $S_t = \Sigma_{t|t-1}^{\text{KK}}$, sowie für $p = q = 1$ den Fall $S_t = \Sigma_{t|t-1}^{(\delta)}$ betrachten, wobei $\Sigma_{t|t-1}^{(\delta)} = (1 + \delta)FS_{t-1|t-1}F^T + Q$, $S_{t|t}^{(\delta)} = \Sigma_{t|t-1}^{(\delta)}V/(Z^2\Sigma_{t|t-1}^{(\delta)} + V)$. Letzteres trägt der Tatsache Rechnung, daß wir, ausgehend von normalverteilten $\Delta\beta$, bei Verwendung des rLS-Filters nicht mehr den

optimalen MSE des klassischen Verfahrens, $\text{tr} \Sigma_{t|t}^{\text{KK}}$, erreichen können, sondern die Robustifizierung gerade so wählen, daß im idealen Modell der $\text{MSE}(\beta^{\text{rLS}}) = (1 + \delta) \text{tr} \Sigma_{t|t}^{\text{KK}}$ beträgt; dementsprechend sind auch die Kalman-Gains M_t zu modifizieren auf $M_t^{(\delta)}$.

An dieser Stelle ist es allerdings nicht gelungen, die Behauptung analytisch nachzuweisen, die jeweiligen Dichten von Prognose- und Filterfehler lägen in entsprechenden r -Umgebungen von Normalverteilungen.

Stattdessen wollen wir nun numerisch für $p = q = 1$ überprüfen, ob die Behauptung zutrifft.

8.4.4 numerische Auswertung

untersuchte Modelle

Im Rahmen einer kleinen Studie haben wir die Frage, ob die entsprechenden Dichten der Prognose-/Filterfehler innerhalb von r -Umgebungen von Normalverteilungen liegen, anhand der in Tabelle 8.1 dargestellten Modelle untersucht.

Tabelle 8.1: **untersuchte Modelle**

Beispiel no	F	Z	Q	V	a_0	r_1
8.1	1	1	1	1	0	5%
8.1	1	1	1	1	0	10%
8.1	1	1	1	1	0	20%
8.2	1	1	1/100	100	0	10%
8.3	1	1	100	1/100	0	10%
8.4	0.5	1	1	1	0	10%
8.5	0.5	-1	1	1	0	10%

Radien für $t = 1$

In einer ersten Studie betrachten wir den Fall $\Delta\beta_0 \sim \mathcal{N}_1(0, \Sigma)$, $\Sigma = \lim_t \Sigma_{t|t-1}^{[\delta]}$. Dazu fügen wir das Subskript $a = c$ für Verwendung der δ -Kompensation in Σ ein und $a = 0$ für entsprechend unkompensiertes Σ . Wir berechnen auf einem Gitter von r -Werten und für $i = 1, 2$ die Dichten

$$p_{i|1;r}^{\text{rLS}}(u) = p^{\Delta\beta_{i|1;r}^{\text{rLS}}}(u), \quad p_{i|1;a}^{\mathcal{N}}(u) := p^{\Delta\beta_{i|1;a}^{\text{KK}}}(u)$$

der entsprechend auf einen Radius r kalibrierten rLS-Filter/Prognosefehler sowie der zum Vergleich antretenden Normalverteilungen, indem wir die gemeinsame Dichte von $(\Delta\beta_1, \Delta y_1)$ im idealen Modell auf einem äquidistanten Gitter (β_i, y_j) der Dimension 401×401 auswerten. Anschließend ermitteln wir den "Abstand"

$$r_0(i|1; a; r) := r_0(p_{i|1;r}^{\text{rLS}}, p_{i|1;a}^{\mathcal{N}})$$

von $p^{\text{rLS}}(u)$ und $p^{\mathcal{N}}(u)$,

$$r_0(q, p) := \inf\{0 \leq r \leq 1 \mid q(u) \geq (1-r)p[\lambda]\}. \quad (8.62)$$

In Abbildung 8.3 stellen wir den Verlauf der Abbildung $r \mapsto r_0(r)$ im Steady-State Modell aus Beispiel 1.4.1 sowie in den Beispielen 8.2 und 8.3 aus Tabelle 8.1 dar. In Beispiel 8.3 halten die Filterfehler die Bedingung $r_0 < r$ nicht ein, in den anderen Beispielen schon. Wie gewünscht erhalten wir aber zumindest für den Prognosefehler $\Delta\beta_{2|1}^{\text{rLS}}$ in allen untersuchten Modellen $r_0(r) < r$. Ein ähnliches Verhalten haben wir bei allen untersuchten Modellen feststellen können; bestenfalls für sehr kleine r ($r < 0.015$) gab es Fälle, daß r_0 bei den Prognosefehlern geringfügig größer war, was jedoch wohl auf numerische Ungenauigkeiten zurückzuführen ist.

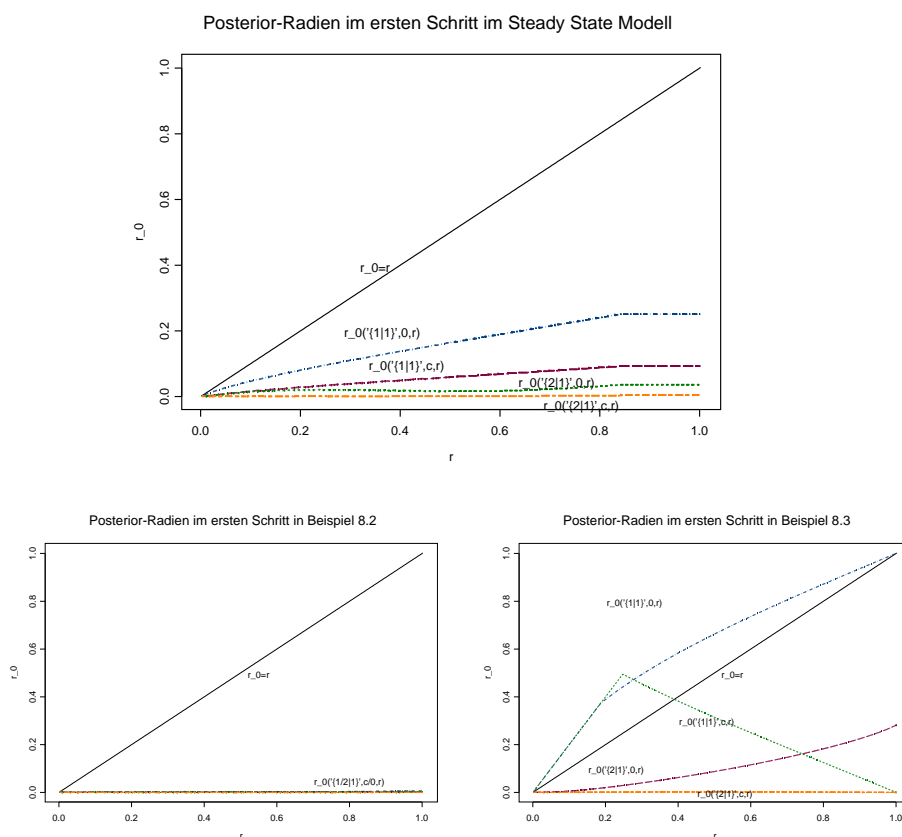


Abbildung 8.3: $r_0(r)$ im Steady-State Modell aus Beispiel 1.4.1 sowie in Beispielen 8.2, 8.3

Radien für $t \geq 1$

In einer zweiten Studie ermitteln wir für ein fixiertes r_1 rekursiv die Dichten von $\Delta\beta_{t+i|t;a;r_1}^{\text{rLS}}$ für $i = 0, 1$ $t = 1, \dots, 50$ ⁸ und kontrollieren jeweils, ob $r_{0,t}(r_1) < r_1$. Daneben lassen wir jeweils den exakten SO-optimalen Filter, der sich mit den exakten bedingten Erwartungswerten bestimmt, mitlaufen und ermitteln auch dessen “Abstand” $r_0(r_1)$ zur Normalverteilung. In Abbildung 8.4

⁸Nach 50 Beobachtungen hat sich in allen untersuchten Modellen $\Sigma_{t|t-1}$ und ebenso b_t , M_t stabilisiert.

stellen wir den Verlauf von $r_0(r)$ im Steady-State Modell aus Beispiel 1.4.1 bei Kalibrierung auf $r_1 = 10\%$, $r_1 = 20\%$ resp. $r_1 = 5\%$, sowie im Modell aus Beispiel 8.2 mit $r_1 = 10\%$ dar.

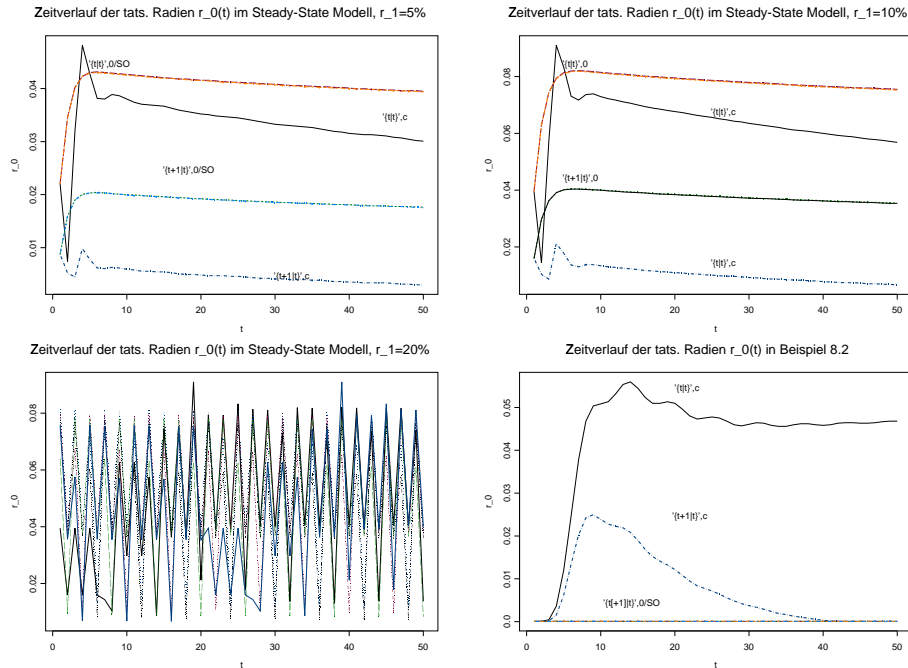


Abbildung 8.4: $r_0(r)$ im Zeitverlauf im Steady-State Modell aus Beispiel 1.4.1 bei Kalibrierung auf $r_1 = 10\%$, $r_1 = 20\%$ resp. $r_1 = 5\%$, sowie im Modell aus Beispiel 8.2 mit $r_1 = 10\%$

In $t = 50$ vergleichen wir für $r = 10\%$ die Funktionen in y , $\hat{f}_a^{rLS}(y)$ und $\hat{f}^{SO}(y)$, $E[\Delta\beta_50|\Delta y_50 = y]$ sowie die besten linearen Verfahren $f_a^{KK}(y)$ mit und ohne δ -Kompensation und stellen sie alle in Abbildung 8.5 dar. Die Nichtlinearität des bedingten Erwartungswertes kommt erst weit jenseits der Clippingsschwelle zum Tragen; der unkompenzierte rLS-Filter $\hat{f}_0^{rLS}(y)$ läßt sich nicht von $\hat{f}^{SO}(y)$ unterscheiden. Dieser Effekt war auch in den übrigen Beispielen zu beobachten.

Anschließend gehen wir im fixierten Zeitpunkt $t = 50$ vor wie in der ersten Studie, das heißt wir ermitteln jeweils den “Abstand” $r_{50}(r)$ zur Normalverteilung auf einem r -Gitter.

In Abbildung 8.6 stellen wir für $r_1 = 10\%$ den Verlauf von $r_{50}(r)$ im Steady-State Modell aus Beispiel 1.4.1 dar. Hier erfüllen die kompensierten Filter- und Prognosefehler für kleine r nicht die Bedingung $r_{50} < r$, die unkompenzierten allerdings immer. Für die übrigen Beispiele ergaben sich entsprechende Darstellungen.

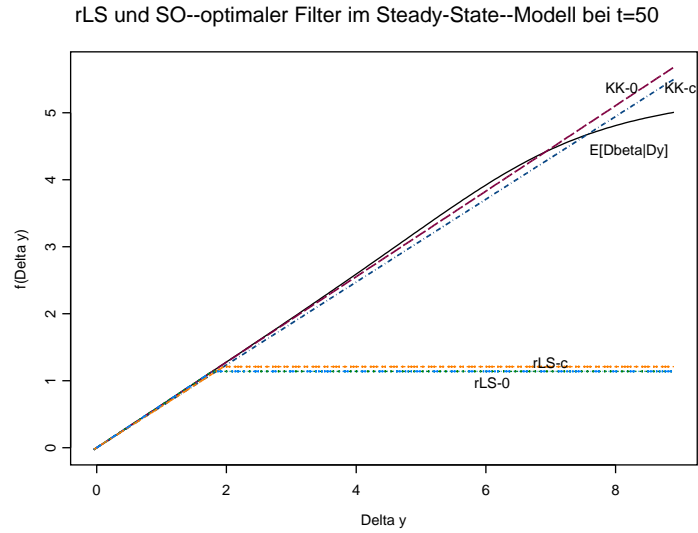


Abbildung 8.5: $\hat{f}^{rLS}(y)$ und $\hat{f}^{SO}(y)$ in $t = 50$ im Steady-State Modell aus Beispiel 1.4.1

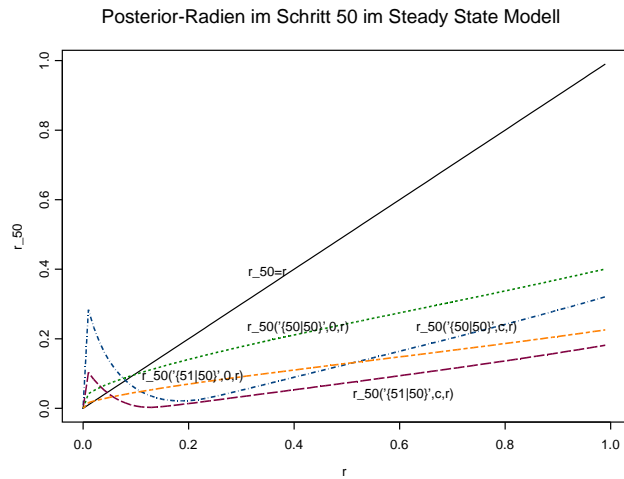


Abbildung 8.6: $r_{50}(r)$ im Steady-State Modell aus Beispiel 1.4.1

numerische Aspekte

Die numerischen Fehler, die wir bei diesen Auswertungen eingehen, genügen dabei folgenden Abschätzungen

Abschneidefehler: Weil wir anstelle des vollen \mathbb{R} die Integrale jeweils nur auf $[-a\sigma; a\sigma]$ auswerten, brauchen wir eine Abschätzung für die Integrale auf $\mathbb{R} \setminus [-a\sigma; a\sigma]$. Für die Integranden $p_1 = \varphi(x/\sigma)/\sigma$, $p_2 = (x-c)\varphi(x/\sigma)/\sigma$ und $p_3 = x^2\varphi(x/\sigma)/\sigma$ mit $\varphi(x)$ der Dichte von $\mathcal{N}(0,1)$ ergeben sich mit

der Mill's Ratio für $a > 0$

$$\int_a^\infty \varphi(x) dx \leq \varphi(a)/a$$

Fehlerabschätzungen $A_1 = 2\varphi(a)/a$, $A_2 = |c|\varphi(a)/a$, $A_3 = (a^2+1)\varphi(a)/a$

Integrationsfehler: Wir verwenden das Simpson-Verfahren, machen dabei also — vgl. etwa Stoer (1989), p. 118 — einen Fehler Kh^4 , $h = 2a/N$ die Diskretisierungstiefe, $K \leq \frac{a}{90} \sup_x |\frac{d^4}{dx^4}p(x)|$, so daß mit

$$\begin{aligned} \frac{d^3}{dx^3}\varphi(x) &= (-x^3 + 3x)\varphi(x), \\ \frac{d^4}{dx^4}\varphi(x) &= (x^4 - 6x^2 + 3)\varphi(x), \\ \frac{d^5}{dx^5}\varphi(x) &= (-x^5 + 10x^3 - 15x)\varphi(x), \\ \frac{d^6}{dx^6}\varphi(x) &= (x^6 - 15x^4 + 45x^2 - 15)\varphi(x) \end{aligned}$$

sich die entsprechenden K_i abschätzen lassen durch

$$\begin{aligned} K_1 &\leq 3\varphi(0)a/(90\sigma^4) \doteq 0.0133a/\sigma^4 \\ K_2 &\leq K_1 + \frac{d^5}{dx^5}\varphi(0.6167)a/(90\sigma^4) \doteq 0.0391a/\sigma^4 \\ K_3 &\leq 15\varphi(0)a/(90\sigma^4) \doteq 0.0317a/\sigma^4 \end{aligned}$$

“optimales” a : Mit diesen beiden Fehlern A_i und $K_i h^4$ können wir nun versuchen, zu gegebenem N ein optimales a zu finden, so daß beide Fehler gleichmäßig klein sind. Für $N = 400$ ergibt sich, daß für $0.2 \leq \sigma$ der Wert $a = 6.5$ sowohl A_i als auch K_i kleiner als 10^{-6} werden läßt.

8.5 Parametrisierung des Grades der Robustheit

Die Untersuchungen der letzten Abschnitte haben deutlich gemacht, daß für den Grad der Robustheit, den wir für ein Filterverfahren erreichen wollen, zumindest folgende drei Maßzahlen in Frage kommen:

- die Clippingschwelle b respektive $\tilde{\rho} = 1/b$
- der Kontaminationsradius r
- der Effizienzverlust δ

Offenbar kann man mit einer der drei Maßzahlen die zwei anderen eindeutig bestimmen, und es herrscht folgende Monotonie $b \downarrow \Rightarrow r, \delta \uparrow$, $r \uparrow \Leftrightarrow \delta \uparrow$. Wie die drei Größen exakt zusammenhängen, ist in Abbildung 8.7 für das Steady-State Modell aus Beispiel 1.4.1 im Grenzfall, also

$$S = \lim_t \Sigma_t^{\text{KK}} = (\sqrt{5} + 1)/2 \doteq 1.618, \quad \text{resp. } M = (\sqrt{5} - 1)/2 \doteq 0.618$$

abgetragen.

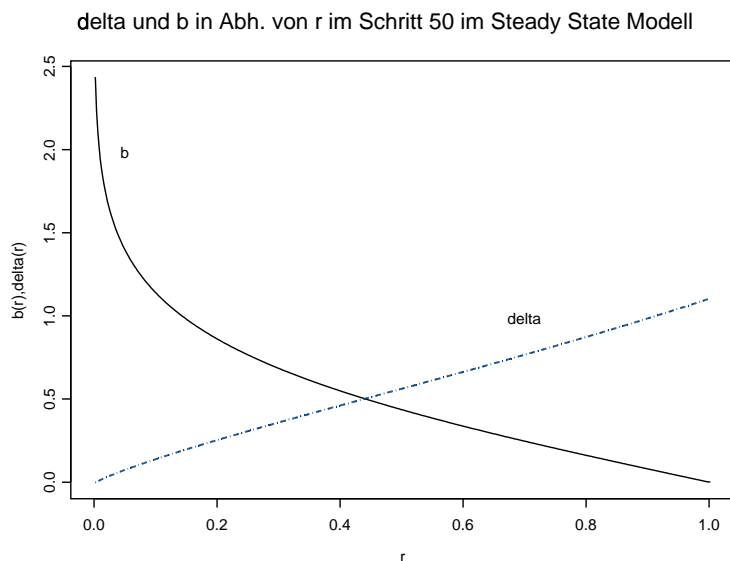


Abbildung 8.7: Zusammenhang r_{so} , δ , $b^{\text{so}} = 1/\tilde{\rho}^{\text{so}}$

8.6 “Lemma 5”–Ansatz für AO’s im reduzierten Modell

Sehr viel schwieriger als im SO–Modell stellt sich die Situation im AO–Modell dar, da hier nicht die Beobachtungen selbst kontaminiert werden, sondern man eine kontaminierte Faltung beobachtet. Dies führt auf Probleme, die im allgemeinen nicht mehr geschlossen lösbar sind.

8.6.1 Problemstellung

Es geht darum, eine Funktion $f(Y)$ zu finden, die $E_{\text{id}}[|X - f(Y)|^2]$ minimiert unter der Nebenbedingung, daß $|E_{\text{cont}}[X - f(Y)]| \leq b$. Dabei legen wir die Kontamination fest als eine AO–Konvexkontamination mit Radius r , das heißt, wir setzen mit einem kontaminierenden Maß \tilde{P}^ε

$$E_{\text{real}}[f(Y)] := (1 - r) E_{\text{id}}[f(Y)] + r E_{P^X * \tilde{P}^\varepsilon}[f(Y)].$$

Da wir wieder keine Kontamination in X betrachten, ist (3) äquivalent mit

$$|E_{\text{id}}[X] - E_{\text{real}}[f]| \leq b \quad (8.63)$$

8.6.2 Lösung

Zunächst können wir wegen der verlangten Meßbarkeit in Y von f und der Existenz des zweiten Momentes von X das Problem (2) in Abschnitt 8.1.2 zurückführen auf

$$|E_{\text{id}}[|E_{\text{id}}[X|Y] - f(Y)|^2] \stackrel{!}{=} \min_f \quad \text{NB (8.63)}. \quad (8.64)$$

Da sich die Kontamination nun im ε -Raum abspielt, werten wir den Erwartungswert in (63) in zwei Schritten aus und erhalten

$$\mathbf{E}_{\text{real}}[f(Y)] = \mathbf{E}_{\hat{P}_\varepsilon} \left[\mathbf{E}_{P^X} [f(X + \varepsilon) | \varepsilon] \right]. \quad (8.65)$$

Den bedingten Erwartungswert $\mathbf{E}_{P^X} [f(X + \varepsilon) | \varepsilon]$ bezeichnen wir mit f_ε . Wie in Abschnitt 8.2 wollen wir nun eine Äquivalenz vom Typ

$$\sup |f_\varepsilon - \mathbf{E}_{\text{id}}[X]| \leq \tilde{b} \iff |\mathbf{E}_{\text{real}}[f(Y)] - \mathbf{E}_{\text{id}}[X]| \leq b \quad (8.66)$$

herleiten, wobei wie dort \tilde{b} nur von b , dem Kontaminationsradius r und dem Idealbias $\mathbf{E}_{\text{id}}[X - f(Y)]$ abhängen möge.

Es ist klar, daß für $\sup_{P^Y | \varepsilon = \epsilon} f_\varepsilon(\epsilon) = \tilde{b} + \mathbf{E}_{\text{id}}[X]$, $\inf_{P^Y | \varepsilon = \epsilon} f_\varepsilon(\epsilon) = -\tilde{b} + \mathbf{E}_{\text{id}}[X]$ gilt $|\mathbf{E}_{\text{real}}[f(Y)] - \mathbf{E}_{\text{id}}[X]| \leq b$.

Im folgenden beschränken wir uns, wie im SO-Fall, auf im idealen Modell unverzerrte Verfahren, nehmen also (4) an, so daß

$$|\mathbf{E}_{\text{id}}[X] - \mathbf{E}_{\text{real}}[f(Y)]| = r |\mathbf{E}_{\hat{P}_\varepsilon}[f_\varepsilon] - \mathbf{E}_{\text{id}}[X]|.$$

Wie beim SO-Problem sehen wir, daß dann gilt

$$\sup |f_\varepsilon - \mathbf{E}_{\text{id}}[X]| = b/r \iff |\mathbf{E}_{\text{real}}[f(Y)] - \mathbf{E}_{\text{id}}[X]| \leq b, \quad (8.67)$$

also $\tilde{b} = b/r$.

8.6.3 Existenz und Eindeutigkeit einer Lösung

Wir wollen zunächst die Frage beantworten, ob überhaupt eine Lösung zu (64) existiert und ob diese eindeutig ist. Dazu formulieren wir (64) um zu

$$\mathbf{E}_{\text{id}}[|d(Y)|^2] \stackrel{!}{=} \min_{d \in L_2(P^Y)} \text{NB} \quad (8.68)$$

$$\sup |\mathbf{E}_{\text{id}}[d + \mathbf{E}_{\text{id}}[X|Y]| \varepsilon] - \mathbf{E}_{\text{id}}[X]| \leq \tilde{b} \quad (8.69)$$

und das optimale \hat{f}^{AO} zu (64) ist $d + \mathbf{E}_{\text{id}}[X|Y]$.

Im folgenden wollen wir wie bisher $\mathbf{E}_{\text{id}}[X|Y] - \mathbf{E}_{\text{id}}[X]$ mit $D(Y)$ oder kurz D abkürzen. Offenbar stellt (68) ein Minimum-Norm-Problem auf einer Teilmenge A des $L_2(P^Y)$ dar. Dabei ist

$$A := \{d \in L_2(P^Y) : \sup |\mathbf{E}_{\text{id}}[d + D | \varepsilon]| \leq \tilde{b}\}.$$

A ist konvex, denn mit $d_\lambda := (1 - \lambda)d_1 + \lambda d_2$ und $d_1, d_2 \in A$, $\lambda \in [0, 1]$ gilt:

$$\begin{aligned} & \sup |\mathbf{E}_{\text{id}}[d_\lambda + D | \varepsilon]| = \\ &= \sup |\mathbf{E}_{\text{id}}[(1 - \lambda)(d_1 + D) + \lambda(d_2 + D) | \varepsilon]| \leq \\ &\leq (1 - \lambda) \sup |\mathbf{E}_{\text{id}}[d_1 + D | \varepsilon]| + \lambda \sup |\mathbf{E}_{\text{id}}[d_2 + D | \varepsilon]| \leq b \end{aligned}$$

A ist beschränkt (klar) und schwach abgeschlossen in $L_2(P^Y)$:

Ohne Einschränkung sei $D = 0$. Sei $d_n \in A$, $\|d_n - d\|_{L_2(P^Y)} \rightarrow 0$. Mit Jensen für bedingte Erwartungswerte folgt

$$\mathbf{E} |\mathbf{E}[d_n | \varepsilon] - \mathbf{E}[d | \varepsilon]|^2 \leq \|d_n - d\|_{L_2(P^Y)}^2,$$

also die Konvergenz der bedingten Erwartungswerte im $L_2(P^Y)$. Hiermit ist Proposition C.3.1 für $x_n = \mathbf{E}[d_n | \varepsilon]$, $x = \mathbf{E}[d | \varepsilon]$ anwendbar, und es folgt die schwache (und starke) Abgeschlossenheit von A . Daher existiert genau ein Element kleinster Norm in A .

8.6.4 Lösungsansatz mit Lagrangetechniken

Da damit (64) offenbar ein konvexes Optimierungsproblem ist, liegt es nahe, zur Lösung wie im Beweis zu Theorem 8.3.1 ein Optimierungstheorem aus Rieder (1994) zu verwenden. Wir geben dazu die Identifikationen für eine Anwendung von Rieders Theorem B.2.1 an:

$$X \rightsquigarrow L_2(P^Y), \quad Z \rightsquigarrow (L_2(P^\varepsilon))^2, \quad C \rightsquigarrow \{h \in L_2(P^\varepsilon) : h \geq 0 [P^\varepsilon]\}, \quad A \rightsquigarrow L_2(P^Y), \quad G \rightsquigarrow (\text{Eid}[\cdot + D | \varepsilon]^\tau, -\text{Eid}[\cdot + D | \varepsilon]^\tau)^\tau, \quad f \rightsquigarrow \|\cdot\|_{L_2(P^Y)}^2, \quad z_0 \rightsquigarrow (b, b)^\tau$$

Allerdings ist hier $\overset{\circ}{C} = \emptyset$, so daß die Existenz stetiger Funktionale als Lagrange-Multiplikatoren nicht zur Verfügung steht.

8.6.5 Formulierung mit Dichten

Zur Existenz von Dichten müssen wir die Dominiertheit der gemeinsamen Verteilung von Y und ε im idealen Modell durch ein Produktmaß $\mu_1 \otimes \mu_2$ voraussetzen.

Dann aber stellt sich (64) mit Dichten so dar:

$$\int |f(y) - D(y)|^2 p^X(y - \varepsilon) p^\varepsilon(\varepsilon) \mu_1(d\varepsilon) \mu_2(dy) = \min_f \quad \text{NB: (8.71)} \quad (8.70)$$

$$|\int f(y) p^X(y - \varepsilon) \mu_2(dy)| \leq b \quad [\mu_1(d\varepsilon)] \quad (8.71)$$

8.6.6 Eigenschaften der Lösung

Da f in der Nebenbedingung mit p^X gefaltet wird, erbt es alle Glattheitseigenschaften von p^X — sofern Differentiation und Integration vertauschbar sind. Ist nämlich f beschränkt und p^X differenzierbar mit beschränkter Ableitung, so ist dies auch

$$f_\varepsilon(e) = \int f(y) p^X(y - e) \mu_2(dy). \quad (8.72)$$

Damit kann es — sofern $\text{E}[\text{E}[X|Y]|\varepsilon = \varepsilon]$ das Niveau b mit einer positiven Steigung überquert — auch kein Y -meßbares f geben, das die Nebenbedingung “maximal” erfüllt, also so, daß

$$f_\varepsilon(\varepsilon) = \text{E}[\text{E}[X|Y]|\varepsilon = \varepsilon] \min\left\{1, \frac{b}{|\text{E}[\text{E}[X|Y]|\varepsilon = \varepsilon]|}\right\},$$

denn ein solches f würde die hergeleitete Glattheit von $\text{E}[f|\varepsilon]$ verletzen.

Ebensowenig ist die klassische Lösung $\text{E}[X|Y]$ zulässig, weil i.a.

$$\sup_{P^\varepsilon} |\text{E}[\text{E}[X|Y]|\varepsilon]| = \infty.$$

Hingegen ist die SO-Lösung immer zulässig, da $\sup_{P^\varepsilon} |\text{E}[f|\varepsilon]| \leq \sup |f| = b$, aber im allgemeinen suboptimal.

8.6.7 numerische Lösung

Diskretisierungsstellen

Wir wollen die Lösung numerisch ermitteln, indem wir das Problem diskretisieren und es dann mit numerischen Verfahren bearbeiten. Als Gitterpunkte seien

dazu in "y"-Richtung $-\infty < t_0 < t_1 < \dots < t_i < \dots < t_n < t_{n+1} < \infty$ gewählt, die Inkremente sollen $d_i := t_i - t_{i-1}$, $i = 1, \dots, n+1$ heißen.

Auch die "ε"-Richtung sei diskretisiert, und zwar durch $-\infty < \varepsilon_1 < \varepsilon_2 < \dots < \varepsilon_j < \dots < \varepsilon_{m-1} < \varepsilon_m < \infty$.

Als mögliche Diskretisierungsfolgen werden im folgenden äquidistante t_i 's und ε_j 's gewählt, oder sie werden gemäß Gaußquantilen gesetzt.

erste Diskretisierung

Als ersten Diskretisierungsversuch wollen wir einen Polygonzug ansetzen, wobei wir als Basis lineare Teilstücke $f|_{[t_{i-1}; t_i]}(x) = \alpha_i x + \beta_i$ für $i = 1, \dots, n+1$ verwenden wollen. Außerdem soll an den Rändern der Diskretisierung gelten:

$$f(t) = -f(t') = b \quad \text{für } t' \leq t_0 \text{ und } t \geq t_{n+1}. \quad (8.73)$$

Wegen der Stetigkeitsforderung und den Randbedingungen bleiben aber nur n Freiheitsgrade übrig, nämlich $\alpha_1, \dots, \alpha_n$; die sonstigen Parameter lassen sich auf diese zurückführen, wie folgendes Lemma zeigt:

Lemma 8.6.1 *Seien $-\infty < t_0 < t_1 < \dots < t_n < t_{n+1} < \infty$ Diskretisierungspunkte, $d_i := t_i - t_{i-1}$, $i = 1, \dots, n+1$. Sei $f: \mathbb{R} \rightarrow \mathbb{R}$ ein stetiger Polygonzug mit $f(t_{n+1}) = -f(t_0) = b$ und $f|_{[t_{i-1}; t_i]}(x) = \alpha_i x + \beta_i$ für $i = 1, \dots, n+1$, so gilt*

$$\beta_i = -\alpha_i t_{i-1} + \sum_{j=1}^{i-1} \alpha_j d_j - b \quad i = 1, \dots, n+1 \quad (8.74)$$

$$\alpha_{n+1} = 2b/d_{n+1} - \sum_{j=1}^n \alpha_j d_j / d_{n+1} \quad (8.75)$$

$$\beta_{n+1} = \sum_{j=1}^n \alpha_j d_j / d_{n+1} t_{n+1} - b \frac{t_n + t_{n+1}}{d_{n+1}} \quad (8.76)$$

BEWEIS: $i = 1$: Es gilt nach Voraussetzung $-b = f(t_0) = \alpha_1 t_0 + \beta_1$, also $\beta_1 = -b - \alpha_1 t_0$; $i \rightarrow i+1$: Wegen der Stetigkeit von f gilt

$$\begin{aligned} f(t_i) &= \alpha_{i+1} t_i + \beta_{i+1} = \alpha_i t_i + \beta_i = \alpha_i t_i - \alpha_i t_{i-1} + \sum_{j=1}^{i-1} \alpha_j d_j - b = \\ &= \sum_{j=1}^i \alpha_j d_j - b, \end{aligned}$$

also $\beta_{i+1} = -\alpha_{i+1} t_i + \sum_{j=1}^i \alpha_j d_j - b$. Außerdem gilt

$$b = f(t_{n+1}) = \alpha_{n+1} t_{n+1} + \beta_{n+1} = \alpha_{n+1} d_{n+1} + \sum_{j=1}^n \alpha_j d_j - b$$

und damit $\alpha_{n+1} = 2b/d_{n+1} - \sum_{j=1}^n \alpha_j d_j / d_{n+1}$. Schließlich gilt

$$\beta_{n+1} = -\alpha_{n+1} t_n + \sum_{j=1}^n \alpha_j d_j - b = \sum_{j=1}^n \alpha_j d_j / d_{n+1} t_{n+1} - b \frac{t_n + t_{n+1}}{d_{n+1}}$$

////

Wir wollen (74) für $i = 1, \dots, n$ abkürzen zu

$$\beta_i = G_i \alpha - b \quad \text{mit} \quad G_i := (d_1, \dots, d_{i-1}, -t_{i-1}, 0, \dots, 0). \quad (8.77)$$

Dazu sei weiter $G := (G_1^\tau, \dots, G_n^\tau)^\tau$ und

$$G_{n+1} := -(d_1, \dots, d_n)/d_{n+1}, \text{ so da\ss } \alpha_{n+1} = G_{n+1}\alpha + 2b/d_{n+1} \quad (8.78)$$

Schlie\sslich wird

$$\beta_{n+1} = -\alpha_{n+1}t_{n+1} + b \quad (8.79)$$

eine weitere Diskretisierung

Als Basis wollen wir nun sogenannte *Tentfunktionen* verwenden:

$$f_i(t) := \mathbf{I}_{(t_{i-1}; t_i]} \frac{t - t_{i-1}}{t_i - t_{i-1}} + \mathbf{I}_{(t_i; t_{i+1}]} \frac{t - t_{i+1}}{t_i - t_{i+1}} \quad (8.80)$$

An den R\u00e4ndern soll wie in (73) schlie\sslich der Funktionswert konstant $\mp b$ sein, daher kommen noch die fixen Funktionen

$$f_0 = (-b) \mathbf{I}_{(-\infty; t_0]} - b \mathbf{I}_{(t_0; t_1]} \frac{t - t_1}{t_0 - t_1} \quad (8.81)$$

$$f_{n+1} = b \mathbf{I}_{(t_n; t_{n+1}]} \frac{t - t_n}{t_{n+1} - t_n} + b \mathbf{I}_{(t_{n+1}; \infty)} \quad (8.82)$$

hinzu.

Im folgenden wollen wir $\mathbf{I}_{(t_{i-1}; t_i]}$, $i = 1, \dots, n+1$ durch \mathbf{I}_i , $\mathbf{I}_{(-\infty; t_0]}$ durch \mathbf{I}_0 und $\mathbf{I}_{(t_{n+1}; \infty)}$ durch \mathbf{I}_{n+2} abk\u00fcrzen.

Beispiel: $P^X = \mathcal{N}(0, S)$, $P^\varepsilon = \mathcal{N}(0, V)$

Im Spezialfall, da\ss sowohl X als auch ε normalverteilt sind, kann man noch mehr ausrechnen; hierzu dient folgendes Lemma, dessen Beweis klar ist.

Lemma 8.6.2 *Mit φ der Lebesguedichte von $\mathcal{N}(0, 1)$ und Φ der entsprechenden Verteilung gilt:*

$$\int_a^b (u + vx)^2 \varphi(x) dx = (u^2 + v^2) \Phi(x) - (v^2 x + 2uv) \varphi(x) \Big|_{x=a}^b \quad (8.83)$$

$$\int_a^b \frac{((\gamma - \alpha)x - \beta)^2}{v} \varphi\left(\frac{x}{v}\right) dx = [\beta^2 + v^2(\gamma - \alpha)^2] \Phi\left(\frac{x}{v}\right) - v[(\gamma - \alpha)^2 x - 2\beta(\gamma - \alpha)] \varphi\left(\frac{x}{v}\right) \Big|_{x=a}^b \quad (8.84)$$

$$\int_a^b \frac{1}{s} [\alpha x + \beta] \varphi\left(\frac{x - \varepsilon}{s}\right) dx = [\varepsilon \alpha + \beta] \Phi\left(\frac{x - \varepsilon}{s}\right) - \alpha s \varphi\left(\frac{x - \varepsilon}{s}\right) \Big|_{x=a}^b, \quad (8.85)$$

Im folgenden verwenden wir die Abk\u00fcrzungen:

$$\begin{aligned} T^2 &:= S + V, & M &= S/T^2 \\ \Phi_i &:= \Phi(t_i/T), & \varphi_i &:= \varphi(t_i/T) \\ d\Phi_i &:= \Phi_i - \Phi_{i-1}, & d\varphi_i &:= \varphi_i - \varphi_{i-1}, & dy\varphi_i &:= t_i \varphi_i - t_{i-1} \varphi_{i-1} \\ \Phi_{i,j} &:= \Phi\left(\frac{t_i - \varepsilon_j}{\sqrt{S}}\right), & \varphi_{i,j} &:= \varphi\left(\frac{t_i - \varepsilon_j}{\sqrt{S}}\right) \\ d\Phi_{i,j} &:= \Phi_{i,j} - \Phi_{i-1,j}, & d\varphi_{i,j} &:= \varphi_{i,j} - \varphi_{i-1,j} \end{aligned}$$

Ziel wird es nun sein, das Problem (70)/(71) auf die Gestalt eines quadratischen Optimierungsproblems im \mathbb{R}^n mit linearen Ungleichungs-Nebenbedingungen, genauer sogenannten *box-constraints*, zu bringen, das heißt

$$\alpha^\tau A \alpha + B^\tau \alpha + C = \min_{\alpha \in \mathbb{R}^n} ! \quad \text{NB} \quad D \alpha \leq E, \quad -D \alpha \leq F \quad (8.86)$$

mit $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^n$, $C \in \mathbb{R}$, $D \in \mathbb{R}^{m \times n}$, $E \in \mathbb{R}^m$, $F \in \mathbb{R}^m$. Desweiteren liefert der SO-optimale Filter einen zulässigen Startwert α_0 für α .

Berechnung auf Form (8.86) für die erste Diskretisierung

Die Zielfunktion läßt sich schreiben als

$$\int \frac{[My - f(y)]^2}{T} \varphi\left(\frac{y}{T}\right) dy = I_0 + \sum_{i=1}^n I_i + I_{n+1} + I_{n+2}$$

mit

$$\begin{aligned} I_0 &= \int_{-\infty}^{t_0} \frac{[My+b]^2}{T} \varphi\left(\frac{y}{T}\right) dy \\ I_i &= \int_{t_{i-1}}^{t_i} \frac{[(M-\alpha_i)y - \beta_i(\alpha)]^2}{T} \varphi\left(\frac{y}{T}\right) dy \quad \text{für } i = 1, \dots, n \\ I_{n+1} &= \int_{t_n}^{t_{n+1}} \frac{[(M-\alpha_{n+1}(\alpha))y - \beta_i(\alpha)]^2}{T} \varphi\left(\frac{y}{T}\right) dy \\ I_{n+2} &= \int_{t_{n+1}}^{\infty} \frac{[My-b]^2}{T} \varphi\left(\frac{y}{T}\right) dy \end{aligned}$$

Nach Lemma 8.6.2 ergibt sich

$$\begin{aligned} I_0 &= (b^2 + T^2 M^2) \Phi_0 - T(M^2 t_0 + bM) \varphi_0 =: C_0 \\ I_{n+2} &= (b^2 + T^2 M^2)(1 - \Phi_{n+1}) + T(M^2 t_{n+1} - bM) \varphi_{n+1} =: C_{n+1} \end{aligned}$$

und für $i = 1, \dots, n+1$

$$I_i = (\beta_i^2 + T^2(M - \alpha_i)^2) d\Phi_i - T(M - \alpha_i)^2 d\varphi_i + 2T\beta_i(M - \alpha_i) d\rho_i$$

und damit für $i = 1, \dots, n$ nach (77)

$$\begin{aligned} I_i &= ((G_i \alpha - b)^2 + T^2(M - \alpha_i)^2) d\Phi_i - T(M - \alpha_i)^2 d\varphi_i + \\ &\quad + 2T(G_i \alpha - b)(M - \alpha_i) d\rho_i = \\ &= [(G_i \alpha)^2 + T^2 \alpha_i^2 - 2bG_i \alpha - 2T^2 M \alpha_i + b^2 + T^2 M^2] d\Phi_i - \\ &\quad - T(\alpha_i^2 - 2M \alpha_i + M^2) d\varphi_i + \\ &\quad + 2T(-\alpha_i G_i \alpha + b \alpha_i + M G_i \alpha - Mb) d\rho_i \end{aligned}$$

Bezeichnen wir $\text{diag}[(d\Phi_i)_{i=1, \dots, n}]$ mit $D\Phi$ und analog für die anderen d -Größen, so liefert die Summation über i

$$\begin{aligned} \sum_{i=1}^n I_i &= \alpha^\tau \left[G^\tau D\Phi G + T^2 D\Phi - T D Y \varphi - T D \rho G - T G^\tau D \rho \right] \alpha + \\ &\quad \left[-2MT^2 d\Phi^\tau - 2bd\Phi^\tau G + 2TM d\varphi^\tau + 2TM d\rho^\tau G + 2Tbd\rho^\tau \right] \alpha + \\ &\quad \mathbf{1}^\tau \left[(b^2 + T^2 M^2) d\Phi - TM^2 d\varphi - TMbd\rho \right] = \\ &=: \alpha^\tau A_1 \alpha + \alpha^\tau B_1 + C_1 \end{aligned}$$

Für den Randterm gilt nach (79) und (78):

$$\begin{aligned} I_{n+1} &= ([b - (G_{n+1}\alpha + 2b/d_{n+1})t_{n+1}]^2 + T^2(M - (G_{n+1}\alpha + 2b/d_{n+1}))^2)d\Phi_i - \\ &\quad - T(M - (G_{n+1}\alpha + 2b/d_{n+1}))^2 dy\varphi_i + \\ &\quad + 2T[G_{n+1}\alpha + 2b/d_{n+1}](M - (G_{n+1}\alpha + 2b/d_{n+1}))d\varphi_i = \\ &= L_1(G_{n+1}\alpha)^2 + [4\frac{b}{d_{n+1}}L_1 + L_2]G_{n+1}\alpha + 4\frac{b^2}{d_{n+1}^2}L_1 + 2\frac{b}{d_{n+1}}L_2 + L_3 \end{aligned}$$

mit

$$\begin{aligned} L_1 &:= (t_{n+1}^2 + T^2)d\Phi_{n+1} - Tdy\varphi_{n+1} + 2Tt_{n+1}d\varphi_{n+1} \\ L_2 &:= -2[(bt_{n+1} + M^2T^2)d\Phi_{n+1} - TMdy\varphi_{n+1} + (MTt_{n+1} + Tb)d\varphi_{n+1}] \\ L_3 &:= (b^2 + M^2T^2)d\Phi_{n+1} - TM^2dy\varphi_{n+1} + 2MTbd\varphi_{n+1} \end{aligned}$$

Demzufolge ist $A = A_1 + L_1G_{n+1}G_{n+1}^\tau$, $B = B_1 + [4\frac{b}{d_{n+1}}L_1 + L_2]G_{n+1}$ und $C = C_1 + 4\frac{b^2}{d_{n+1}^2}L_1 + 2\frac{b}{d_{n+1}}L_2 + L_3 + C_0 + C_{n+1}$. Mit der Nebenbedingung gehen wir analog vor:

$$\int f(y)\frac{1}{\sqrt{S}}\varphi\left(\frac{y-\varepsilon_j}{\sqrt{S}}\right)dy = J_{0,j} + \sum_{i=1}^n J_{i,j} + J_{n+1,j} + J_{n+2,j}$$

mit

$$\begin{aligned} J_{0,j} &= \int_{-\infty}^{t_0} (-b)\frac{1}{\sqrt{S}}\varphi\left(\frac{y-\varepsilon_j}{\sqrt{S}}\right)dy \\ J_{i,j} &= \int_{t_{i-1}}^{t_i} [\alpha_i y - \beta_i(\alpha)]\frac{1}{\sqrt{S}}\varphi\left(\frac{y-\varepsilon_j}{\sqrt{S}}\right)dy \quad \text{für } i = 1, \dots, n \\ J_{n+1,j} &= \int_{t_n}^{t_{n+1}} [\alpha_{n+1}(\alpha)y - \beta_i(\alpha)]\frac{1}{\sqrt{S}}\varphi\left(\frac{y-\varepsilon_j}{\sqrt{S}}\right)dy \\ J_{n+2,j} &= \int_{t_{n+1}}^{\infty} b\frac{1}{\sqrt{S}}\varphi\left(\frac{y-\varepsilon_j}{\sqrt{S}}\right)dy \end{aligned}$$

Analog ergibt sich nach Lemma 8.6.2

$$J_{0,j} = -b\Phi_{0,j}, \quad J_{n+2,j} = b(1 - \Phi_{n+1,j})$$

und für $i = 1, \dots, n+1$

$$J_{i,j} = [\varepsilon_j\alpha_i + \beta_i]d\Phi_{i,j} - \alpha_i\sqrt{S}d\varphi_{i,j}$$

und damit für $i = 1, \dots, n$ nach (77)

$$J_{i,j} = [\varepsilon_j\alpha_i + (G_i\alpha - b)]d\Phi_{i,j} - \alpha_i\sqrt{S}d\varphi_{i,j}$$

Für den Randterm gilt nach (79) und (78):

$$\begin{aligned} J_{n+1,j} &= [\varepsilon_j(G_{n+1}\alpha + 2b/d_{n+1}) + b - (G_{n+1}\alpha + 2b/d_{n+1})t_{n+1}]d\Phi_{n+1,j} - \\ &\quad - \sqrt{S}(G_{n+1}\alpha + 2b/d_{n+1})d\varphi_{n+1,j} \end{aligned}$$

Damit bekommen wir, falls wir mit $DD\Phi$ die Matrix mit Einträgen $DD\Phi^\tau = (d\Phi_{i,j})_{i,j}$ und analog $DD\varphi$ einführen, sowie $\mathcal{E} := \text{diag}[\varepsilon_j]$

$$D = \mathcal{E}DD\Phi + DD\Phi G - \sqrt{S}DD\varphi + (\mathcal{E} - t_{n+1}\mathbb{I})d\Phi_{n+1}^\tau, G_{n+1} - \sqrt{S}d\varphi_{n+1}^\tau, G_{n+1}$$

Setzen wir $H = (H_j)$ für die von α unabhängigen Komponenten der $J_{i,j}$, so erhalten wir

$$\begin{aligned} H &= b \left[1 - \Phi_{n+1,j} - \Phi_{0,j} - (\Phi_{n,j} - \Phi_{0,j}) + 2(\varepsilon_j - t_{n+1})/d_{n+1}d\Phi_{n+1,j} + \right. \\ &\quad \left. + d\Phi_{n+1,j} - 2\sqrt{S}/d_{n+1}d\varphi_{n+1,j} \right] = \\ &= b \left[1 - 2\Phi_{n,j} + 2(\varepsilon_j - t_{n+1})/d_{n+1}d\Phi_{n+1,j} - 2\sqrt{S}/d_{n+1}d\varphi_{n+1,j} \right] \end{aligned}$$

und somit $|D\alpha + H| \leq b$, so daß $E = b - H$ und $F = b + H = 2b - E$. Der Startvektor α_0 ergibt sich als $\alpha_{0,i} := df_i^{\text{SO}}/d_i$ mit $df_i^{\text{SO}} = f^{\text{SO}}(t_i) - f^{\text{SO}}(t_{i-1})$ und f^{SO} der SO-optimalen Schätzung für X , also im eindimensionalen, normalverteilten Fall $f^{\text{SO}} = \min\{b, \max\{My, -b\}\}$.

Berechnung auf Form (8.86) für die zweite Diskretisierung

Dieses Mal berechnen wir zunächst die Zielfunktion als

$$\begin{aligned} &\int [My - f(y)]^2 \frac{1}{T} \varphi\left(\frac{y}{T}\right) dy = M^2 \int y^2 \frac{1}{T} \varphi\left(\frac{y}{T}\right) dy - \\ &\quad - 2M \int y f(y) \frac{1}{T} \varphi\left(\frac{y}{T}\right) dy + \int f(y)^2 \frac{1}{T} \varphi\left(\frac{y}{T}\right) dy = \\ &= MT^2 - 2MT^2 \int f'(y) \frac{1}{T} \varphi\left(\frac{y}{T}\right) dy + \int f(y)^2 \frac{1}{T} \varphi\left(\frac{y}{T}\right) dy \end{aligned}$$

mit f' möglicherweise einer Distributionsableitung von f . Dabei gilt

$$\begin{aligned} &\int f'(y) \frac{1}{T} \varphi\left(\frac{y}{T}\right) dy = \sum_{i=1}^n \alpha_i \left(\int_{t_{i-1}}^{t_i} \frac{\varphi\left(\frac{y}{T}\right)}{T d_i} dy - \int_{t_i}^{t_{i+1}} \frac{\varphi\left(\frac{y}{T}\right)}{T d_{i+1}} dy \right) + \\ &\quad + (-b) \int_{t_0}^{t_1} \frac{\varphi\left(\frac{y}{T}\right)}{T(-d_1)} dy + b \int_{t_n}^{t_{n+1}} \frac{\varphi\left(\frac{y}{T}\right)}{T d_{n+1}} dy = \\ &= \sum_{i=1}^n \alpha_i \left(\frac{d\Phi_i}{d_i} - \frac{d\Phi_{i+1}}{d_{i+1}} \right) + \frac{b d\Phi_1}{d_1} + \frac{b d\Phi_{n+1}}{d_{n+1}} =: \sum_{i=1}^n \alpha_i B_{1,i} + C_1 \end{aligned}$$

$$\begin{aligned} &\int f(y)^2 \frac{1}{T} \varphi\left(\frac{y}{T}\right) dy = \int \left(\sum_{i=1}^n \alpha_i \left[\frac{I_i(y-t_{i-1})}{d_i} - \frac{I_{i+1}(y-t_{i+1})}{d_{i+1}} \right] - \right. \\ &\quad \left. - \frac{b I_1(y-t_1)}{d_1} - b I_0 + \frac{b I_{n+1}(y-t_n)}{d_{n+1}} + b I_{n+2} \right)^2 \frac{\varphi\left(\frac{y}{T}\right)}{T} dy = \\ &= \int \left\{ \sum_{i=1}^n \left(\alpha_i^2 \left[\frac{I_i(y-t_{i-1})^2}{d_i^2} + \frac{I_{i+1}(y-t_{i+1})^2}{d_{i+1}^2} \right] - \right. \right. \\ &\quad \left. \left. - \alpha_i \alpha_{i-1} I_{i>1} \frac{I_i(y-t_{i-1})(y-t_i)}{d_i^2} - \alpha_i \alpha_{i+1} I_{i<n} \frac{I_i(y-t_{i+1})(y-t_i)}{d_{i+1}^2} \right) + \right. \\ &\quad \left. + 2b \left[\frac{I_1(y-t_1)(y-t_0)}{d_1^2} \alpha_1 + \frac{I_{n+1}(y-t_n)(y-t_{n+1})}{d_{n+1}^2} \alpha_n \right] + \right. \end{aligned}$$

$$\begin{aligned}
& + b^2 \left[\frac{I_1(y-t_1)^2}{d_1^2} + I_0 + \frac{I_{n+1}(y-t_n)^2}{d_{n+1}^2} + I_{n+2} \right] \left\} \frac{\varphi(\frac{y}{T})}{T} dy = \\
& = \sum_{i=1}^n Z_{1,i} + Z_{2,i} + Z_{3,i} + Z_{4,i} + Z_5 + Z_6 + Z_7 + Z_8 + Z_9
\end{aligned}$$

mit

$$\begin{aligned}
Z_{1,i} & := \alpha_i^2 \frac{1}{d_i^2} [(T^2 + t_{i-1}^2) d\Phi_i - T dy \varphi_i + 2t_{i-1} T d\varphi_i] \\
Z_{2,i} & := \alpha_i^2 \frac{1}{d_{i+1}^2} [(T^2 + t_{i+1}^2) d\Phi_{i+1} - T dy \varphi_i + 2t_{i+1} T d\varphi_{i+1}] \\
S_i & := -\alpha_i \alpha_{i-1} \frac{1}{d_i^2} [(T^2 + t_{i-1} t_i) d\Phi_i - T dy \varphi_i + (t_i + t_{i-1}) T d\varphi_i] \\
Z_{3,i} & := I_{i>1} S_i, \quad Z_{4,i} := I_{i<n} S_{i+1} \\
Z_5 & := 2b\alpha_1 \frac{1}{d_1^2} [(T^2 + t_0 t_1) d\Phi_1 - T dy \varphi_1 + (t_0 + t_1) T d\varphi_1] \\
Z_6 & := 2b\alpha_n \frac{1}{d_{n+1}^2} [(T^2 + t_{n+1} t_n) d\Phi_{n+1} - T dy \varphi_{n+1} + (t_n + t_{n+1}) T d\varphi_{n+1}] \\
Z_7 & := b^2 \frac{1}{d_1^2} [(T^2 + t_1^2) d\Phi_1 - T dy \varphi_1 + 2t_1 T d\varphi_1] \\
Z_8 & := b^2 \frac{1}{d_{n+1}^2} [(T^2 + t_n^2) d\Phi_{n+1} - T dy \varphi_{n+1} + 2t_n T d\varphi_{n+1}] \\
Z_9 & := b^2 (\Phi_0 + 1 - \Phi_{n+1})
\end{aligned}$$

Die Matrix A , die so entsteht, ist offenbar tridiagonal, was sich durch die speziellen Tentfunktionen als Basis erklärt. Als Hauptdiagonale der Matrix A ergibt sich $U := (Z_{1,i} + Z_{2,i})_{i=1, \dots, n}$, als obere und untere Nebendiagonale $V := (Z_{3,i})_{i=1, \dots, n-1}$. Für den Vektor B gilt $B = (B_i)_{i=1, \dots, n} = -2MT^2 B_{1,i} + Z_5 I_{i=1} + Z_6 I_{i=n}$, und $C = Z_7 + Z_8 + Z_9 - 2MC_1 + MT^2$. Für die Nebenbedingung rechnen wir:

$$\begin{aligned}
& \int f(y) \frac{1}{\sqrt{S}} \varphi\left(\frac{y-\varepsilon_j}{\sqrt{S}}\right) dy = \\
& = \int \left(\sum_{i=1}^n \alpha_i f_i(y) + f_0 + f_{n+1} \right) \frac{1}{\sqrt{S}} \varphi\left(\frac{y-\varepsilon_j}{\sqrt{S}}\right) dy = \\
& = \sum_{i=1}^n \frac{\alpha_i}{d_i} [(\varepsilon_j - t_{i-1}) d\Phi_{i,j} - \sqrt{S} d\varphi_{i,j}] - \frac{\alpha_i}{d_{i+1}} [(\varepsilon_j - t_{i+1}) d\Phi_{i+1,j} - \sqrt{S} d\varphi_{i+1,j}] + \\
& \quad + \frac{b}{d_1} [(\varepsilon_j - t_1) d\Phi_{1,j} - \sqrt{S} d\varphi_{1,j}] + \frac{b}{d_{n+1}} [(\varepsilon_j - t_n) d\Phi_{n+1,j} - \sqrt{S} d\varphi_{n+1,j}] + \\
& \quad + b(1 - \Phi_{n+1,j} - \Phi_{0,j}) =: \sum_{i=1}^n D_{j,i} \alpha_i + H_j
\end{aligned}$$

und E , F ergeben sich wie in der ersten Diskretisierung als $F = b - H$, $E = b + H$.

Der Startvektor α_0 ergibt sich als $\alpha_{0,i} := f_i^{\text{SO}}$ mit $f_i^{\text{SO}} = f^{\text{SO}}(t_i)$ und f^{SO} wieder als der SO-optimalen Schätzung.

8.6.8 ein Beispiel

Wir betrachten das Beispiel $\varepsilon, X \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}(0,1)$ und stellen die ‘‘Lemma 5’’-optimalen Lösungen f^{SO} und f^{AO} sowohl im y -Raum, als auch im ε -Raum,

also bedingt nach ε , dar. b ist dabei so gewählt, daß für die SO-optimalen Lösung ein Effizienzverlust von 10% eintritt.

Als numerisches Auswertungsgitter in ε - und in y -Richtung verwenden wir auf $[-5; 5] \times [-7; 7]$ ein äquidistantes Gitter der Dimension 400×400 . Das quadratische Optimierungsproblem ist dabei mit Unterstützung von C. Zillober mithilfe K. Schittkowskis Routine QL0001 aus der Bibliothek IMSL gelöst worden. Das Ergebnis findet sich in Abbildung 8.8 grafisch dargestellt.

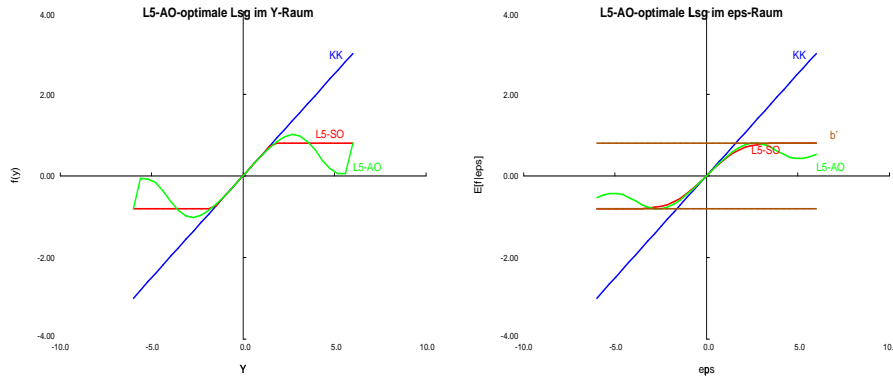


Abbildung 8.8: Es sind die Lösungen zu Problem (64) unter Nebenbedingung (16) bzw (65) zu einheitlichem $b = 1.27$ sowohl im y -Raum, als auch im ε -Raum dargestellt. Das Abknicken am Rand auf $\pm b$ im y -Raum suggeriert, daß die optimale Lösung am Rand auf 0 zurückgeht; dies hatten wir aber durch die Wahl unserer approximierenden Funktionen ausgeschlossen.

8.7 Minimax-Resultat für AO-Situation im reduzierten Modell

Wie im SO-Fall würden wir gerne ein Paar aus ungünstigster Verteilung \tilde{P}^ε für ε und dazu bestem Schätzer $f_1(Y)$ ermitteln, so daß diese beiden einen Sattelpunkt bilden.

8.7.1 Problemstellung

Im Unterschied zu Abschnitt 8.3 werden nun nicht die Y 's sondern die ε 's kontaminiert:

Formulierung des Minimax-Problems

Gesucht ist eine Schätzung $f_1(\hat{Y})$ für X , für die gilt

$$\sup_{\mathcal{V}_r(W)} E_P[|X - f(\hat{Y})|^2] = \inf_f \sup_{\mathcal{V}_r(W)} E_P[|X - f(\hat{Y})|^2] \quad (8.87)$$

mit $\mathcal{V}_r(W)$ aus (10)

Proposition 8.7.1 *Es gilt*

$$\inf_f \sup_{\mathcal{V}_r(W)} \mathbb{E}_P[|X - f(\hat{Y})|^2] = \sup_{\mathcal{V}_r(W)} \inf_f \mathbb{E}_P[|X - f(\hat{Y})|^2]$$

BEWEIS: wie Bemerkung 8.3.2 (b)

////

Nehmen wir an, die Verteilungen \hat{P}^Y für beliebige Wahlen von $\hat{P}^\varepsilon \in U_r(P^\varepsilon, W)$ seien dominiert durch ein σ -finites Maß $\mu \in \mathcal{M}_1(\mathbb{B}^p)$, also

$$\mu \gg \pi_2[\mathcal{V}_r(W)] \quad (8.88)$$

mit der Projektion π_2 auf die Koordinaten $(p+1, \dots, 2p)$ von (X, Y) . Unter dieser Voraussetzung lautet (87), mit Dichten formuliert:

Gesucht ist eine Schätzung $f_1(\hat{Y})$ für X , für die gilt

$$\begin{aligned} & \sup_{\tilde{P}^\varepsilon} \int |y - \varepsilon - f_1(y)|^2 p^X(y - \varepsilon) [r\tilde{P}^\varepsilon(d\varepsilon) + (1-r)P^\varepsilon(d\varepsilon)] \mu(dy) \\ & \stackrel{!}{=} \infsup_f \int |y - \varepsilon - f(y)|^2 p^X(y - \varepsilon) [r\tilde{P}^\varepsilon(d\varepsilon) + (1-r)P^\varepsilon(d\varepsilon)] \mu(dy), \end{aligned} \quad (8.89)$$

wobei wir $Y = X + \varepsilon$ benutzen und \tilde{P}^ε die Menge $\mathcal{M}_1(\mathbb{B}^k)$ durchläuft.

Bemerkung 8.7.2 (a) Wir können uns auf eine wesentliche Dominiiertheit von $\pi_2[\mathcal{V}_r(W)]$ durch ein σ -finites Produktmaß μ zurückziehen.

Dabei bedeute wesentlich, daß für jedes f und für jedes $P \in \mathcal{V}_r(W)$ ein weiteres $P' \in \mathcal{V}_r(W)$ existiert, dessen Randverteilungen $\pi_2 P'$ durch μ dominiert sind, so daß

$$\mathbb{E}_{P'}[|X - f(\hat{Y})|^2] \geq \mathbb{E}_P[|X - f(\hat{Y})|^2]$$

(b) (88) ist sicherlich erfüllt, falls P^X Lebesgue-stetig mit beschränkter Dichte p^X ist.

(c) Aus (88) folgt, daß wir für jedes $P \in \mathcal{V}_r(W)$ die gemeinsame Verteilung von (ε, \hat{Y}) schreiben können als

$$P(d\varepsilon, dy) = p^X(y - \varepsilon) \hat{P}^\varepsilon(d\varepsilon) \mu(dy) \quad (8.90)$$

mit p^X der μ -Dichte von P^X bzw.

$$P(d\varepsilon, dy) = p^X(y - \varepsilon) [r\tilde{P}^\varepsilon(d\varepsilon) + (1-r)P^\varepsilon(d\varepsilon)] \mu(dy) \quad (8.91)$$

mit dem kontaminierenden Maß \tilde{P}^ε .

(d) Wie in der Dominiiertheits-Voraussetzung im SO-Fall, (25), müssen wir (88) nur im idealen Modell nachweisen.

(e) Genauso gut wie X können wir auch ε auf Basis von Y zu schätzen versuchen, denn für den optimalen Schätzer für X aufgrund von Y , $\hat{f}(Y)$, gilt

$$\hat{f}(Y) = Y - \hat{g}(Y) \quad (8.92)$$

mit $\hat{g}(Y)$, dem optimalen Schätzer für ε mithilfe von Y .

Formulierung als funktionalanalytisches Problem

Wegen Proposition 8.7.1 wird wie im SO-Fall aus (89) das Problem, $E_P[|X - E_P[X|Y]|^2]$ über $P \in \mathcal{V}_r(W)$ zu maximieren.

Dies schreibt sich als

$$\int \frac{\left(\int (y-\epsilon) p^X(y-\epsilon) [r\tilde{P}^\epsilon(d\epsilon) + (1-r)P^\epsilon(d\epsilon)] \right)^2}{\int p^X(y-\epsilon) [r\tilde{P}^\epsilon(d\epsilon) + (1-r)P^\epsilon(d\epsilon)]} \mu(dy) \stackrel{!}{=} \inf_{\tilde{P}^\epsilon \in \mathcal{M}_1(\mathbb{B}^q)}, \quad (8.93)$$

Dieses können wir nun optimieren, indem wir längs Folgen $(\tilde{P}_n^\epsilon)_n$ argumentieren, längs denen wir mit μ_1 -Dichten p^ϵ , \tilde{p}^ϵ argumentieren dürfen, wobei μ_1 ein diese Folge dominierendes, σ -finites Maß ist.

Als nächstes zeigen wir, daß das Optimum nicht von der Wahl des Maßes μ_1 abhängt: Wir nehmen an, es gäbe bei Verwendung von $\mu_1^{(i)}$, $i = 1, 2$ zwei verschiedene Maße $P_i = P_i(\tilde{P}_i^\epsilon) \in \mathcal{V}_r(W)$ mit

$$E_{P_1}[|X - E_{P_1}[X|Y]|^2] = E_{P_2}[|X - E_{P_2}[X|Y]|^2] = \min_P.$$

Analog zur Argumentation für (42) erhalten wir die strikte Konvexität von $F(\tilde{P}^\epsilon) = E_{P(\tilde{P}^\epsilon)}[|E_{P(\tilde{P}^\epsilon)}[X|Y]|^2]$ in \tilde{P}^ϵ , so daß $E_{P_3}[|X - E_{P_3}[X|Y]|^2]$ mit $P_3 = \frac{1}{2}(\tilde{P}_1^\epsilon + \tilde{P}_2^\epsilon)$ noch kleiner wäre.

Mit einem ähnlichem Argument folgt im Fall, daß $\mathcal{L}(X) = \mathcal{L}(-X)$ und $\mathcal{L}(\epsilon) = \mathcal{L}(-\epsilon)$, aus (93) bereits, daß das ungünstigste \tilde{P}^ϵ ebenfalls symmetrisch sein muß, denn wir sehen, daß der Wert von (93) unverändert bleibt, wenn wir von $\tilde{P}_1^\epsilon = \mathcal{L}(\tilde{\epsilon}_1)$ übergehen zu $\tilde{P}_2^\epsilon = \mathcal{L}(-\tilde{\epsilon}_1)$. Weil aber das Problem strikt konvex in \tilde{P}^ϵ ist, folgt, daß (93) beim Übergang zum symmetrisierten $\tilde{P}_s^\epsilon = \frac{1}{2}(\tilde{P}_1^\epsilon + \tilde{P}_2^\epsilon)$ kleiner werden muß.

Falls außerdem $E_P|\epsilon|^2$ auf $\mathcal{V}_r(W)$ beschränkt ist, können wir, wie in Bemerkung 8.7.2(e) dargestellt, ausnutzen, daß für festes $P \in \mathcal{V}_r(W)$ gilt

$$E_P[X|Y] = Y - E_P[\epsilon|Y], \quad E_P[|X - E_P[X|Y]|^2] = E_P[|\epsilon - E_P[\epsilon|Y]|^2],$$

was unter der Beschränktheit von $E_P|\epsilon|^2$ zu

$$E_P[|X - E_P[X|Y]|^2] = E_P[|\epsilon|^2] - E_P[|E_P[\epsilon|Y]|^2]$$

wird und wir somit (93) lösen, indem wir $E_P[|E_P[\epsilon|Y]|^2]$ über $P \in \mathcal{V}_r(W)$ minimieren, also

$$\int \frac{\left(\int \epsilon p^X(y-\epsilon) [r\tilde{p}^\epsilon(\epsilon) + (1-r)p^\epsilon(\epsilon)] \mu_1(d\epsilon) \right)^2}{\int p^X(y-\epsilon) [r\tilde{p}^\epsilon(\epsilon) + (1-r)p^\epsilon(\epsilon)] \mu_1(d\epsilon)} \mu(dy) \stackrel{!}{=} \inf_{\tilde{P}^\epsilon}, \quad (8.94)$$

was den Vorzug hat, daß die unbekannte Dichte $\tilde{p}^\epsilon(\epsilon)$ nur in einer Variable, ϵ , und nicht in zweien, wie $y - x$ in (93), ausgewertet wird.

8.7.2 Zusammenhang zur Bayes-Schätzung

Lokationsparameters bei kontaminierter Priorverteilung, wie wir im vereinfachten Modell (1) verdeutlichen wollen:

Formulierung als Bayes-Problem

Wie in Bemerkung 8.7.2(e) gesehen, können wir ebenso gut wie X auch ε auf Basis von Y zu schätzen versuchen. Im Bayes-schen Setup gilt für dieses Problem, daß bei fester (Prior-)Verteilung P^ε von ε und Beobachtung gemäß der Lokationsfamilie $P^{Y|\varepsilon=\varepsilon}(\cdot) = P^X(\cdot - \varepsilon)$ der optimale Schätzer für ε auf Basis der Beobachtung $Y = y$ gerade der Posterior-Erwartungswert von $\mathcal{L}(\varepsilon|Y = y)$ ist.

Unserem Problem entspricht in Bayes-Sichtweise das Problem, das Bayes-Risiko über alle möglichen Priorverteilungen \hat{P}^ε der Konvexkontaminationsumgebung $U_r(P^\varepsilon)$ zu minimieren.

Zusammenhang zur Fisher-Information im Spezialfall $X \sim \mathcal{N}_p(0, \Sigma)$

Schick und Mitter (1994) weisen darauf hin, daß dieses Problem geschlossen noch ungelöst ist und geben einige Referenzen für Teillösungen.

Proposition 8.7.3 *Für den Fall, daß X normalverteilt ist, kann man bei quadratischer Verlustfunktion das jeweils minimale Bayes-Risiko $r(P^\varepsilon)$ bei beliebiger Priorverteilung P^ε schreiben als⁹*

$$r(P^\varepsilon) = \text{tr } \Sigma - \text{tr}(\mathcal{I}(\mathcal{N}_p(0, \Sigma) * P^\varepsilon)), \quad (8.95)$$

wobei \mathcal{I} die multivariate Lokations-Fisher-Information im Modell

$$\mathcal{P} := \{(\mathcal{N}_p(0, \Sigma) * P^\varepsilon)(\cdot - \theta) \mid \theta \in \mathbb{R}^k\},$$

gemäß Abschnitt B.2 Formel (18) ist.

BEWEIS: Um dies nachzuweisen, notieren wir, daß für φ_Σ , die Dichte von $\mathcal{N}_k(0, \Sigma)$,

$$\begin{aligned} |\varphi_\Sigma(x+h) - \varphi_\Sigma(x)| &= \left| \int_0^1 (x+sh)^\tau \Sigma^{-2} h \varphi_\Sigma(x+sh) \lambda^k(dx) \right| \leq \\ &\leq [(h^\tau \Sigma^{-3/2} h) / ((2\pi)^k \det \Sigma)]^{\frac{1}{2}} \leq K_1 |h|, \end{aligned}$$

gilt, denn $|x| \exp(-|x|^2/2)$ wird maximal für $|x| = 1$, und entsprechend

$$\begin{aligned} |\Sigma^{-1}(x+h)\varphi_\Sigma(x+h) - \Sigma^{-1}x\varphi_\Sigma(x)| &\leq |\Sigma^{-1}x| |\varphi_\Sigma(x+h) - \varphi_\Sigma(x)| + |h| \varphi_\Sigma(x+h) \leq \\ &\leq K_2 |h|, \end{aligned}$$

weil $|x|^2 \exp(-|x|^2/2)$ beschränkt ist. Daher können wir wegen dominierter Konvergenz Integration und Differentiation vertauschen, und die Dichte von $\mathcal{N}(0, \Sigma) * P^\varepsilon$, $p^Y(y) = \int p^X(y - \varepsilon) P^\varepsilon(d\varepsilon)$, ist zweimal stetig differenzierbar. Daher dürfen wir die Lokations- L_2 -Ableitung einführen,

$$\Lambda^Y = \Lambda^Y(y, P^\varepsilon) = -\nabla p^Y(y, P^\varepsilon) / p^Y(y, P^\varepsilon)$$

Die entsprechenden Terme \hat{p}^Y , \tilde{p}^Y , $\hat{\Lambda}^Y$, $\tilde{\Lambda}^Y$ sind über die entsprechenden \hat{P}^ε bzw. \tilde{P}^ε definiert.

⁹Dieser Zusammenhang geht im Eindimensionalen auf Brown (1971) zurück, und wurde unabhängig voneinander als Ausgangspunkt für Untersuchungen in Bickel (1981), Marazzi (1985), Levit (1981) genommen.

Der Bayes-optimaler Schätzer für X bei quadratischer Verlustfunktion ist der Posterior-Erwartungswert. Mit

$$p^{Y|\varepsilon=\epsilon}(y) = \frac{1}{((2\pi)^p \det \Sigma)^{1/2}} \exp\left(-\frac{1}{2}(y - \epsilon)^\tau \Sigma^{-1}(y - \epsilon)\right)$$

der Dichte von $\mathcal{N}_p(\epsilon, \Sigma)$ schreibt sich dieser im Fall, daß $X \sim \mathcal{N}(0, \Sigma)$, als

$$\begin{aligned} \mathbb{E}[\varepsilon|Y=y] &= \frac{\int \varepsilon p^{Y|\varepsilon}(y) P^\varepsilon(d\varepsilon)}{\int p^{Y|\varepsilon}(y) P^\varepsilon(d\varepsilon)} = \Sigma \frac{\int \Sigma^{-1}(\varepsilon - y + y) p^{Y|\varepsilon}(y) P^\varepsilon(d\varepsilon)}{\int p^{Y|\varepsilon}(y) P^\varepsilon(d\varepsilon)} = \\ &= y + \Sigma \frac{\frac{\partial}{\partial y} \int p^{Y|\varepsilon}(y) P^\varepsilon(d\varepsilon)}{\int p^{Y|\varepsilon}(y) P^\varepsilon(d\varepsilon)} = y - \Sigma \Lambda^Y(y) \end{aligned} \quad (8.96)$$

Weiter gilt für irgendeine in Y differenzierbare Funktion $T(Y)$, deren bedingter Erwartungswert $\mathbb{E}[|T'(Y)||\varepsilon] < \infty$ für alle ε ist, wobei wir das Argument von T weglassen:

$$\begin{aligned} r(T, P^\varepsilon) &:= \int \mathbb{E}[|T - \varepsilon|^2 | \varepsilon] P^\varepsilon(d\varepsilon) = \\ &= \int \mathbb{E}[|T - Y|^2 | \varepsilon] + \mathbb{E}[|Y - \varepsilon|^2 | \varepsilon] + 2 \mathbb{E}[(T - Y)^\tau (Y - \varepsilon) | \varepsilon] P^\varepsilon(d\varepsilon) = \\ &= \mathbb{E}[|T - Y|^2] + \text{tr} \Sigma - 2 \int (T - y)^\tau \Sigma \nabla_Y p^{Y|\varepsilon}(y) P^\varepsilon(d\varepsilon) = \quad (8.97) \\ &= \mathbb{E}[|T - Y|^2] + \text{tr} \Sigma + 2 \int \sum_{i,j=1}^p \frac{\partial}{\partial y_j} [\Sigma_{i,j} (T_i - Y_i)] p^{Y|\varepsilon}(y) P^\varepsilon(d\varepsilon) = \\ &= \mathbb{E}[|T - Y|^2] - \text{tr} \Sigma + 2 \int \sum_{j=1}^p \frac{\partial}{\partial y_j} [(\Sigma T)_j] p^{Y|\varepsilon}(y) P^\varepsilon(d\varepsilon) = \\ &= \mathbb{E}[|T - Y|^2] - \text{tr} \Sigma + 2 \int \sum_{j=1}^p \mathbb{E}\left[\frac{\partial}{\partial y_j} [(\Sigma T)_j] | \varepsilon\right] P^\varepsilon(d\varepsilon) \end{aligned} \quad (8.98)$$

wobei der Schluß von (97) auf die nächste Zeile mit koordinatenweiser partieller Integration folgt.

Wegen der oben nachgewiesenen Vertauschbarkeit von Integration und Differentiation sehen wir ein, daß

$$0 = \frac{\partial}{\partial y} \frac{\partial}{\partial y^\tau} 1 = \frac{\partial}{\partial y} \frac{\partial}{\partial y^\tau} \int p^Y(y) dy = \int \frac{\frac{\partial}{\partial y} \frac{\partial}{\partial y^\tau} p^Y(y)}{p^Y(y)} p^Y(y) dy.$$

Damit erhalten wir, wobei wir die Argumente von f^Y weglassen,

$$\frac{\partial}{\partial y^\tau} (y - \Lambda^Y) = \mathbb{I}_p + \frac{\frac{\partial^2}{\partial y \partial y^\tau} p^Y}{p^Y} - \Lambda^Y \Lambda^{Y \tau}. \quad (8.99)$$

Setzen wir nun für $T(Y) = \mathbb{E}[\varepsilon|Y]$, so folgt mit dem bisherigen, daß

$$\begin{aligned} \int \sum_{j=1}^p \mathbb{E}\left[\frac{\partial}{\partial y_j} [(\Sigma T)_j] | \varepsilon\right] P^\varepsilon(d\varepsilon) &= \sum_{i,j=1}^p \mathbb{E}\left[\frac{\partial}{\partial y_j} \Sigma_{i,j} (Y_i + \Lambda_i^Y)\right] = \\ &\stackrel{(8.99)}{=} \text{tr} \Sigma + \mathbb{E}[\Lambda^{Y \tau} \Sigma \Lambda^Y] = \text{tr} \Sigma + \text{tr}(\mathcal{I}(\mathcal{N}_p(0, \Sigma) * P^\varepsilon)), \end{aligned} \quad (8.100)$$

$$\begin{aligned}
r(P^\varepsilon) &:= r(\mathbb{E}[\varepsilon|Y], P^\varepsilon) \stackrel{(8.96)}{=} \mathbb{E}[|Y + \Sigma\Lambda^Y - \varepsilon|^2] = \\
&\stackrel{(8.98)}{=} \mathbb{E}[\Lambda^Y, \tau\Sigma\Lambda^Y] - \text{tr}\Sigma + 2 \int \sum_{j=1}^p \mathbb{E}\left[\frac{\partial}{\partial y_j} [(\Sigma T(Y))_j] \mid \varepsilon\right] P^\varepsilon(d\varepsilon) = \\
&\stackrel{(8.100)}{=} \text{tr}\Sigma - \text{tr}[\mathcal{I}(\mathcal{N}_p(0, \Sigma) * P^\varepsilon)].
\end{aligned}$$

////

Daher läuft unser Problem darauf hinaus, die Fisher-Information $\text{tr}(\mathcal{I}(\hat{P}^Y))$ über alle $\hat{P}^Y \in \mathcal{N}(0, \Sigma) * U_r(P^\varepsilon)$ zu minimieren.

Teilaspekte der Lösung im Spezialfall $X \sim \mathcal{N}_1(0, \sigma^2)$

Immerhin zeigen Bickel und Collins (1983) in ihren Theoremen 1 und 3, daß man für $X \sim \mathcal{N}(0, 1)$ noch folgende Aussagen treffen kann:

- Es gibt ein eindeutiges kontaminierendes Maß $\tilde{P}_1^\varepsilon \in \mathcal{M}_1(\mathbb{B})$, so daß dann $\hat{P}_1^Y := \mathcal{N}(0, 1) * [(1-r)P^\varepsilon + r\tilde{P}_1^\varepsilon]$ die Fisher-Information $\mathcal{I}(\hat{P}^Y)$ über alle $\hat{P}^Y \in \mathcal{N}(0, 1) * U_r(P^\varepsilon)$ minimiert.
- Die Dichte \hat{p}_1^Y von \hat{Y}_1 ist absolutstetig mit Ableitung $(\hat{p}_1^Y)'$ und $\hat{\Lambda}_1^Y := -(\hat{p}_1^Y)'/\hat{p}_1^Y$ existiert fast überall (Lebesgue).
- Der Träger von \tilde{P}_1^ε ist enthalten in der Menge $\{\varepsilon \mid J(\hat{\Lambda}_1^Y, \varepsilon) = \mathcal{I}(\hat{P}_1^Y)\}$ mit

$$J(\psi(Y), \varepsilon) = \begin{cases} \mathbb{E} [[2\hat{\Lambda}_1^Y(Y - \varepsilon) - \psi(Y)]\psi(Y) \mid \varepsilon] & \text{für } \mathbb{E}[\psi(Y)^2 \mid \varepsilon] < \infty \\ \infty & \text{sonst} \end{cases}$$

Dabei bedeutet $\Lambda_1^Y(Y - \varepsilon)$ die Auswertung von Λ_1^Y an der Stelle $Y - \varepsilon$.

- Der Träger von \tilde{P}_1^ε ist eine abzählbar unendliche Menge in \mathbb{R} mit $\pm\infty$ als einzig möglichen Häufungspunkten.

Unserem Problem entspricht dabei Beispiel (ii)(b) in Bickel und Collins (1983), wobei deren π_0 gerade unser P^ε ist. Da wir im allgemeinen nicht mit einer Varianz von 1 arbeiten, müssen wir unser Problem zunächst umstandardisieren, indem wir das Modell

$$\tilde{Y} = \tilde{X} + \tilde{\varepsilon} \tag{8.101}$$

mit $\tilde{\cdot} := \cdot/\sigma$ und $\sigma^2 = \text{Var} X$ betrachten. Die sich in (101) ergebende ungünstigste Verteilung \tilde{P}_0^ε wird in (1) zu $\tilde{P}^\varepsilon(d\varepsilon) := \tilde{P}_0^\varepsilon(\sigma d\varepsilon)$ und der entsprechende Schätzer \tilde{T} dann zu $T := \sigma\tilde{T}$.

Bemerkung 8.7.4 Mallows (1980) vermutet, daß die (positiven) Trägerpunkte von \tilde{P}_0^ε auf einem äquidistanten Gitter $\gamma_k = hk$, $h \in \mathbb{R}_+$, $k \in \mathbb{N}$ liegen und \tilde{P}_0^ε eine geometrische Wahrscheinlichkeit ρ^k darauf wirft; laut Schick und Mitter (1994) stellt Donoho in einer dem Autor nicht zugänglichen Arbeit (Donoho (1978)) die leicht modifizierte Vermutung mit Trägerpunkten der Form

$$\text{supp}(\tilde{P}_1^\varepsilon) = \{\gamma_k \mid \gamma_k = a + hk\}, \quad k \in \mathbb{N} \tag{8.102}$$

für gewisse $h \in \mathbb{R}_+$, $a \in \mathbb{R}_+$ auf.

8.7.3 Zusammenhang zum “Lemma 5”–Ansatz

Wie in Bickel und Collins (1983) in der Motivation zu deren Beispiel (ii)(b) dargestellt, ist die Lösung zum Problem, das minimale Bayes–Risiko über alle Prior–Verteilungen $\hat{P}^\varepsilon \in \mathcal{U}(r, P^\varepsilon)$ zu bestimmen, äquivalent zur Bestimmung des minimalen Bayes–Risikos über alle Prior–Verteilungen $\tilde{P}^\varepsilon \in \mathcal{M}_1(\mathbb{B})$ unter einer Schranke an das Bayes–Risiko in einer speziellen Prior–Verteilung P^ε .

Genausogut kann man natürlich auch argumentieren, um zu zeigen, daß das Problem äquivalent zur Bestimmung des minimalen Bayes–Risikos in einer speziellen Prior–Verteilung P^ε unter einer Schranke an das maximale Bayes–Risiko über alle Prior–Verteilungen $\tilde{P}^\varepsilon \in \mathcal{M}_1(\mathbb{B})$ ist.

Ein Bayes–Schätzer $f(Y)$ kann aber dann und nur dann ein endliches maximales Bayes–Risiko besitzen, wenn er einen beschränkten, bedingten Erwartungswert $E[f(Y)|\varepsilon]$ besitzt. Dies aber ist genau Nebenbedingung zu Problem (64).

Also muß jede Lösung eines Problems im “Lemma 5”–Ansatz mit einem gewissen b genau einer Lösung des Minimax–Problems entsprechen.

Bemerkung 8.7.5 (a) Dies gilt nicht nur für das Problem bei normalverteiltem X , sondern für beliebiges P^X , W , so daß (88) in Kraft ist, und für beliebige Dimension.

(b) Wir haben hiermit die Möglichkeit jede Minimax–Lösung zu gegebenem Radius r durch Lösung des “Lemma 5”–Problems für eine gewisse Schranke b zu ermitteln. Wir kennen jedoch — außer dessen Monotonie — nicht den funktionalen Zusammenhang $b \mapsto r(b)$. Zudem können wir auf diese Weise nicht die zugehörige, ungünstigste Kontamination \tilde{P}^ε ermitteln.

8.7.4 numerische Berechnungen

Wir wissen bereits, daß die SO–optimale Sattelpunkts–Lösung (\hat{P}_0, f_0) zum Radius r im AO–Problem nicht zulässig ist, aber als einen Ansatz können wir versuchen, \hat{P}_0 , genauer \hat{P}_0^Y durch $\pi_2 P_n$, $P_n \in \mathcal{V}_r(W)$ zu approximieren.

Als Beispiel wollen wir $P^X = P^\varepsilon = \mathcal{N}(0, 1)$ bei $r = 0.1$ betrachten. Wir verwenden dazu als Ansatz Verteilungen gemäß der Vermutung (102) von Donoho mit Parametern a , h , k , ρ .

Da wir hier ein Problem vor uns haben, dessen geometrische Struktur wir nicht kennen, verwenden wir den 3–Phasen–Algorithmus C.2.1 zum Auffinden des ungünstigsten \tilde{P}^ε :

Algorithmus 8.7.6

1. tue von $i = 1$ bis n_1

(a) ziehe unabhängig $a_i \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}(0, 2.25)$, $h_i \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}(0, 1/4)$, $\rho_i \stackrel{\text{u.i.v.}}{\sim} \text{ufo}(0, 1)$,

(b) bestimme $\tilde{P}_i^\varepsilon := \tilde{P}^\varepsilon(a_i, h_i, \rho_i)$,

(c) berechne \hat{P}_i mit Randverteilungen

$$\hat{P}_i^{(1)} := \mathcal{N}(0, 1), \quad \hat{P}_i^{(2)} := \hat{P}^X * [0.9\mathcal{N}(0, 1) + 0.1\tilde{P}_i^\varepsilon],$$

(d) werte $\text{MSE}_{\hat{P}_i}[E_{\hat{P}_i}[X|Y]]$ aus

2. fertige eine “Top- n_2 -Liste” mit den “besten” $\hat{P}_{(i_1)}, \dots, \hat{P}_{(i_{n_2})}$ an;

3. tue von $j = 1$ bis n_3
 - (a) ziehe mit $P(I_j = i_l) = c/(n_2 + 1 - l + j)$ ein "Zentrum" I_j aus der "Top- n_2 -Liste",
 - (b) ziehe unabhängig a_j, h_j, ρ_j mit Varianz proportional zu $1/j$ und Mittelwert a_{I_j}, h_{I_j}, \dots
 - (c) datiere die "Top- n_2 -Liste" auf

Ein Durchlauf dieses Algorithmus' für $n_1 = n_3 = 10000$ und $n_2 = 200$ ergab als optimale Werte für a, h und ρ

$$\hat{a} = 3.098, \quad \hat{h} = 0.0731, \quad \hat{\rho} = 0.0289$$

was Darstellung 8.9 ergibt.

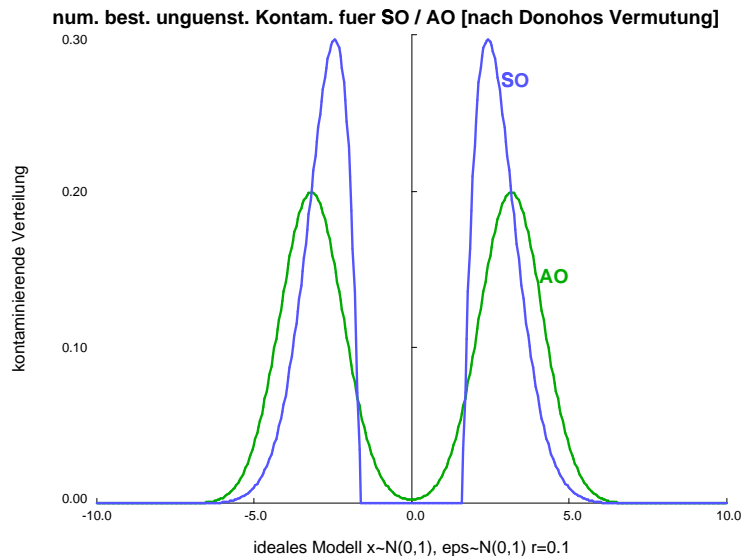


Abbildung 8.9: Im Fall $\epsilon, X \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}(0,1)$ stellen wir für Radius $r = 0.1$ der ungünstigsten SO-Kontamination \tilde{P}_0^Y aus Theorem 8.3.1 die numerisch als ungünstigst gefundene AO-Kontamination $\tilde{P}_1^Y = \mathcal{N}(0,1) * \tilde{P}_1^\epsilon$ gegenüber; bei der Suche nach \tilde{P}_1^Y verwenden wir den 3-Phasen-Algorithmus C.2.1 und unterstellen Donohos Vermutung (102). \tilde{P}_1^Y kann aus theoretischen Gründen \tilde{P}_0^Y nicht erreichen, kommt diesem aber im Rahmen der Möglichkeiten nahe.

8.8 Optimalität des rIC

Während die Untersuchungen in den Abschnitten 8.2, 8.3 und 8.4 eine (fast-)Optimalität des rLS-Filters für SO-Umgebungen gezeigt haben, soll nun untersucht werden, in welchem Sinn man auch den rIC-Filter als optimal innerhalb des Filterproblems auffassen kann. Wir werden hier nur einen "Lemma 5"-Ansatz für das SO-Problem präsentieren.

8.8.1 eine Cramér–Rao–Schranke für stochastische Parameter

Ein wesentliches Argument bei der Herleitung des RIC in Kapitel 3 war es, das Filterproblem als ein Parameter–Schätzproblem mit stochastischem Parameter zu interpretieren. Weil beim Filtern aber immer nur eine neue Beobachtung verarbeitet wird, besteht die Schwierigkeit darin, die asymptotischen Argumente wie lokale Fisher–Konsistenz, vgl. Lemma 9.4.7 bzw. Lemma 10.2.13, die im parametrischen Modell im eigentlichen Sinn — also mit deterministischen Parameter — zur Verfügung steht, zu ersetzen durch eine Argumentation, die nur mit einer Beobachtung auskommt. Für diese Zwecke wird in Abschnitt B.4 eine generalisierte Fisher–Information für das Schätzen eines stochastischen Parameters definiert. Mit dieser erhalten wir folgende

Proposition 8.8.1 *Seien $X, Y \in \mathbb{R}^p$ – bzw. \mathbb{R}^q – dimensionale Zufallsvariablen, deren gemeinsame Verteilung $P^{X,Y}$ durch das Produktmaß $\mu = \mu_1 \otimes \mu_2$ mit gewissen $\mu_{1[2]} \in \mathcal{M}_\sigma(\mathbb{B}^{p[q]})$ dominiert werde. Dabei seien $E|X|^2$, sowie die verallgemeinerte Fisher–Information $\mathcal{I}_{X|Y}$ von X gegeben Y endlich¹⁰. Definieren wir dann*

$$\Lambda_x(x, y) = \frac{\nabla_x p^{X|Y=y}}{p^{X|Y=y}}, \quad (8.103)$$

so gilt für alle Y –meßbaren Schätzer f für X , so daß $|f\Lambda_x| \in L_1(P^{X,Y})$ ist¹¹, die Korrelationsbedingung

$$E[(f(Y) - X)\Lambda_x(X, Y)^\tau] = \mathbb{I}_p \quad (8.104)$$

und die verallgemeinerte Cramér–Rao–Schranke¹²:

$$\text{MSE}(f(Y)) = \text{tr} \mathcal{I}_{X|Y}^{-1} + E|f(Y) - X - \mathcal{I}_{X|Y}^{-1} \Lambda_x|^2 \geq \text{tr} \mathcal{I}_{X|Y}^{-1}. \quad (8.105)$$

Dabei ist Gleichheit nur durch den One–Step–Schätzer

$$f_h := X + \mathcal{I}_{X|Y}^{-1} \Lambda_x \quad (8.106)$$

erreichbar und dieser ist nur zugelassen, falls er meßbar in Y ist. In jedem Fall ist der Schätzer

$$f_h^0 := E[X + \mathcal{I}_{X|Y}^{-1} \Lambda_x | Y] \quad (8.107)$$

optimal unter allen zulässigen Schätzern.

BEWEIS: Mit partieller Integration erhalten wir

$$\begin{aligned} \int x \nabla_x^\tau p^{X|Y=y}(x, y) \lambda^k(dx) P^Y(dy) &= -\mathbb{I}_p, \\ \int f(y) \nabla_x^\tau p^{X|Y=y}(x, y) \lambda^k(dx) P^Y(dy) &= 0 \end{aligned}$$

¹⁰Dabei heiße $\mathcal{I}_{X|Y}$ endlich, falls $\tilde{\mathcal{I}}_{X|Y}$ in (29) in Abschnitt B.4 endlich ist.

¹¹Solche Schätzer heißen im folgenden *zulässig*.

¹²Wie in der klassischen Cramér–Rao–Schranke gilt

$$E[(f(Y) - X)(f(Y) - X)^\tau] \succeq \mathcal{I}_{X|Y}^{-1}$$

und daher folgt (104). Für (105) sehen wir ein, daß

$$\begin{aligned} 0 &\preceq \int [f(y) - x - \mathcal{I}_{X|Y}^{-1} \Lambda_x][\dots]^\tau p^{X|Y=y}(dx) \lambda^k(dx) P^Y(dy) = \\ &= \int [f(y) - x][\dots]^\tau P^{X,Y}(dx, dy) - \mathcal{I}_{X|Y}^{-1} \end{aligned}$$

wobei “=” nur gelten kann, falls $f(Y) - X = \mathcal{I}_{X|Y}^{-1} \Lambda_x$ fast sicher $P^{X,Y}(dx, dy)$. Die Aussage über die Optimalität von (107) folgt aus (105) und der Optimalität des bedingten Erwartungswertes. ////

Bemerkung 8.8.2 Dies liefert eine allgemeine mehrdimensionale Formulierung der Cramér–Rao–Schranke ohne den Umweg über das *Transformed Innovation Distribution Model* aus Masreliez und Martin (1977).

Korollar 8.8.3 (a) *Gleichheit in (105) kann nur erreicht werden, wenn die faktorisierte, bedingte Verteilung $P^Y(dy)$ –fast–sicher von der Form $P^{X|Y=y} = \mathcal{N}(f_h(y), \mathcal{I}_{X|Y}^{-1})$ ist.*

(b) *Ist $Y = ZX + \varepsilon$ für eine Matrix $Z \in \mathbb{R}^{q \times p}$, X , ε stochastisch unabhängig, $\varepsilon \sim \mathcal{N}(0, V)$, so wird Gleichheit in (105) nur erreicht, wenn $(X, \varepsilon) \sim \mathcal{N}_{p+q}(0, \begin{pmatrix} \Sigma & 0 \\ 0 & V \end{pmatrix})$; in diesem Fall ist gerade $f_h(y) = My$.*

BEWEIS: (a): Wir wissen, daß $P^{X,Y}$ –fast sicher

$$\mathcal{I}_{X|Y}(f_h(y) - x) = \Lambda_x = \nabla_x \log p^{X|Y=y}.$$

Integration nach $\lambda^k(dx)$ zusammen mit der Tatsache, daß $\int p^{X|Y=y} \lambda^k(dx) = 1$ liefert die Behauptung; für (b) schreiben wir

$$\Lambda_x = \frac{\nabla_x(p^\varepsilon(y-Zx))}{p^\varepsilon(y-Zx)} + \frac{\nabla_x p^X(x)}{p^X(x)} = Z^\tau V^{-1}(y - Zx) + \frac{\nabla_x p^X(x)}{p^X(x)},$$

so daß

$$f_h = (\mathbb{I}_p - \mathcal{I}_{X|Y}^{-1} Z^\tau V^{-1} Z)X + \mathcal{I}_{X|Y}^{-1} \frac{\nabla_x p^X(x)}{p^X(x)} + Z^\tau V^{-1} y.$$

Damit f_h meßbar in Y ist, muß also $\frac{\nabla_x p^X(x)}{p^X(x)} = (Z^\tau V^{-1} Z - \mathcal{I}_{X|Y})X$ sein, also gilt $X \sim \mathcal{N}_p(0, \Sigma)$ mit $\Sigma := -(Z^\tau V^{-1} Z - \mathcal{I}_{X|Y})$. Dann aber muß notgedrungen $f_h = E[X|Y]$ sein, also

$$f_h(y) = (Z^\tau V^{-1} Z + \Sigma^{-1})^{-1} Z^\tau V^{-1} y = \hat{M}y$$

////

Mit Argumenten analog zu Proposition 8.8.1 erhalten wir sogar faktorisiert nach $Y = y$:

Proposition 8.8.4 *Seien X , Y wie in Proposition 8.8.1; dann gilt mit*

$$I_{X|Y} = I_{X|Y}(Y) = E[\Lambda_x \Lambda_x^\tau | Y] \tag{8.108}$$

$P^Y(dy)$ –fast sicher

$$E[f(Y) - X][\dots]^\tau | Y \succeq I_{X|Y} \tag{8.109}$$

BEWEIS: Wieder mit partieller Integration erhalten wir $P^Y(dy)$ -fast sicher

$$\begin{aligned}\int x \nabla_x^\tau p^{X|Y=y}(x, y) \lambda^k(dx) &= -\mathbb{I}_p, \\ \int f(y) \nabla_x^\tau p^{X|Y=y}(x, y) \lambda^k(dx) &= 0\end{aligned}$$

und daher

$$\begin{aligned}0 &\leq \mathbb{E} \left[[f(Y) - X - I_{X|Y}^{-1} \Lambda_x][\dots]^\tau | Y \right] = \\ &= \mathbb{E} \left[[f(Y) - X][\dots]^\tau | Y \right] - I_{X|Y}^{-1}\end{aligned}$$

////

8.8.2 “fast”-Optimalität des rIC

Wir wollen nun den mithilfe der Cramér-Rao-Schranke ausgezeichneten Filter $\mathbb{E}[X - I_{X|Y}^{-1} \Lambda_x | Y]$ aus (107) als Ausgangspunkt für die Argumentation nehmen: Betrachtet man das Problem, zu einem zulässigen Startschätzer \tilde{X}_0 für X den besten zulässigen One-Step $\hat{X} = \tilde{X}_0 + f^{\text{OS}}(Y)$, $Y = ZX + \varepsilon$ unter der Nebenbedingung $\sup |f^{\text{OS}}| < b$ zu finden, so können wir das Problem nach Proposition 8.8.4 darauf zurückführen,

$$\mathbb{E} \left[|\tilde{X}_0 + f(Y) - X - I_{X|Y}^{-1} \Lambda_x|^2 | Y \right] = \min_f !$$

zu lösen, was zunächst nach Lemma C.2.2 auf

$$\hat{X} = \tilde{X}_0 + (\mathbb{E}[X - \tilde{X}_0 + I_{X|Y}^{-1} \Lambda_x | Y]) \min \left\{ 1, \frac{b}{|\mathbb{E}[X - \tilde{X}_0 + I_{X|Y}^{-1} \Lambda_x | Y]|} \right\}$$

führt. Machen wir aber dabei schon den Ansatz, daß

$$\tilde{X}_0 = X_0 + A \Lambda_{X_0} w, \quad w = \min \left\{ 1, \frac{b_0}{|A \Lambda_{X_0}|} \right\} \quad (8.110)$$

wobei wir $\Lambda_{X_{[0]}}$ für $\Lambda_x(X_{[0]}(Y), Y)$ schreiben, so ist die Frage nach dem optimalen A_0 , so daß

$$\mathbb{E}[|X - X_0 - A \Lambda_{X_0} w|^2] = \min_A !$$

Im Unterschied zum klassischen, parametrischen Fall ist hier Bedingung (104) wenig hilfreich, denn diese mittelt über alle Werte von X aus, so daß nicht nur biasfreie, sondern jeder zulässige Schätzer in Y (104) erfüllt.

Wie im Beweis von Proposition 8.8.4 sehen wir, daß das Minimum an derselben Stelle A_0 angenommen wird wie das Minimum von

$$\mathbb{E}[|X - X_0 + I_{X|Y}^{-1} \Lambda_x - A \Lambda_{X_0} w|^2] = \min_A !$$

Im folgenden machen wir die Annahme, Λ_X sei zweimal differenzierbar in X und werden benötigen, daß Λ_X “fast” von einem Gaußschen X stammt, in dem Sinn, daß $\|\Lambda_X''\|_\infty < \delta$ ist — vgl. auch den Stetigkeitssatz aus Martin (1979).

$$\Lambda_{X_0} = \Lambda_X + \nabla_x^\tau \Lambda_X (X_0 - X) + r_2,$$

mit Restglied $\sup_{x,y} |r_2| \leq \delta |X - X_0|^2$. Wir bemerken aber, daß wir die Ableitung von Λ_x im noch zu bestimmenden Filter nicht verwenden, sondern nur auszunutzen, daß $E[\nabla_x^\tau \Lambda_X | Y = y] = -I_{X|Y}^{-1}$ und $E[x_i \partial_{x_i} \Lambda_X | Y] = 0$, was jeweils mit partieller Integration folgt, um

$$\begin{aligned} & E[X - X_0 + I_{X|Y}^{-1} \Lambda_X - A \Lambda_{X_0} w | Y] = \\ &= E[X - X_0 + I_{X|Y}^{-1} (\Lambda_X - \Lambda_{X_0}) + (I_{X|Y}^{-1} - Aw) \Lambda_{X_0} | Y] = \\ &= E[r_2 + (I_{X|Y}^{-1} - Aw) \Lambda_{X_0} | Y] \end{aligned}$$

zu erhalten. Setzen wir

$$B := X + I_{X|Y}^{-1} \Lambda_X, \quad C := X_0 + A \Lambda_{X_0} w,$$

so ergibt sich wegen der Y -Meßbarkeit von C

$$E|B - C|^2 = E|B|^2 - 2E[E[B - C|Y]^\tau C] - E|C|^2,$$

wobei $E|B|^2$ irrelevant für die Optimierung in A ist. Wenn wir r_2 vernachlässigen, ist daher zu minimieren

$$E|Aw \Lambda_{X_0}|^2 - 2E\Lambda_{X_0}^\tau I_{X|Y}^{-1} A \Lambda_{X_0} w = \min_A !,$$

was äquivalent ist zu

$$E|(Aw - I_{X|Y}^{-1}) \Lambda_{X_0}|^2 = \min_A ! \quad (8.111)$$

Ist nun X_0 so gewählt, daß $E[I_{X|Y}^{-1} \Lambda_{X_0} \Lambda_{X_0}^\tau] = \mathbb{I}_p$ — was im normalverteilten Fall automatisch erfüllt ist —, und verlangen wir Optimalität¹³ im Sinne der leicht modifizierten Normalgleichung

$$(A_0 w - I_{X|Y}^{-1}) \Lambda_{X_0} \perp A \Lambda_{X_0} \quad \text{in } L_2^p(P) \quad \forall A \in \mathbb{R}^{p \times p} \quad (8.112)$$

$$\iff E[(A_0 w - I_{X|Y}^{-1}) \Lambda_{X_0} \Lambda_{X_0}^\tau] = 0 \quad (8.113)$$

so erhalten wir gerade die Korrelationsbedingung

$$E[A \Lambda_{X_0} w \Lambda_{X_0} w] \stackrel{!}{=} \mathbb{I}_p \quad (8.114)$$

Insgesamt ergibt sich

$$\hat{X} = X_0 + A \Lambda_{X_0} w + ([X_0 + I_{X|Y}^{-1} \Lambda_{X_0}] - [X_0 + A w \Lambda_{X_0}]) w_2, \quad (8.115)$$

$$w_2 = \min\left\{1, \frac{b}{|[X_0 + I_{X|Y}^{-1} \Lambda_{X_0}] - [X_0 + A w \Lambda_{X_0}]|}\right\}, \quad (8.116)$$

also als Korrektur zum Startwert X_0 eine Summe aus einem geclippten, umstandardisierten Inkrement, das dem AO-Teil $A \Lambda_{X_0} w_{AO}$ im rIC-Filter entspricht, und einem ebenfalls gestutzten Term, der die Abweichung dieses robustifizierten Inkrements zum klassisch optimalen darstellt, was im rIC-Filter dem IO-Teil $A \Lambda_{X_0} w_{IO}$ entspricht.

Durch geeignete Wahl von b , b_0 und der zum Stutzen betrachteten Norm erhält man die verschiedenen rIC-Varianten.

¹³Im Hinblick auf das vernachlässigte r_2 und die Verwendung von $A \Lambda_{X_0}$ statt $A \Lambda_{X_0} w$ als Tangenten müßte man korrekter von δ -Optimalität sprechen.

Teil II

robuste Schätzung der
Hyper-Parameter

Im ersten Teil der Arbeit haben wir durchweg die Kenntnis des Zustandsraummodells in Gestalt der Hyper-Parameter F_t, Z_t, Q_t, V_t vorausgesetzt. In vielen Anwendungsbeispielen — vor allem im ökonomischen Bereich, vgl. etwa Burmeister und Wall (1982), Watson (1989), aber auch Pecchi (1993) und Brand (1998) — sind jedoch nur die Dimensionen von Zustand und Beobachtung, sowie die Matrix Z bekannt.

Dafür geht man aber von einer Zeitinvarianz der Hyper-Parameter aus, und es gilt, diese aus den Beobachtungen als Nebenparameter mitzuschätzen.

In dem nun folgenden zweiten Teil dieser Arbeit geht es dann auch genau um diese Frage. Ein wichtiges Verfahren in diesem Kontext stammt aus einer Arbeit von Shumway und Stoffer (1982), die zur Hyper-Parameter-Schätzung den EM-Algorithmus verwenden.

Es wird in dieser Arbeit allerdings nicht begründet, warum der EM-Algorithmus auch in der Situation des Zustandsraummodells mit abhängigen Beobachtungen anwendbar ist.

Selbst bei unabhängigen Beobachtungen wäre der EM-Algorithmus aus Robustheitsgesichtspunkten zumindest zu hinterfragen, weshalb wir uns in Kapitel 9 zunächst mit den statistischen Eigenschaften des EM-Algorithmus' aus Sicht der Robustheit beschäftigen und Alternativen hierzu aufzeigen.

In Kapitel 10 sind dann weitere Überlegungen fällig, die es ermöglichen die Techniken aus Kapitel 9 auf Modelle mit Zeitabhängigkeit, insbesondere das Zustandsraummodell, zu übertragen.

Hierbei stellt sich die Frage, ob wir bei unbekanntem Systemmatrizen diese überhaupt aus den Beobachtungen schätzen können, sowie nach einer Optimalitätstheorie hierfür.

Kapitel 9

der EM–Algorithmus aus Sicht der asymptotischen Statistik

9.1 Motivation

Als einfach umzusetzendes Verfahren zur Schätzung von Parametern bei fehlenden Daten erfreut sich der EM–Algorithmus großer Beliebtheit. Das Prinzip hinter diesem Verfahren ist dabei die Maximierung der bedingten Likelihood.

Aus asymptotischer Sicht allerdings ist eher die L_2 –Ableitung des Modells interessant, denn sie liefert eine Charakterisierung asymptotisch optimaler Verfahren im Sinn von Minimaxtheorem und Faltungssatz — vgl. Abschnitt A.5.

Bei fehlenden Beobachtungen kann man nun die L_2 –Differenzierbarkeit des entsprechend gröber filtrierte Modells untersuchen und daraus auf die LAN–Eigenschaft schließen. Dazu liefern wir in Abschnitt 9.2.1 zunächst mit dem Wurzeldichtenkalkül eine effiziente Notation. Mit dieser führen wir dann den Begriff der L_2 –Differenzierbarkeit ein. Es erweist sich, daß sich die Eigenschaft der L_2 –Differenzierbarkeit bei einer Vergrößerung des Modells — wie zum Beispiel durch Missings — erhält, was es uns ermöglicht, in Abschnitt 9.4 entsprechende Cramér–Rao–Schranken, sowie notwendige asymptotische Entwicklungen für Minimax–Schätzer und Maximin–Tests herzuleiten.

Die L_2 –Ableitung ist dann auch wieder die Basis für die Definition von Influenzkurven in diesem Kontext in Definition 9.4.4, so daß man auch in diesem Kontext robust–optimale Verfahren charakterisieren kann. Im Beispiel eines bivariaten, normalen Skalenmodells bei teilweise fehlender zweiter Beobachtungskordinate untersuchen wir mit diesen Mitteln in Abschnitt 9.6 die aus dem EM–Algorithmus hervorgehende Schätzfolge.

9.2 Glattheit von parametrischen Modellen

Bevor wir auf den EM–Algorithmus eingehen, benötigen wir Begriffe aus der asymptotischen Statistik wie L_2 –Differenzierbarkeit, Loglikelihoodentwicklung und LAN–Eigenschaft, die mit den Namen Le Cam und Hájek verbunden sind.

9.2.1 Wurzeldichtenkalkül

Wir wollen bei unseren Überlegungen das elegante, auf Le Cam zurückgehende Wurzeldichtenkalkül aus Kapitel 2 in Rieder (1994) verwenden, um das explizite Konditionieren auf Nichtsingulär-Bereiche von Maßen so weit wie möglich zu vermeiden. Wir notieren daher:

Definition 9.2.1 Sei (Ω, \mathcal{A}) ein meßbarer Raum. Dann definieren wir für $k \in \mathbb{N}$ den reellen Hilbertraum der Wurzeldichten $\mathcal{L}_2^k(\mathcal{A})$ als die Menge der Paare

$$\mathcal{L}_2^k(\mathcal{A}) := \{\xi\sqrt{dP} \mid \xi \in L_2^k(P), P \in \mathcal{M}_b(\mathcal{A})\} \quad (9.1)$$

Auf dieser Menge identifizieren wir Paare $\xi\sqrt{dP}$, $\eta\sqrt{dQ}$ vermöge der Äquivalenzrelation

$$\xi\sqrt{dP} \equiv \eta\sqrt{dQ} \iff \int |\xi\sqrt{p} - \eta\sqrt{q}|^2 d\mu = 0, \quad (9.2)$$

wobei $|\cdot|$ die euklidische Norm im \mathbb{R}^k und $\mu \in \mathcal{M}_b(\mathcal{A})$ irgendein P und Q dominierendes Maß ist und p, q die μ -Dichten von P und Q .

Auf der Menge der Äquivalenzklassen erklären wir die Vektorraumoperationen als

$$\alpha\xi\sqrt{dP} + \beta\eta\sqrt{dQ} := (\alpha\xi\sqrt{p} + \beta\eta\sqrt{q})\sqrt{d\mu} \quad (9.3)$$

für $\alpha, \beta \in \mathbb{R}$. Schließlich definieren wir ein Skalarprodukt auf diesem Vektorraum als

$$\langle \xi\sqrt{dP} \mid \eta\sqrt{dQ} \rangle := \int \xi^\tau \eta \sqrt{pq} d\mu \quad (9.4)$$

9.2.2 L_2 -Differenzierbarkeit in glatten Modellen

Wir wollen dieses Kalkül im Zusammenhang mit glatten parametrischen Familien verwenden. Sei also das Modell \mathcal{P} gegeben als

$$\mathcal{P} = \{P_\theta \mid \theta \in \Theta\} \subset \mathcal{M}_1(\mathcal{A}), \quad (9.5)$$

eine Familie von Wahrscheinlichkeitsmaßen auf dem meßbaren Raum (Ω, \mathcal{A}) ; diese wird parametrisiert durch θ aus einem offenen k -dimensionalen Parameterbereich $\Theta \subset \mathbb{R}^k$. Dementsprechend schreiben wir auch $E_\theta[\cdot]$ für $E_{P_\theta}[\cdot]$. Mit dem Wurzeldichtenkalkül können wir L_2 -Differenzierbarkeit von \mathcal{P} definieren:

Definition 9.2.2 Das Modell \mathcal{P} heißt L_2 -differenzierbar in θ , falls es eine Funktion $\Lambda_\theta \in L_2^k(P_\theta)$ gibt, so daß für $t \rightarrow 0$

$$\|\sqrt{dP_{\theta+t}} - \sqrt{dP_\theta}(1 + \frac{1}{2}t^\tau \Lambda_\theta)\|_{\mathcal{L}_2(\mathcal{A})} = o(t) \quad (9.6)$$

und für alle $t \in \mathbb{R}^k$

$$t^\tau \Lambda_\theta \stackrel{[P_\theta]}{=} 0 \implies t = 0. \quad (9.7)$$

Die Funktion Λ_θ heißt L_2 -Ableitung, und wir nennen die $k \times k$ -Matrix \mathcal{I}_θ

$$\mathcal{I}_\theta := E_\theta[\Lambda_\theta \Lambda_\theta^\tau] \quad (9.8)$$

Fisher-Information von \mathcal{P} in θ .

9.3 Glattheit von Modellen für gröbere σ -Algebren

9.3.1 Wurzeldichtenkalkül für gröbere σ -Algebren

Diese Begriffe werden jetzt umgesetzt für die Situation, in der nur eine gröbere σ -Algebra zur Verfügung steht. Sei dazu weiter (Ω, \mathcal{A}) unser meßbarer Raum, $\mathcal{U} \subset \mathcal{A}$ eine Unter- σ -Algebra. Dann schreiben wir $\mathcal{P}^{\mathcal{U}}$ für das restringierte Modell

$$\mathcal{P}^{\mathcal{U}} = \{P_{\theta}^{\mathcal{U}} \mid \theta \in \Theta\} \subset \mathcal{M}_1(\mathcal{U}), \quad (9.9)$$

wobei wir jedes $P_{\theta}^{\mathcal{U}}$ auffassen können als Restriktion eines Maßes $P_{\theta} \in \mathcal{M}_1(\mathcal{A})$ auf $\mathcal{M}_1(\mathcal{U})$. Ebenso verstehen wir zu einem dominierenden Maß $\mu \in \mathcal{M}_1(\mathcal{A})$ unter $p_{\theta}^{\mathcal{U}}$ die μ -Dichte der Einschränkung von P_{θ} auf $\mathcal{M}_1(\mathcal{U})$. Diese läßt sich auch als bedingter Erwartungswert schreiben: $p_{\theta}^{\mathcal{U}} = E[p_{\theta}|\mathcal{U}]$, wie man einfach mithilfe der Radon-Nikodym-Gleichung sieht. Dementsprechend fassen wir auch $dP_{\theta}^{\mathcal{U}}$ auf als $E_{\mu}[dP_{\theta}|\mathcal{U}]$. Eine weitere wichtige Konsequenz ergibt sich:

Lemma 9.3.1 Sei X eine $(\Omega, \mathcal{A}) - (\mathbb{R}^k, \mathbb{B}^k)$ -meßbare Zufallsgröße. Dann gilt

$$E_{\mu}[X dP_{\theta}|\mathcal{U}] = E_{P_{\theta}}[X|\mathcal{U}] E_{\mu}[dP_{\theta}|\mathcal{U}] \quad (9.10)$$

BEWEIS: Wir gehen zu μ -Dichten über und sehen, daß für alle $A \in \mathcal{U}$ gilt:

$$\begin{aligned} \int_A E_{\mu}[X p_{\theta}|\mathcal{U}] d\mu &= \int_A X p_{\theta} d\mu = \int_A X dP_{\theta} = \\ &= \int_A E_{P_{\theta}}[X|\mathcal{U}] dP_{\theta} = \int_A E_{P_{\theta}}[X|\mathcal{U}] dP_{\theta}^{\mathcal{U}} = \int_A E[X|\mathcal{U}] p_{\theta}^{\mathcal{U}} d\mu \end{aligned}$$

////

9.3.2 L_2 -Differenzierbarkeit für gröbere σ -Algebren

Im unabhängig identisch verteilten Fall überträgt sich die L_2 -Differenzierbarkeit auf das restringierte Modell, denn es gilt

Theorem 9.3.2 Ist \mathcal{P} L_2 -differenzierbar in θ mit Ableitung Λ_{θ} und gilt für $t \in \mathbb{R}^k$

$$t^{\tau} E_{\theta}[\Lambda_{\theta}|\mathcal{U}] \stackrel{[P_{\theta}]}{=} 0 \implies t = 0, \quad (9.11)$$

so ist es auch $\mathcal{P}^{\mathcal{U}}$ mit Ableitung $\Lambda_{\theta}^{\mathcal{U}} := E_{\mu}[\Lambda_{\theta}|\mathcal{U}]$, wobei μ ein W -Maß ist, das alle P_{θ_n} längs einer konvergenten Folge $\theta_n = \theta + h_n$, $h_n \rightarrow 0$, $h_n/|h_n| \rightarrow h$, sowie P_{θ_0} mit $\theta_0 = \theta$ dominiert. Diese Ableitung ist eindeutig $P_{\theta_0}^{\mathcal{U}}$ -f.s.

BEWEIS: Rieder und Ruckdeschel (2000a)

////

Bemerkung 9.3.3 (a) Dieses Resultat findet sich in Witting (1985), dort Satz 1.193. Allerdings wird im Beweis dort der Satz von Pratt verwendet, ohne daß dessen Voraussetzungen nachgewiesen werden. Durch eine andere Argumentation mit dem Satz von Vitali kann man diese Lücke zwar umgehen, der hier zitierte Beweis jedoch verwendet elementarer das Vallée-Poussin Kriterium.

(b) Einen gänzlich anderen Beweis dieser Aussage, der sich allerdings nicht ohne weiteres auf den Fall von Dreiecksschemata übertragen läßt, geben Le Cam und Yang (1988), Proposition 4.

9.3.3 L_2 -Differenzierbarkeit für Dreiecksschemata

Bei der Behandlung von Missings wollen wir zwei Herangehensweisen präsentieren: Einmal Missings an zufälligen Beobachtungen und einmal an bekannten, deterministischen Stellen. Bei letzterer liegt bereits bei den einfachsten anvisierten Beispielen keine Situation identisch verteilter Daten mehr vor: einige Beobachtungen werden "vollständig" vorliegen, andere nicht. Daher müssen wir den Begriff der L_2 -Differenzierbarkeit zumindest auf Dreiecksschemata wie im zentralen Grenzwertsatz von Lindeberg-Feller verallgemeinern. Um aber trotzdem von Optimalität reden zu können, verlangen wir jedoch im Unterschied zur Darstellung in Rieder (1994), daß die "standardisierten" Fisher-Informationen konvergieren. Wir verschenken aber nichts, wenn wir statt einer Standardisierung mit $1/\sqrt{n}$ auch die Standardisierung mit einer deterministischen Folge symmetrischer, positiv definiter Matrizen $(s_n)_n \subset \mathbb{R}^{k \times k}$, $s_n \rightarrow 0$ zulassen und somit den Fall der Regression mit deterministischen Regressoren abdecken. In letzterem Fall wäre $s_n := (\sum_{i=1}^{i_n} x_{n,i} x_{n,i}^\tau)^{-\frac{1}{2}}$.

Sei also

$$\mathcal{P}_{n,i} = \{P_{n,i,\theta} \mid \theta \in \Theta\} \subset \mathcal{M}_1(\mathcal{A}_{n,i}) \quad (9.12)$$

ein Dreiecksschema parametrisierter Wahrscheinlichkeiten auf einer Folge von Stichprobenräumen $(\Omega_{n,i}, \mathcal{A}_{n,i})$ mit offener Parametermenge $\Theta \subset \mathbb{R}^k$.

Definition 9.3.4 *Ein parametrisiertes Dreiecksschema $\mathcal{P}_{n,i}$ heie L_2 -differenzierbar in θ , falls es ein Dreiecksschema von Funktionen $\Lambda_{n,i,\theta} \in L_2^k(P_{n,i,\theta})$ gibt, so da gilt*

(L2-1)

$$\mathbb{E}_{n,i,\theta}[\Lambda_{n,i,\theta}] = 0 \quad \forall i = 1, \dots, i_n, \forall n \geq 1 \quad (9.13)$$

(L2-2) *Es gibt eine deterministische Folge symmetrischer, positiv definiter Matrizen $(s_n)_n \subset \mathbb{R}^{n \times n}$, $s_n \rightarrow 0$, so da mit einer endlichen, symmetrischen, positiv definiten Matrix $\mathcal{I}_\theta \in \mathbb{R}^{n \times n}$*

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{i_n} \mathbb{E}_{n,i,\theta}[s_n \Lambda_{n,i,\theta} \Lambda_{n,i,\theta}^\tau s_n] = \mathcal{I}_\theta \quad (9.14)$$

(L2-3) *Fr alle $\varepsilon \in \mathbb{R}_{>0}$ und alle $t \in \mathbb{R}^k$ gilt mit s_n aus (L2-2)*

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{i_n} \mathbb{E}_{n,i,\theta}[\mathbb{I}_{\{|t^\tau s_n \Lambda_{n,i,\theta}| > \varepsilon\}} |t^\tau s_n \Lambda_{n,i,\theta}|^2] = 0 \quad (9.15)$$

(L2-4) *Fr alle $t \in \mathbb{R}^k$ gilt mit*

$$\theta_n := \theta + s_n t \quad (9.16)$$

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{i_n} \|\sqrt{dP_{n,i,\theta_n}} - \sqrt{dP_{n,i,\theta}}(1 + \frac{1}{2} t^\tau s_n \Lambda_{n,i,\theta})\|_{\mathcal{L}_2(\mathcal{A}_{n,i})}^2 = 0 \quad (9.17)$$

Die Matrix \mathcal{I}_θ heie Fisher-Information in θ von $\mathcal{P}_{n,i}$ und das Schema $\Lambda_{n,i,\theta}$ heie L_2 -Ableitung von $\mathcal{P}_{n,i}$ in θ .

Die korrespondierenden Produktmaße auf $(\times_i \Omega_{n,i}, \otimes_i \mathcal{A}_{n,i})$ seien

$$P_\theta^{(n)} := \bigotimes_{i=1}^{i_n} P_{n,i,\theta}, \quad P_{\theta_n}^{(n)} := \bigotimes_{i=1}^{i_n} P_{n,i,\theta_n}, \quad (9.18)$$

und $x_{n,i} : \times_j \Omega_{n,j} \rightarrow \Omega_{n,i}$ sei die i -te kanonische Projektion. Hiermit gilt:

Theorem 9.3.5 *Falls das parametrisierte Dreiecksschema $\mathcal{P}_{n,i}$ in θ L_2 -differenzierbar ist mit L_2 -Ableitung $\Lambda_{n,i,\theta}$ und Fisher-Information \mathcal{I}_θ , so gilt für die Loglikelihoods der Alternativen, die durch (16) und (18) definiert sind*

$$\log \frac{dP_{\theta_n}^{(n)}}{dP_\theta^{(n)}} = t^\tau s_n \sum_{i=1}^{i_n} \Lambda_{n,i,\theta}(x_{n,i}) - \frac{1}{2} t^\tau \mathcal{I}_\theta t + o_{P_\theta^{(n)}}(n^0), \quad (9.19)$$

wobei

$$\left(s_n \sum_{i=1}^{i_n} \Lambda_{n,i,\theta}(x_{n,i}) \right) \circ P_\theta^{(n)} \xrightarrow{w} \mathcal{N}_k(0, \mathcal{I}_\theta) \quad (9.20)$$

BEWEIS: Rieder (1994), Theorem 2.2.9; wie sich die hier verwendete, andere Standardisierung umsetzt, ist klar. ////

9.4 Folgerungen für unvollständige Beobachtungen

Im letzten Abschnitt haben wir gesehen, daß für L_2 -differenzierbare Modelle die LAN-Eigenschaft folgt und damit die starken Optimalitätsresultate der asymptotischen Statistik wie der Faltungssatz A.5.5 und das as. Minimax-Theorem A.5.9 zur Verfügung stehen.

Andererseits haben wir in Abschnitt 9.3.2 gezeigt, daß im unabhängig identisch verteilten Fall sich die L_2 -Differenzierbarkeit auch auf gröber meßbare Modelle überträgt, und die Optimalitätsresultate somit dort ebenfalls gelten. Dies wollen wir im folgenden ausnutzen, um die Optimalitätsresultate auch auf die Situation von unvollständig beobachteten Messungen auszudehnen.

Da aber im allgemeinen nicht alle Beobachtungen im gleichen Maße unvollständig beobachtet werden, liegt es nahe, gleich m verschiedene "Grade" an Unvollständigkeit zu betrachten. Formal könnte man sagen, die Beobachtungen stammten aus m verschiedenen Modellen, wobei ein Index-Mechanismus bestimmt, aus welchem Modell die jeweilige Beobachtung stammt.

Um die Frage der L_2 -Differenzierbarkeit in dieser Situation zu beantworten müssen wir daher — einem Vorschlag im Seminar in Bayreuth aus dem WS 99/00 folgend — zunächst das m -Stichprobenproblem untersuchen.

9.4.1 L_2 -Differenzierbarkeit im m -Stichprobenproblem

Seien \mathcal{P}_l $l = 1, \dots, m$ L_2 -differenzierbare Modelle auf dem Maßraum (Ω, \mathcal{A}) mit einheitlichem Parameterraum Θ . Die L_2 -Ableitungen seien $\Lambda_{\theta,l}$, und in Abweichung zu Definition 9.2.2 dürfen die einzelnen Fisher-Informationen $\mathcal{I}_{\theta,l} =$

$E_\theta[\Lambda_{\theta,l}\Lambda_{\theta,l}^\tau]$ singular sein.

Wir beobachten Paare $(V, Y)_i$, $i = 1, \dots, m$ mit $V_i \in 1, \dots, m$ und

$$Y_i := \sum_{l=1}^m \mathbf{I}_{\{V_i=l\}} X_{i,l}, \quad X_{i,l} \stackrel{\text{u.i.v.}}{\sim} P_{\theta,l}, \quad P_{\theta,l} \in \mathcal{P}_l. \quad (9.21)$$

Ist nun der Auswahlmechanismus $\{V_i\}$ stochastisch unabhängig von den $X_{i,l}$ und sind die $V_i \stackrel{\text{u.i.v.}}{\sim} \text{Mult}(m, \underline{p})$ mit $\underline{p} = (p_1, \dots, p_m)$ bekannt, $p_l \geq 0$, $\sum_l p_l = 1$, so gilt:

Proposition 9.4.1 *Die Paare $(V, Y)_i$ sind identisch verteilt, und das Modell $\mathcal{Q} := \mathcal{L}_\theta((V, Y))$ ist L_2 -differenzierbar, sofern $\sum_l p_l \mathcal{I}_{\theta,l} \succ 0$ ist.*

BEWEIS: Zunächst zur Unabhängigkeit: Seien $n \in \mathbb{N}$, $1 \leq i_1 < \dots < i_n$; da die Mengen $\times_{j=1}^n A_{i_j}$, $A_{i_j} = I_{i_j} \times J_{i_j}$ mit $I_{i_j} \in \{1, \dots, m\}$, $J_{i_j} \in \mathcal{A}$, einen durchschnittstabilen Erzeuger von $(2^{\{1, \dots, m\}} \otimes \mathcal{A})^{\otimes n}$ bilden, genügt es die Unabhängigkeit für solche $\times_{j=1}^n A_{i_j}$ nachzuweisen.

$$\begin{aligned} P\left(\bigcap_{j=1}^n (V, Y)_{i_j} \in A_{i_j}\right) &= P\left(\bigcap_{j=1}^n [V_{i_j} = I_{i_j} \cap Y_{i_j} \in J_{i_j}]\right) = \\ &= P\left(\bigcap_{j=1}^n [V_{i_j} = I_{i_j} \cap X_{i_j, I_{i_j}} \in J_{i_j}]\right) = \prod_{j=1}^n P(V_{i_j} = I_{i_j} \cap X_{i_j, I_{i_j}} \in J_{i_j}) = \\ &= \prod_{j=1}^n P(V_{i_j} = I_{i_j} \cap Y_{i_j} \in J_{i_j}) = \prod_{j=1}^n P((V, Y)_{i_j} \in A_{i_j}) \end{aligned}$$

Andererseits gilt

$$\begin{aligned} P((V, Y)_i \in A_1) &= P(V_i = I_1 \cap Y_i \in J_1) = \\ &= P(V_i = I_1)P(X_{i, I_1} \in J_1) = p_{I_1} P_{\theta, I_1}(J_1), \end{aligned}$$

so daß die $(V, Y)_i$ identisch verteilt sind.

Sei ν das Zählmaß auf \mathbb{Z} . Dann gilt für ein generisches Element aus $dQ \in \mathcal{Q}_1$

$$Q(dv, dy) = dQ_\theta(dv, dy) = \sum_{l=1}^m p_l \mathbf{I}_{\{v=l\}} P_{\theta,l}(dy) \nu(dv)$$

Damit aber gilt in $\mathcal{L}_2(\mathcal{A})$ mit $\bar{\Lambda}_\theta(v, y) = \sum_{l=1}^m \mathbf{I}_{\{v=l\}} \Lambda_{\theta,l}(y)$

$$\begin{aligned} &\|\sqrt{dQ_{\theta+h}} - \sqrt{dQ_\theta}\left(1 + \frac{h^\tau}{2} \bar{\Lambda}_\theta\right)\|_{\mathcal{L}_2(\mathcal{A})} = \\ &= \left\| \sum_{l=1}^m \sqrt{p_l} \mathbf{I}_{\{v=l\}} [\sqrt{dP_{\theta+h,l}} - \sqrt{dP_{\theta,l}}\left(1 + \frac{h^\tau}{2} \bar{\Lambda}_{\theta,l}\right)] \right\|_{\mathcal{L}_2(\mathcal{A})} = o(|h|) \end{aligned}$$

Als Fisher-Information erhalten wir

$$\begin{aligned} \mathcal{I}_\theta &= E_{Q_\theta}[\bar{\Lambda}_\theta \bar{\Lambda}_\theta^\tau] = E_{Q_\theta}\left[\sum_{l=1}^m \mathbf{I}_{\{V=l\}} \Lambda_{\theta,l} \Lambda_{\theta,l}^\tau\right] = \\ &= \sum_{l=1}^m p_l E_{P_{\theta,l}}[\Lambda_{\theta,l} \Lambda_{\theta,l}^\tau] = \sum_{l=1}^m p_l \mathcal{I}_{\theta,l} \succ 0 \end{aligned}$$

////

9.4.2 Situation mit Missings als m -Stichprobenproblem

Wir gehen zunächst aus von unabhängig identisch verteilten Beobachtungen X_i , $X_i : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$, aus einem L_2 -differenzierbaren Modell

$$\mathcal{P} = \{P_\theta \in \mathcal{M}_1(\mathcal{A}) \mid \theta \in \Theta\}. \quad (9.22)$$

Die Situation mit m verschiedenen Typen von Missings fassen wir nun als spezielles m -Stichprobenproblem auf, wobei die \mathcal{P}_l nun die Gestalt $\mathcal{P}^{\mathcal{U}_l}$ haben, wobei \mathcal{U}_l Unter- σ -Algebren zu $\sigma(X)$ sind. Dies verallgemeinert die Situation, in der einzelne Beobachtungen ganz fehlen, was dann der trivialen σ -Algebra entspricht. Dennoch wollen wir solche unvollständigen Beobachtungen kurz Missings nennen.

In dieser Modellierung können wir bei jeder einzelnen Beobachtung entscheiden, ob es sich um ein Missing handelt, und wenn ja, gemäß welcher Unter- σ -Algebra die Beobachtung meßbar ist. Formal bilden wir also folgendes Modell: Es gibt einen idealen Beobachtungsprozeß von Beobachtungen X_i aus \mathcal{P} ; wir beobachten aber Paare (V_i, Y_i) , wobei die V_i über die Feinheit der Meßbarkeit der Y_i entscheiden.

Die Unter- σ -Algebren \mathcal{U}_l $l = 1, \dots, m$ denken wir uns als erzeugt durch ein m -Tupel $F = (F_1, \dots, F_m)^\tau$ X_i -meßbarer, möglicherweise nicht injektiver Funktionen $F_l : (\Omega', \mathcal{A}') \rightarrow (\Xi_l, \mathcal{B}_l)$; das vollständige Fehlen einer Beobachtung wird durch $\Xi = \{\perp\}$, $\mathcal{B} = \{\emptyset, \Xi\}$ dargestellt¹.

Bis hierhin haben wir außer der Meßbarkeit von (V_i, Y_i) noch keine Aussagen über die Variablen V_i gemacht. Halten wir uns an die Terminologie von Little und Rubin (1987), so wollen wir uns im folgenden auf die Situation beschränken, daß die Zeitpunkte, an denen Missings vorliegen durch einen u.i.v. Zufallsmechanismus entschieden wird, der vollständig unabhängig vom X -Prozeß ist [missing completely at random] (MCAR); damit bleiben wir in der Situation unabhängig identisch verteilter Daten, was das Kalkül erheblich vereinfacht; wir sind also in der Situation des vorigen Abschnittes, und das Modell mit Missings \mathcal{Q}_n schreibt sich als

$$\mathcal{Q}_n = \{Q_{n,\theta} = \mathcal{L}_\theta(V, Y)^{\otimes n} \mid \theta \in \Theta\}$$

Bemerkung 9.4.2 Die Tatsache, daß sich bei unabhängig identisch verteilten vollständigen Daten die LA(M)N-Eigenschaft auf gröber filtrierte Modelle überträgt, wird bereits in Le Cam und Yang (1988) bewiesen. In den Setup dieser Arbeit fallen auch die hier gemachten Aussagen; die Autoren geben überdies Bedingungen an, unter denen sogar die LA(M)N-Eigenschaft im gröber filtrierten Modell ausreicht, ohne Rückgriff auf die stärkere L_2 -Differenzierbarkeit zu nehmen.

Wir wollen aber hier im Kontext dieser L_2 -Differenzierbarkeit bleiben, denn dieser liefert uns zugleich mit der L_2 -Ableitung den Ansatzpunkt für den Begriff der Influenzkurve in Definition 9.4.4.

¹Die Verallgemeinerung auf die nicht identisch verteilte, unabhängige Situation, $X_i : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')_i$, $m = m_i$ und $F_l = F_{l,i}$, $F_{l,i} : (\Omega', \mathcal{A}')_i \rightarrow (\Xi_{l,i}, \mathcal{B}_{l,i})$ ist klar, würde aber die Notation erheblich verkomplizieren.

9.4.3 Umsetzung der allgemeinen asymptotischen Theorie

Im Fall von (MCAR)–Missings wollen wir nun Auswirkungen auf die Inferenz in einem L_2 –differenzierbaren, parametrischen Modell $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ betrachten. Als unmittelbare Folgerung von Theorem 9.3.2 und Proposition 9.4.1 erhalten wir:

Proposition 9.4.3 *Das Modell $\mathcal{P}^{\mathcal{U}} := \mathcal{Q} = \{\mathcal{L}_\theta(V, Y)\}$ ist in $\theta \in \Theta$ L_2 –differenzierbar mit L_2 –Ableitung $\Lambda_\theta^{\mathcal{U}}$ und Fisher–Information $\mathcal{I}_\theta^{\mathcal{U}} < \infty$, falls \mathcal{P} in θ L_2 –differenzierbar ist mit endlicher Fisher–Information \mathcal{I}_θ und $0 \prec \mathcal{I}_\theta^{\mathcal{U}} < \infty$. Dabei ist*

$$\Lambda_\theta^{\mathcal{U}} = \Lambda_\theta^{\mathcal{U}}(v, y) = \sum_l \mathbb{I}_{\{v=l\}} \Lambda_\theta^{\mathcal{U}_l}(y),$$

und als $\mathcal{I}_\theta^{\mathcal{U}}$ ergibt sich $\sum_l p_l \mathcal{I}_\theta^{\mathcal{U}_l}$

Definition 9.4.4 *Sei $\mathcal{P}^{\mathcal{U}}$ in $\theta \in \Theta$ L_2 –differenzierbar mit L_2 –Ableitung $\Lambda_\theta^{\mathcal{U}}$ und Fisher–Information $\mathcal{I}_\theta^{\mathcal{U}}$, und sei $D \in \mathbb{R}^{p \times k}$, $\text{rk } D = p$. Gilt für $\psi_\theta^{(1)}, \dots, \psi_\theta^{(m)}$*

- (a) $\psi_\theta^{(l)} \in L_2^k(P_\theta^{\mathcal{U}_l})$, $l = 1, \dots, m$
- (b) $\mathbb{E}_\theta[\psi_\theta^{(l)}] = 0$, $l = 1, \dots, m$
- (c) $\sum_l p_l \mathbb{E}_\theta[\psi_\theta^{(l)} \Lambda_\theta^{\mathcal{U}_l \tau}] = \mathbb{I}_k$ resp. D ,

dann heie der Ausdruck

$$\psi_\theta(v, y) := \sum_l \mathbb{I}_{\{v=l\}} \psi_\theta^{(l)}(y) \tag{9.23}$$

[partielle] Influenzkurve in θ . Die Menge aller [partiellen] Influenzkurven in θ heie $\Psi_2(\theta)$ resp. $\Psi_2^D(\theta)$.

Bemerkung 9.4.5 Wegen der Mebarkeit (a) ist (c) äquivalent mit

$$\sum_l p_l \mathbb{E}_\theta[\psi_\theta^{(l)} \Lambda_\theta^{\mathcal{U}_l \tau}] = \mathbb{I}_k \text{ resp. } D$$

Definition 9.4.6 *Sei $\tau : \Theta \rightarrow \mathbb{R}^p$ eine in $\theta \in \Theta$ differenzierbare Abbildung mit $D = d\tau|_\theta$, $\text{rk } D = p$. Ein asymptotischer Schätzer $S = (S_n)$ für $\tau(\theta)$, $S_n : (\tilde{\Omega}^n, \tilde{\mathcal{A}}^{\otimes n}) \rightarrow (\mathbb{R}^p, \mathbb{B}^p)$ heie asymptotisch linear in $Q_\theta \in \mathcal{P}^{\mathcal{U}}$, falls es eine partielle Influenzkurve $\psi_\theta \in \Psi_2^D(\theta)$ gibt, so da*

$$s_n^{-1}(S_n - \tau(\theta)) = s_n \sum_{i=1}^n \psi_\theta(v_i, y_i) + o_{Q_{n,\theta}}(n^0) \tag{9.24}$$

Die Begründung für Bedingung (c) in Definition 9.4.4 ist folgende:

Lemma 9.4.7 *Sei S ein asymptotischer Schätzer für $\tau(\theta)$ mit Entwicklung (24) für ein ψ_θ mit Eigenschaften (a), (b), aus Definition 9.4.4. Dann gilt (c) genau dann, wenn*

$$s_n^{-1}(S_n - \tau(\theta)) \circ Q_{n,\theta+s_n t_n} \xrightarrow{w} \mathcal{N}(Dt, \mathcal{C}_\theta(\psi_\theta)) \tag{9.25}$$

für $t_n \rightarrow t \in \mathbb{R}^k$, wobei

$$\mathcal{C}_\theta(\psi_\theta) = \sum_l p_l \text{Cov}_\theta[\psi_\theta^{(l)}]. \tag{9.26}$$

BEWEIS: Rieder (1994), Lemma 4.2.18. ////

Proposition 9.4.8 Sei $S = (S_n)$ ein in $Q_\theta \in \mathcal{P}^{\mathcal{U}}$ asymptotisch linearer Schätzer mit Influenzkurve $\psi_\theta \in \Psi_2^D(\theta)$. Dann gilt

(a)

$$\mathcal{C}_\theta(\psi_\theta^{\mathcal{U}}) = \sum_l p_l \text{Cov}_\theta[\psi_\theta^{(l)}] \succeq D[\sum_l p_l \mathcal{I}_\theta^{\mathcal{U}_l}]^{-1} D^\tau \quad (9.27)$$

mit “=” $\iff \psi_\theta^{\mathcal{U}} = \hat{\psi}_\theta := D\mathcal{I}_\theta^{\mathcal{U}}^{-1}\Lambda_\theta^{\mathcal{U}}$ (b) $s_n^{-1}(S_n - \tau(\theta))$ ist regulär für $D = d\tau|_\theta$ mit normaler Grenzverteilung

$$\mathcal{N}(0, \mathcal{C}_\theta(\psi_\theta)) = \mathcal{N}(0, \mathcal{C}_\theta(\psi_\theta) - \mathcal{C}_\theta(\hat{\psi}_\theta)) * \mathcal{N}(0, \mathcal{C}_\theta(\hat{\psi}_\theta)) \quad (9.28)$$

Die Grenzverteilung $\mathcal{N}(0, \mathcal{C}_\theta(\hat{\psi}_\theta))$ wird dann und nur dann angenommen, wenn $\psi_\theta = \hat{\psi}_\theta$.

(c) Sei die zugrundeliegende Verlustfunktion² $\ell \in \mathbf{L}$, ℓ stetig λ^p -f.s.. Dann ist

$$\begin{aligned} & \lim_{b \rightarrow \infty} \lim_{c \rightarrow \infty} \lim_{n \rightarrow \infty} \sup_{t \in \Theta, |t| \leq c} \int b \wedge \ell(s_n^{-1}[S_n - \tau(\theta + s_n t)]) dQ_{n, \theta + s_n t} = \\ & = \int \ell d\mathcal{N}(0, \mathcal{C}_\theta(\psi_\theta)) \geq \int \ell d\mathcal{N}(0, \mathcal{C}_\theta(\hat{\psi}_\theta)) \end{aligned} \quad (9.29)$$

und die untere Schranke wird angenommen, wenn $\psi_\theta = \hat{\psi}_\theta$. Ist $\ell \in \mathbf{L}$ monoton quadratisch, d.h.

$$\ell(z) = \begin{cases} v(z^\tau A z) & \text{für } |z| < \infty \\ \sup \ell & \text{für } |z| = \infty \end{cases} \quad (9.30)$$

für eine monotone Funktion $v : [0, \infty] \rightarrow [0, \infty]$ und eine positiv definite Matrix A , und nicht konstant, so kann die untere Schranke nur durch $\psi_\theta = \hat{\psi}_\theta$ angenommen werden.

BEWEIS: Rieder (1994), Proposition 4.2.20 ////

Noch allgemeiner gilt

Proposition 9.4.9 Wir betrachten das Problem, $\tau(\theta)$ in den Modellen $\tilde{Q}_n = \{Q_{n, \theta + s_n t} | t \in \mathbb{R}^k\} \subset \mathcal{M}_1(\mathcal{A}_n)$ zu schätzen.

(a) Die Standardisierung $R = (R_n)$ eines asymptotischen Schätzers $S = (S_n)$ für $\tau(\theta)$, $R_n = s_n^{-1}(S_n - \tau(\theta))$ sei regulär für $D = d\tau|_\theta$ mit Grenzverteilung $M \in \mathcal{M}_1(\mathbb{B}^k)$. Dann gibt es eine Wahrscheinlichkeit $M_0 \in \mathcal{M}_1(\mathbb{B}^k)$, so daß

$$M = M_0 * \mathcal{N}(0, \mathcal{C}_\theta(\hat{\psi}_\theta)) \quad (9.31)$$

Die Standardisierung $R^* = (R_n^*)$ eines regulären Schätzers S^* erreicht die Grenzverteilung $M^* = \mathcal{N}(0, \mathcal{C}_\theta(\hat{\psi}_\theta))$ dann und nur dann, wenn S^* ein asymptotisch linearer Schätzer mit Influenzkurve $\psi_\theta = \hat{\psi}_\theta$ ist.

²vgl. Definition A.5.7

(b) Sei $\ell \in L$,

$$\psi_0 = \int \ell dN(0, \mathcal{C}_\theta(\hat{\psi}_\theta)). \quad (9.32)$$

Dann gilt

$$\lim_{b \rightarrow \infty} \lim_{c \rightarrow \infty} \liminf_{n \rightarrow \infty} \inf_{S \in \mathcal{S}} \sup_{t \in \Theta, |t| \leq c} \int b \wedge \ell(s_n^{-1}[S_n - \tau(\theta + s_n t)]) dQ_{n, \theta + s_n t} \geq \psi_0 \quad (9.33)$$

Ist ℓ λ^p -f.s. stetig und der Schätzer S^* asymptotisch linear in P_θ , mit Influenzkurve $\hat{\psi}_\theta$, so gilt

$$\lim_{b \rightarrow \infty} \lim_{c \rightarrow \infty} \lim_{n \rightarrow \infty} \sup_{t \in \Theta, |t| \leq c} \int b \wedge \ell(s_n^{-1}[S_n - \tau(\theta + s_n t)]) dQ_{n, \theta + s_n t} = \psi_0 \quad (9.34)$$

BEWEIS: Rieder (1994), Proposition 4.2.19

////

9.5 EM-Algorithmus

Im vorigen Abschnitt haben wir gesehen, wie sich die Situation von unvollständigen Beobachtungen in den Kontext der allgemeinen asymptotischen Theorie einordnen läßt.

Dies kann man dazu verwenden, um zu beurteilen, wie gut ein Verfahren zur Parameterschätzung in dieser Situation [asymptotisch] ist, und um gegebenenfalls optimale Verfahren zu konstruieren.

Für den Fall unvollständiger Beobachtungen ist der EM-Algorithmus ein recht universell einsetzbares und oft einfach durchzuführendes Verfahren. Wir wollen nun aus der Warte der Asymptotik beurteilen, ob der EM-Algorithmus sich auch mit den theoretisch optimalen Verfahren messen kann, und untersuchen, ob ein entsprechend der asymptotischen Theorie optimales Verfahren aufwendiger zu realisieren ist oder nicht.

9.5.1 "klassischer" EM-Algorithmus

Wir beschränken uns hier darauf, den EM-Algorithmus aus Dempster et al. (1977) für unabhängig identisch verteilte Beobachtungen darzustellen. Auf die zahlreichen Verallgemeinerungen wie GEM-, ECM-, ECME- und OSL-Algorithmen gehen wir hier nicht ein, sondern verweisen auf die umfangreiche Literatur zu diesem Thema, vgl. die Monografien Little und Rubin (1987) und McLachlan und Krishnan (1997).

Zunächst gehen wir nicht auf die Regularitätsvoraussetzungen ein und verwenden die bisher eingeführte Notation anstelle derjenigen von Dempster et al. (1977).

Wir definieren $Q(\theta, \tilde{\theta})(v, y) := \sum_l I_{\{v=l\}} E_{\tilde{\theta}}[\log p_\theta(y) | \mathcal{U}_l]$ mit $p_\theta(x)$ der Dichte von P_θ . Dann lautet der EM-Algorithmus

Algorithmus 9.5.1

(-1) Vorbereitung:

Berechnen von $Q(\theta, \tilde{\theta})(v, y)$ als Funktionen in $(\theta, \tilde{\theta}, y, v)$.

(0) Initialisierung:

$$j := 0;$$

Bestimmen eines Startschätzers $\theta^{(j)}$ für θ .

(1) E-Schritt:

Berechnung des bedingten Erwartungswertes an $P_{\theta^{(j)}}$, $\sum_i Q(\theta, \theta^{(j)})(v_i, y_i)$ als Funktion in θ .

(2) M-Schritt:

Maximieren der bedingten Likelihood

$$\theta^{(j+1)} := \operatorname{argmax}_{\theta} \sum_i Q(\theta, \theta^{(j)})(y_i) \quad (9.35)$$

$j := j + 1$, gehe zu (1) bis Konvergenz.

grundlegende Eigenschaften

Der EM-Algorithmus besitzt folgende, einfach zu beweisende Eigenschaften, die man z.B. in McLachlan und Krishnan (1997) bewiesen findet:

- Wegen der Konkavität der \log -Funktion und der Jensen-Ungleichung für bedingte Erwartungswerte ist die Folge der bedingten Loglikelihoods, $L(\theta^{(j)}) := \sum_i \log p^{\mathcal{M}_i \theta^{(j)}}(y_i)$ monoton wachsend.
- Falls die Folge der $\theta^{(j)}$ gegen ein θ_0 konvergiert und $Q(s, t)$ in $t = \theta_0$, $s = \theta_0$ und $L(t)$ in $t = \theta_0$ differenzierbar sind, so muß notgedrungen $d/dt L(\theta_0) = 0$ sein.

Leider ist damit aber noch nicht viel gewonnen, denn

- es ist nicht klar, ob mit der Konvergenz der $L(\theta^{(j)})$ auch die Folge der $\theta^{(j)}$ mitkonvergiert – vgl. McLachlan und Krishnan (1997), Example 3.3.
- selbst wenn die Folge der $\theta^{(j)}$ konvergiert, ist nicht klar, ob es sich um ein lokales Maximum, Minimum oder einen Sattelpunkt handelt – vgl. McLachlan und Krishnan (1997), Example 3.1. und Example 3.2.
- selbst wenn die Folge der $\theta^{(j)}$ gegen ein lokales Maximum konvergiert, ist nicht sicher, ob es sich um das globale Maximum handelt.
- selbst wenn die Folge der $\theta^{(j)}$ gegen das globale Maximum konvergiert, ist nicht sicher, ob dieses ein konsistenter Schätzer für den Parameter ist – vgl. Le Cam (1986), §17.6, pp. 621–625.
- selbst wenn die Folge der $\theta^{(j)}$ ein konsistenter Schätzer für den Parameter im idealen Modell ist, muß dies unter kleinsten Modellabweichungen nicht mehr gelten — der Grund für die Entstehung der robusten Statistik.

Zur Konvergenz des Algorithmus hat Wu (1983) folgendes gezeigt:

Theorem 9.5.2 Sei $\theta^{(j)}$ eine Folge, die durch den EM-Algorithmus erzeugt wird. Gilt dann mit $S := \{\theta \in \Theta^o \mid \theta \text{ ist stationärer Punkt von } L\}$

$$M(\theta^{(j)}) \cap S^c \text{ ist abgeschlossen,} \quad (9.36)$$

so sind alle Häufungspunkte von $\{\theta^{(j)}\}$ stationäre Punkte von L und $L(\theta^{(j)}) \uparrow L(\theta_0)$ für einen stationären Punkt θ_0 .

Eigenschaft (36) ist erfüllt, falls Q stetig in (s, t) ist.

Allerdings beschäftigen sich alle dem Autor bekannten Arbeiten über die Konvergenz des EM-Algorithmus' nur mit der Frage, ob die Folge der $\theta^{(j)}$ gegen das globale Maximum der bedingten Loglikelihood konvergiert und gegebenenfalls, wie schnell sie dies tut. Die Fragestellung, welche statistischen Eigenschaften diese Schätzfolge hat, wird weitestgehend ausgeklammert.

9.5.2 statistische Eigenschaften des klassischen EM-Algorithmus'

Wir gehen von nun an bereits von der Konvergenz der $\theta^{(j)}$ gegen die globale Maximalstelle der Funktion L aus. Damit müssen wir uns fragen, unter welchen Umständen der Maximum-Likelihood-Schätzer konsistent ist, und wenn er dies ist, wie er sich auch auf Umgebungen des idealen Modells verhält.

Daß sich der ML-Schätzer in vielen Situationen als optimal im Sinn der Cramér-Rao-Schranke herausstellt, verdankt er weniger der Eigenschaft, Maximalstelle der Likelihood oder Nullstelle deren Ableitung zu sein, als vielmehr der Tatsache, daß unter Regularitätsvoraussetzungen des zugrundeliegenden Modells diese Nullstelle eine Entwicklung der Gestalt

$$\sqrt{n}(S_n - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\theta(x_i) + o_P(n^0)$$

besitzt, mit $\psi_\theta = \mathcal{I}_\theta^{-1} \Lambda_\theta$ der klassisch optimalen Influenzkurve; in diesem Fall handelt es sich nach Definition 9.4.6 also um einen asymptotisch linearen Schätzer, der dann aber die im Sinne der Cramér-Rao-Ungleichung optimale Varianz annimmt. Genauer gilt

Theorem 9.5.3 *Gilt für das Modell $\mathcal{P} = \{P_{\theta,n}; \theta \in \Theta\}$*

(a) *Das Modell \mathcal{P} besitzt die LAMN-Eigenschaft in $\theta = \theta_0$ mit asymptotisch suffizienter Statistik W_n und Standardisierung T_n .*

(b) *Die Folge $s_n^{-1}(\theta^{\text{ML}} - \theta_0)$ ist straff unter $P_{\theta,n}$*

(c) *Für jedes $\varepsilon > 0$ und $\alpha > 0$ gilt*

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P_{\theta_0} \left[\sup_{|h_1 - h_2| \leq \delta} \left| \log \frac{dP_{\theta_2, n}}{dP_{\theta_0, n}} - \log \frac{dP_{\theta_1, n}}{dP_{\theta_0, n}} \right| > \varepsilon, h_i \in D_\alpha \right] = 0, \quad (9.37)$$

wobei wir $D_\alpha := \{h \in \mathbb{R}^k : |h| \leq \alpha\}$ setzen und abkürzend schreiben $\theta_i := \theta_0 + s_n h_i$, so haben wir

$$s_n^{-1}(\theta^{\text{ML}} - \theta_0) = T_n^{-\frac{1}{2}}(\theta_0) W_n(\theta_0) + o_{P_{\theta_0, n}}(n^0) \quad (9.38)$$

Ist insbesondere das Modell auch noch L_2 -differenzierbar mit Standardisierungsfolge $s_n = \frac{1}{\sqrt{n}}$ und besitzt sogar die LAN-Eigenschaft, so gilt

$$\sqrt{n}(\theta^{\text{ML}} - \theta_0) = \frac{1}{\sqrt{n}} \mathcal{I}_\theta^{-1} \sum_{i=1}^n \Lambda_{n,i} + o_{P_{\theta_0, n}}(n^0) \quad (9.39)$$

BEWEIS: Jeganathan (1982), Theorem 7.

////

Bemerkung 9.5.4 • Andere Bedingungen für die Konsistenz des ML-Schätzers im unabhängig identisch verteilten Fall lassen sich aus Huber (1967), Le Cam (1970), sowie Wald (1949) und Wolfowitz (1949) ablesen.

- Der Rückzug auf L_2 -differenzierbare und damit asymptotisch normale Modelle schließt Fälle zunächst aus, in denen das ML-Verfahren mit noch besserer Konvergenzrate als $1/\sqrt{n}$ konsistent ist, wie z.B. die $ufo(0, \theta)$ -Familie.

Verläßt man aber das ideale Modell, und betrachtet die Schätzfolge auf Umgebungssystemen, so gilt für diese wieder ein (nichtparametrischer) asymptotischer Minimaxsatz mit asymptotisch normalverteilten Verfahren als Optimum längs ungünstigsten "Richtungen" innerhalb der Umgebungen — vgl. Remark 4.3.3 "Smooth Parametric Model" in Rieder (1994).

Auf die Stabilitätseigenschaften des ML-Schätzers kommen wir später noch zu sprechen.

9.5.3 "One-Step" EM-Algorithmus

asymptotische Eigenschaften des ML-Prinzips

Als Einstieg soll anhand von Diagramm 9.1 skizziert werden, wie man die statistischen Eigenschaften des ML-Prinzips herleitet. Ziel ist die effiziente Schätzung

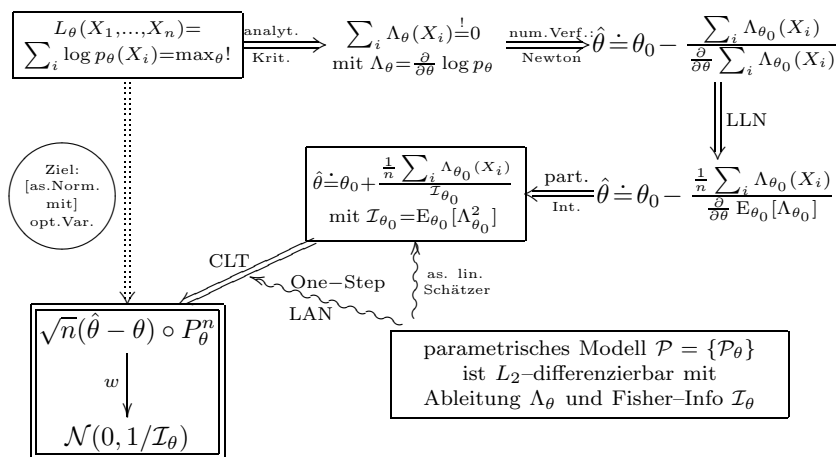


Abbildung 9.1: schem. Darstellung der Herleitung der as. Normalität des MLE

eines Parameters, also eine möglichst kleine Varianz des Verfahrens möglichst ohne Bias. Dieses Problem ist im allgemeinen nur asymptotisch in den Griff zu bekommen, und argumentiert man mehrdimensional, bietet sich im wesentlichen die dargestellte Argumentation an, die wir hier der Einfachheit halber nur eindimensional ausführen:

Wir starten mit der Likelihood des Modells, diese ist zu maximieren. Wir gehen dazu zum Logarithmus über, leiten nach dem Parameter ab, suchen eine Nullstelle, was im allgemeinen nur numerisch geht — zum Beispiel mit dem

Newton-Verfahren — und gelangen zum Gauss-Newton-Verfahren, argumentieren dann mit dem Gesetz der großen Zahlen im Nenner und erhalten das Fisher-Scoring, integrieren den Nenner partiell und erhalten eine asymptotische Entwicklung in einer Summe stochastisch unabhängiger Inkremente, von wo aus wir mit Lindeberg-Lévy auf die Normalverteilung schließen können.

Für eine Übersicht zu numerischen Optimierungsverfahren, die zu diesem Zweck in der Statistik verwendet werden, verweisen wir auf Berndt et al. (1974), und im Spezialfall des Zustandsraummodells auf Gupta und Mehra (1974).

Bei all diesen Schlüssen benötigen wir massiv die Glattheit der Likelihood. Andererseits kann diese Glattheit der Likelihood nicht wirklich notwendig sein, da ja diese als Dichte nur fast sicher bestimmt ist.

Einen anderen, direkteren Zugang, bietet die Bedingung an das Modell, L_2 -differenzierbar zu sein. Diese liefert dann sofort die Existenz der asymptotisch linearen Entwicklung und zugleich über die dann geltende LAN-Eigenschaft sogar eine gewisse Gleichmäßigkeit der asymptotischen Normalität.

Eine weitere, vom ML-Prinzip unabhängige Konstruktion, die des One-Steps, ermöglicht es nun, ohne weitere Voraussetzungen einen Schätzer mit der gewünschten asymptotischen Entwicklung zu erzielen:

Idee des “One-Step”

Daß man die optimale Entwicklung aus Proposition 9.4.8 auch anders erreichen kann, als mit einem M-Schätzer legt folgendes heuristisches Argument nahe: Wir definieren zu einem Startschätzer θ^0 mit $\sqrt{n}(\theta^0 - \theta) = O_P(n^0)$ und einer Influenzkurve ψ_θ den *One-Step-Schätzer*

$$\hat{\theta} := \theta^0 + \frac{1}{n} \sum_{i=1}^n \psi_{\theta^0}(x_i) \quad (9.40)$$

Dann “gilt”

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\theta(x_i) - \sqrt{n}(\theta^0 - \theta) &= \frac{1}{\sqrt{n}} \left(\sum_{i=1}^n \psi_{\theta^0}(x_i) - \psi_\theta(x_i) \right) \doteq \\ &\doteq \sqrt{n} \int \psi_{\theta^0} - \psi_\theta dP_\theta = -\sqrt{n} \int \psi_{\theta^0} (dP_{\theta^0} - dP_\theta) \doteq \\ &\doteq -\sqrt{n} \int \psi_{\theta^0} \Lambda_{\theta^0}^T(\theta^0 - \theta) dP_{\theta^0} = -\sqrt{n}(\theta^0 - \theta), \end{aligned}$$

also $\sqrt{n}(\hat{\theta} - \theta) \doteq \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\theta(x_i)$.

Genauer läßt sich nach diesem Prinzip bei Vorliegen von L_2 -Differenzierbarkeit ein Schätzer mit der gewünschten Entwicklung sogar auf Umgebungssystemen $\mathcal{U}(\theta)$,

$$\mathcal{U}(\theta) = \{U(\theta, r) \mid r \in [0, \infty)\} \quad (9.41)$$

und $U(\theta, r)$ eine “Umgebung” mit Radius r um P_θ , gewinnen:

Man nehme irgendeinen Schätzer θ^0 , der nur Werte in Θ annimmt und \sqrt{n} -konsistent auf $\mathcal{U}(\theta)$ ist, so daß für alle $r \in (0, \infty)$

$$\lim_{M \rightarrow \infty} \limsup_n \sup \{Q_n^{(n)}(\sqrt{n}|\hat{\theta}_n - \theta| > M \mid Q_{n,i} \in U(\theta, r/\sqrt{n})\} = 0. \quad (9.42)$$

Diesen verwende man als Startschätzer und definiere damit — notfalls mit einer noch einmal regularisierten Version — den One-Step-Schätzer (40). Mathematisch abgesichert findet sich dieses Argument in Rieder (1994), Kapitel 6.

Argumente für einen “One-Step”-EM-Algorithmus

Nach diesem Prinzip verfahren wir nun bei der Konstruktion eines “One-Step”-EM-Algorithmus’. Dabei haben wir folgende Argumente

- Für den Fall der klassischen Scores-Funktion $\hat{\psi}_\theta := \mathcal{I}_\theta^{-1} \Lambda_\theta$ kommt man dabei mit der L_2 -Differenzierbarkeit ohne weitere Voraussetzungen aus: Als Startschätzer kann man stets Minimum-Distanzschätzer zur Kolmogoroff- oder zur Cramér-von Mises-Metrik verwenden — vgl. Rieder (1994), Theoreme 6.3.7, 6.3.8 —, und zu diesen gibt es stets L_2 -Approximationen $\hat{\psi}_{n,\theta}$ für $\hat{\psi}_\theta$, deren Suprema von der Größenordnung $o(\sqrt[4]{n})$ sind, so daß die entsprechende One-step-Entwicklung gilt — vgl. Rieder (1994), Lemma 6.4.4.
- Unter Regularitätsannahmen — nämlich daß

$$-\frac{1}{n} \sum \dot{\psi}_\theta(x) \doteq - \int \dot{\psi}_\theta dP_\theta \stackrel{!}{=} - \int \psi_\theta \Lambda_\theta^\tau dP_\theta = \mathbb{I},$$

fällt diese Konstruktion mit dem ersten Schritt eines Newton-Raphson-Verfahrens zusammen.

- Da wir aber in Theorem 9.3.2 gezeigt haben, daß sich die L_2 -Differenzierbarkeit auf gröbere σ -Algebren überträgt, gilt dies auch für das Modell $\mathcal{P}^{\mathcal{U}}$ aus Abschnitt 9.4.3, was die Konsistenz des Verfahrens zeigt.

Definition “One-Step”-EM-Algorithmus

Wir schlagen also folgenden Algorithmus vor:

Algorithmus 9.5.5

(-1) Vorbereitung:

Berechnen von Λ_θ und $E_\theta[\Lambda_\theta|\mathcal{U}_i]$ sowie $\mathcal{I}_\theta^{\mathcal{U}_i}$, und damit $\mathcal{I}_\theta = \sum_{j=1}^k p_j \mathcal{I}_\theta^{\mathcal{U}_j}$ sowie \mathcal{I}_θ^{-1} als Funktionen in (θ, x) bzw. θ .

(0) Initialisierung:

Bestimmen eines \sqrt{n} -konsistenten Startschätzers θ^0 für θ .

(1) E-Schritt:

Werte-Belegung von $E_\theta[\Lambda_\theta(x_i)|\mathcal{U}_j]$.

(2) M-Schritt:

Auswertung des One-Steps

$$\hat{\theta} := \theta^0 + \frac{1}{n} \mathcal{I}_{\theta^0}^{-1} \sum_{i=1}^n \sum_{j=1}^k I_{V_i=j} E[\Lambda_{\theta^0}(x_i)|\mathcal{U}_j], \quad (9.43)$$

Im Fall, daß man nicht am Parameter selbst, sondern an einer möglicherweise nicht injektiven, differenzierbaren Parametertransformation $\tau(\theta)$ interessiert ist, führt dies auf folgende Modifikation von (43)

$$\hat{\theta} := \theta^0 + \frac{1}{n} D_{\theta^0} \mathcal{I}_{\theta^0}^{-1} \sum_{i=1}^n \sum_{j=1}^k \mathbf{I}_{V_i=j} E[\Lambda_{\theta^0}(x_i) | \mathcal{U}_j], \quad (9.44)$$

wobei D_{θ} die Jakobische von τ in θ ist. Ein wichtiger Spezialfall ist das Schätzen eines Haupt-Parameters bei Vorliegen eines Nebenparameters; dann ist $\theta = (\theta^{(1)\tau}, \theta^{(2)\tau})\tau$, mit $\theta^{(1)}$ dem Haupt- und $\theta^{(2)}$ dem Nebenparameter und $\tau(\theta) = \theta^{(1)}$ — vgl. Beispiel 9.6.4.

Bemerkung 9.5.6 Ein “One-Step”-Verfahren bei unvollständiger Information wurde bereits im Spezialfall der zufälligen Zensierung von Yang (1996) vorgeschlagen — allerdings ohne auf den Zusammenhang zum EM-Algorithmus hinzuweisen. Diese Arbeit benützt beim Übergang vom vollständig beobachteten Modell zum unvollständig beobachteten die Resultate von Le Cam und Yang (1988).

9.5.4 “klassischer” vs. “One-Step” EM-Algorithmus

Im direkten Vergleich der beiden Verfahren stellen wir fest

- Die Schwierigkeit, den bedingten Erwartungswert auszuwerten, liegt in beiden Verfahren vor; aufgrund der iterativen Struktur im klassischen Fall muß man ihn dort aber mehrfach auswerten.
- Auf der “One-Step”-Seite schlägt zusätzlich die Berechnung der Fisher-Information \mathcal{I} zu Buche, auf die man beim klassischen Verfahren verzichten kann.
- Die numerischen Schwierigkeiten, die unter Umständen im klassischen Verfahren bei der Maximierung der bedingten Likelihood auftreten, werden im One-Step-Verfahren vermieden.
- In den Beispielen 9.6.3 und 9.6.4 erzielen beide Verfahren die optimale Limes-Varianz.

Unter lokaler Identifizierbarkeit des Modells³ liefert im L_2 -differenzierbaren Rahmen das “One-Step” Verfahren in jedem Fall einen konsistenten, asymptotisch effizienten Schätzer, während wir für die Konsistenz und die Optimalität des bedingten ML-Schätzers zusätzliche Voraussetzungen wie (37) benötigen.

9.5.5 robuster EM-Algorithmus

Eine generelle Eigenschaft der ML-Schätzung ist es, daß, wann immer die Scorefunktion nicht beschränkt ist, eine einzelne Beobachtung unbeschränkten Einfluß auf die Parameterschätzung hat; das Verfahren ist nicht stetig in der schwachen Topologie!

Ist man sich des Modells nicht vollkommen sicher, sollten daher robuste Schätzer

³also $\det \mathcal{I} > 0$

verwendet werden, bei denen solche Effekte nicht auftreten.

Hat man einmal das Konzept asymptotisch linearer Schätzer zur Verfügung, können wir darin auch robuste Optimierungsprobleme wie in Kapitel 5 bzw. 7 in Rieder (1994) stellen.

Betrachtet man die Influenzkurven, wie sie in Definition 9.4.4 eingeführt sind, so stellt man eine formale Analogie zu den Regressions-Influenzkurven in Kapitel 7 fest, in dem Sinn, daß sich die Auswahl-Indikator-Variablen $I_{\{V=i\}}$ als Regressor und der Rest als transformiertes Residuum lesen läßt.

Es stellen sich wie im Regressionskontext folgende Problemansätze

- “Fehlerfreie-Variablen” oder bedingte Umgebungen
- “Fehler-in-den-Variablen” oder unbedingte Umgebungen

Erstere halten die Randverteilung von V fest, also unterliegen die Anteile der unvollständigen Beobachtungen keiner Kontamination; bei letzteren ist es auch zugelassen, daß die Multinomialverteilung der V gestört wird.

Wollen wir nun optimale Influenzkurven bestimmen, in dem Sinn, daß die Spur der Kovarianz minimal ist unter einer Schranke an den Bias, so entstehen für beide Ansätze unterschiedliche Optimierungsprobleme; während erstere analog zu denen der Regression — vgl. Kapitel 7.4 in Rieder (1994) — zu bearbeiten sind, fallen letztere in den allgemeinen Kontext des (mehrdimensionalen) Parameter-Schätzproblems — vgl. Kapitel 5.5 in Rieder (1994). Auf die Details wollen wir aber hier nicht eingehen, da es schlicht darum geht, unter den Identifikationen $X \rightsquigarrow V$, $Y \rightsquigarrow Y$, $K \rightsquigarrow \text{Mult}(m, \underline{p})$, $Q(du|x) = \sum_i I_{\{V=i\}} P_\theta^{\mathcal{U}_i}$ die Resultate aus Rieder (1994) zu spezialisieren.

Exemplarisch geben wir hier nur die Gestalt der Lösung für unbedingte Umgebungen zu folgendem Problem an:

$$\text{tr Cov}_\theta(\psi_\theta) = \min! \quad \psi \in \Psi_2^D(\theta), \quad \|\psi\|_\infty \leq b \quad (9.45)$$

Dazu lassen wir die fixierte Parameterstelle θ in den Indizes weg, und erhalten aus Theorem 5.5.1 in Rieder (1994) als Gestalt der Lösung zu (45) für $b > b_{\min}$ ⁴

$$\hat{\psi} = (A\Lambda - a) \min\left\{1, \frac{b}{|A\Lambda - a|}\right\} \quad (9.46)$$

mit geeigneten Lagrange-Multiplikatoren A und a , die für Eigenschaften (b) und (c) in Definition 9.4.4 sorgen; genauer gilt für $\dim \theta = p$, $D = d\tau \in \mathbb{R}^{p' \times p}$, $\text{rk } D = p' \leq p$, $A \in \mathbb{R}^{p' \times p}$, $A = D\tilde{A}$ für ein $\tilde{A} = \tilde{A}^\tau \succ 0$ und $a = (a_1, \dots, a_k) \in \mathbb{R}^{p' \times k}$, $a = A\zeta$ für ein $\zeta \in \mathbb{R}^{p \times k}$ und

$$\hat{\psi} = \hat{\psi}(v, y) = \sum_{j=1}^m I_{\{V=j\}} (A E[\Lambda | \mathcal{U}_j] - a_j) \min\left\{1, \frac{b}{|A E[\Lambda | \mathcal{U}_j] - a_j|}\right\} \quad (9.47)$$

Hierbei ist vor allem das Bestimmen der Lagrangemultiplikatoren, die eindeutig sind, falls $\mathcal{L}(\Lambda)$ eine Lebesgue-dichte besitzt⁵, ein numerisch nicht triviales Problem — vgl. Anhang D.

⁴ $b_{\min} := \min\{b \mid \exists \psi \in \Psi_2^D : \|\psi\|_\infty \leq b\}$

⁵ vgl. Beweis zu Theorem 5.5.9 in Rieder (1994)

Bemerkung 9.5.7 Ein anderer Robustheitsbegriff liegt der Arbeit von Yang (1990) über die effizient-robuste Schätzung im Modell mit zufälliger Zensierung zugrunde, einem Spezialfall der hier behandelten Situation: Anstatt Varianz und Bias simultan zu betrachten, geht es hier darum, das Schätzerrisiko — übertragen auf das Schätzproblem z.B. die Spur der Kovarianz — auf Umgebungen zu minimieren, ohne dabei die Oszillation des Funktionals — dem Bias im Schätzerkontext — parallel mitzubetrachten.

Diese Arbeit stützt sich wiederum auf Arbeiten von Beran (1981), Beran (1982) und benutzt den Beweis von Le Cam und Yang (1988) für Theorem 9.3.2.

Bemerkung 9.5.8 Ist nur ein \sqrt{n} -konsistenter, robuster Schätzer gesucht, wie es für die Schätzung von Nebenparametern benötigt wird, so erreichen wir sehr einfach folgende, suboptimale, beschränkte Influenzkurve, mithilfe der man ein “One-Step”-Verfahren definieren kann:

$$\tilde{\psi}(v, y) := D\bar{A} \sum_{j=1}^m I_{\{V=j\}} \left(E[\Lambda|\mathcal{U}_j] \min\left\{1, \frac{b}{|E[\Lambda|\mathcal{U}_j]|}\right\} - a_j \right) \quad (9.48)$$

mit

$$\begin{aligned} a_j &= E \left(E[\Lambda|\mathcal{U}_j] \min\left\{1, \frac{b}{|E[\Lambda|\mathcal{U}_j]|}\right\} \right) \\ \bar{A}^{-1} &= E \left(E[\Lambda|\mathcal{U}_j] E[\Lambda|\mathcal{U}_j]^\tau \min\left\{1, \frac{b}{|E[\Lambda|\mathcal{U}_j]|}\right\} \right). \end{aligned}$$

Ein solches $\tilde{\psi}$ werden wir zur Schätzung der Hyper-Parameter im Zustandsraum verwenden.

9.6 Beispiel: bivariates, normales Skalenmodell

Wir geben in diesem Abschnitt Beispiele dafür, wie sich die bisherigen Ergebnisse bei der Schätzung von Parametern beim Vorliegen von Missings umsetzen.

9.6.1 Modell

Seien für $i = 1, \dots, n$

$$X_i \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}_2(0, \Sigma) = \mathcal{N}_2\left(0, \begin{pmatrix} \sigma_1 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2 \end{pmatrix}\right).$$

Offenbar erfüllt die bivariate Normalverteilung die Endlichkeitsvoraussetzung⁶

$$\infty > 2 = \bar{\mathcal{J}}_{\mathcal{N}_2} := \sup \frac{\| \int x \nabla \varphi(x)^\tau d\mathcal{N}_2(dx) \|^2}{\int \varphi^2 d\mathcal{N}_2}, \quad 0 \neq \varphi \in \mathcal{D}_2, \quad [\lambda^2]$$

so daß das Modell L_2 -differenzierbar ist.

Bemerkung 9.6.1 Dieses Modell wird in McLachlan und Krishnan (1997), Ex. 2.1 und Little und Rubin (1987) Ex. 5.15, Ex. 6.3 studiert, allerdings ohne den formalen Nachweis der L_2 -Differenzierbarkeit.

⁶vgl. Theorem B.2.4 und Lemma B.3.2, Lemma B.3.3

Für die späteren Untersuchungen führen wir an dieser Stelle die später noch benötigten Terme ein:

Es gilt mit $d_s = \det \Sigma$, $G := d_s x^\top \Sigma^{-1} x = x_1^2 \sigma_2 + x_2^2 \sigma_1 - 2\sigma_{1,2} x_1 x_2$, wobei wir die Erwartungswerte immer unter P_Σ auswerten:

$$\begin{aligned} \Lambda_\Sigma^{(1)} &:= \Lambda_\Sigma = \begin{pmatrix} \Lambda_{\sigma_1} \\ \Lambda_{\sigma_{1,2}} \\ \Lambda_{\sigma_2} \end{pmatrix} = \frac{1}{2d_s^2} \begin{pmatrix} [\sigma_2(G - d_s) - d_s x_2^2] \\ -2[\sigma_{1,2}(G - d_s) - d_s x_1 x_2] \\ [\sigma_1(G - d_s) - d_s x_1^2] \end{pmatrix} \\ \mathcal{I}_\Sigma^{(1)} &:= E \Lambda_\Sigma^{(1)} \Lambda_\Sigma^{(1)\top} = \frac{1}{2d_s^2} \begin{pmatrix} \sigma_2^2 & -2\sigma_2 \sigma_{1,2} & \sigma_{1,2}^2 \\ -2\sigma_2 \sigma_{1,2} & 2\sigma_1 \sigma_2 + 2\sigma_{1,2}^2 & -2\sigma_1 \sigma_{1,2} \\ \sigma_{1,2}^2 & -2\sigma_1 \sigma_{1,2} & \sigma_1^2 \end{pmatrix} \\ (\mathcal{I}_\Sigma^{(1)})^{-1} &= \begin{pmatrix} 2\sigma_1^2 & 2\sigma_1 \sigma_{1,2} & 2\sigma_{1,2}^2 \\ 2\sigma_1 \sigma_{1,2} & \sigma_1 \sigma_2 + \sigma_{1,2}^2 & 2\sigma_2 \sigma_{1,2} \\ 2\sigma_{1,2}^2 & 2\sigma_2 \sigma_{1,2} & 2\sigma_2^2 \end{pmatrix} \\ \Lambda_\Sigma^{(2)} &:= E \Lambda_\Sigma | X_1 = \begin{pmatrix} E \Lambda_{\sigma_1} | X_1 \\ E \Lambda_{\sigma_{1,2}} | X_1 \\ E \Lambda_{\sigma_2} | X_1 \end{pmatrix} = \frac{1}{2\sigma_1^2} \begin{pmatrix} [X_1^2 - \sigma_1] \\ 0 \\ 0 \end{pmatrix} \\ \mathcal{I}_\Sigma^{(2)} &:= E \Lambda_\Sigma^{(2)} \Lambda_\Sigma^{(2)\top} = \begin{pmatrix} \frac{1}{2\sigma_1^2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ \mathcal{I}_\Sigma^Y &= (1 - \alpha) \mathcal{I}_\Sigma^{(1)} + \alpha \mathcal{I}_\Sigma^{(2)} \end{aligned}$$

Um später die Inverse $(\mathcal{I}_\Sigma^Y)^{-1}$ ausrechnen zu können, notieren wir ohne Beweis

Lemma 9.6.2 (a) Sei $A \in \text{GL}(n)$, $B \in \mathbb{R}^{n \times n}$ so daß $(\mathbb{I}_n + A^{-1}B) \in \text{GL}(n)$; dann ist

$$(A + B)^{-1} = (\mathbb{I}_n + A^{-1}B)^{-1} A^{-1}.$$

(b) Sei $A \in \text{GL}(n)$,

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} A^{(1,1)} & A^{(1,2)} \\ A^{(2,1)} & A^{(2,2)} \end{pmatrix},$$

mit $A_{1,1}, A^{(1,1)} \in \mathbb{R}^{k \times k}$,

$$B \in \mathbb{R}^{n \times n}, \quad B = \begin{pmatrix} B_{1,1} & 0 \\ 0 & 0 \end{pmatrix}, \quad B_{1,1}^{-1} = B^{(1,1)}, \quad \text{mit } B_{1,1} \in \mathbb{R}^{k \times k};$$

dann ist

$$A^{-1}B = \begin{pmatrix} A^{(1,1)} B_{1,1} & 0 \\ A^{(2,1)} B_{1,1} & 0 \end{pmatrix}$$

(c) Seien A, B wie in (b) und sei zusätzlich $C := (\mathbb{I}_k + A^{(1,1)} B_{1,1}) \in \text{GL}(k)$; dann ist

$$(A + B)^{-1} = \begin{pmatrix} C^{-1} & 0 \\ -A^{(2,1)} B_{1,1} C^{-1} \mathbb{I}_{n-k} & \end{pmatrix} A^{-1} = A^{-1} - \begin{pmatrix} \mathbb{I}_k - C^{-1} & 0 \\ A^{(2,1)} B_{1,1} C^{-1} \mathbb{I}_0 & \end{pmatrix} A^{-1}$$

Beispiel 9.6.3 Interessierender Parameter ist $\sigma_{1,2} \in \mathbb{R}$, σ_1, σ_2 seien bekannt. Wir können aber nur einen Anteil $0 < [1 - \alpha]n < n$ der X_i , "vollständig"

beobachten. Genauer entscheidet ein von den X_i unabhängiger Zufallsmechanismus V_i , den wir ebenfalls beobachten, ob wir "vollständige" Beobachtungen durchführen oder nicht; die Wahrscheinlichkeit für eine derartige Realisation von V_i sei $1 - \alpha$. Bei den restlichen X_i steht uns jeweils nur die erste Komponente $X_{i,1}$ als Information zur Verfügung.

Es liegt die Situation des Abschnitts 9.4 vor:

Das so entstehende Modell ist L_2 -differenzierbar nach Proposition 9.4.3, mit $\mathcal{U} = (\sigma(X_{\cdot,1}, X_{\cdot,2}), \sigma(X_{\cdot,1}))^\tau$

$$\Lambda_{\sigma_{1,2}}^{\mathcal{U}} = \begin{pmatrix} \Lambda_{\sigma_{1,2}} \\ \mathbb{E}[\Lambda_{\sigma_{1,2}} | X_1] \end{pmatrix} = \begin{pmatrix} \Lambda_{\sigma_{1,2}} \\ 0 \end{pmatrix}$$

Die Beobachtungen mit fehlender zweiter Koordinate tragen also asymptotisch keine Information zu $\sigma_{1,2}$ bei, und die Fisher-Informationsschranke ist

$$(\mathcal{I}_{\sigma_{1,2}}^Y)^{-1} := \frac{1}{1 - \alpha} \frac{d_s^2}{\sigma_1 \sigma_2 + \sigma_{1,2}^2}.$$

Als unmittelbare Folge ergibt sich eine Optimalitätsaussage für asymptotisch lineare Schätzer:

Ist S_n ein asymptotisch linearer Schätzer in $(V_1, X_1), \dots, (V_n, X_n)$, so erreicht S_n die optimale Varianz im as. Minimax-Theorem A.5.9 gegeben durch die Informationsschranke $(\mathcal{I}_\theta^Y)^{-1}$ dann und nur dann, wenn

$$\psi_{\sigma_{1,2}}^{(1)} = ((\mathcal{I}_\Sigma^{(1)})_{2,2})^{-1} \Lambda_{\sigma_{1,2}} = \frac{\sigma_1 \sigma_2 - \sigma_{1,2}^2}{\sigma_1 \sigma_2 + \sigma_{1,2}^2} [X_1 X_2 + \sigma_{1,2} (1 - X^\tau \Sigma^{-1} X)]$$

und $\psi_{\sigma_{1,2}}^{(2)} = 0$.

Beispiel 9.6.4 Wir betrachten die gleiche Situation wie im vorangegangenen Beispiel, nur seien nun σ_1, σ_2 unbekannt. Wir verwenden Lemma 9.6.2, um zu zeigen

$$(\mathcal{I}_\Sigma^Y)^{-1} = \frac{1}{1 - \alpha} (\mathcal{I}_\Sigma^{(1)})^{-1} - \frac{2\alpha}{1 - \alpha} \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_{1,2} & \sigma_{1,2}^2 \\ \sigma_1 \sigma_{1,2} & \sigma_{1,2}^2 & \sigma_{1,2}^3 / \sigma_1 \\ \sigma_{1,2}^2 & \sigma_{1,2}^3 / \sigma_1 & \sigma_{1,2}^4 / \sigma_1^2 \end{pmatrix}.$$

Die Optimalitätsaussage für asymptotisch lineare Schätzer lautet nun:

Ist S_n ein asymptotisch linearer Schätzer in $(V_1, X_1), \dots, (V_n, X_n)$, so erreicht S_n die optimale Varianz im as. Minimax-Theorem A.5.9 gegeben durch die Informationsschranke

$$D(\mathcal{I}_\Sigma^Y)^{-1} D^\tau = \frac{1}{1 - \alpha} (\sigma_1 \sigma_2 + \sigma_{1,2}^2) - \frac{2\alpha}{1 - \alpha} \sigma_{1,2}^2 \quad \text{mit } D = (0, 1, 0)$$

dann und nur dann, wenn

$$\psi_{\sigma_{1,2}}^{(1)} = D(\mathcal{I}_\Sigma^Y)^{-1} \Lambda_\Sigma = \frac{1}{1 - \alpha} (X_1 X_2 - \sigma_{1,2}) - \frac{\alpha}{1 - \alpha} \sigma_{1,2} (X_1^2 / \sigma_1 - 1)$$

und $\psi_{\sigma_{1,2}}^{(2)} = \sigma_{1,2} (X_1^2 / \sigma_1 - 1)$.

Der Unterschied zum vorangegangenen Beispiel rührt daher, daß wir nun σ_i mit

\sqrt{n} -konsistenten Schätzern $\hat{\sigma}_i$ schätzen müssen und das Modell nicht adaptiv ist.

Wir überzeugen uns hier noch, daß stets $D(\mathcal{I}_\Sigma^Y)^{-1}D^\tau > \mathcal{I}_{\sigma_{1,2}}^{-1}$:

$$\begin{aligned} \frac{1}{1-\alpha} \frac{(\sigma_1\sigma_2 - \sigma_{1,2}^2)}{\sigma_1\sigma_2 - \sigma_{1,2}^2} &\stackrel{!}{\leq} \frac{1}{1-\alpha}(\sigma_1\sigma_2 - \sigma_{1,2}^2) - \frac{2\alpha}{1-\alpha}\sigma_{1,2}^2 \\ \Leftrightarrow 2\alpha\sigma_{1,2}^2(\sigma_1\sigma_2 + \sigma_{1,2}^2) &\stackrel{!}{\leq} (\sigma_1\sigma_2 + \sigma_{1,2}^2) - (\sigma_1\sigma_2 - \sigma_{1,2}^2) \\ \Leftrightarrow 2\alpha\sigma_{1,2}^2(\sigma_1\sigma_2 + \sigma_{1,2}^2) &\stackrel{!}{\leq} 4\sigma_{1,2}^2\sigma_1\sigma_2 \\ \Leftrightarrow \alpha(\sigma_1\sigma_2 + \sigma_{1,2}^2) &\stackrel{!}{\leq} 2\sigma_1\sigma_2 \end{aligned}$$

9.6.2 klassischer EM-Algorithmus für Beispiele 9.6.3 und 9.6.4

Verwendet man in den ersten beiden Beispielen den EM-Algorithmus zur Schätzung von $\sigma_{1,2}$, so ergibt sich folgendes Verfahren,

Algorithmus 9.6.5

(1) Initialisierung:

$$\begin{aligned} j &:= 0 \\ \hat{\sigma}_{1,2}^{(j)} &:= \frac{1}{\sum_i I_{\{V_i=1\}}} \sum_{i=1}^n I_{\{V_i=1\}} X_{1,i} X_{2,i} \\ \hat{\sigma}_1 &:= \begin{cases} \sigma_1, & \text{falls } \sigma_1, \sigma_2 \text{ bekannt} \\ \frac{1}{n} \sum_{i=1}^n X_{1,i}^2 & \text{falls } \sigma_1, \sigma_2 \text{ unbekannt} \end{cases} \end{aligned}$$

(2) E-Schritt:

$$E_{\Sigma^{(j)}}[\Lambda_{\Sigma^{(j)},1,2}(X_i)|X_1] = \begin{cases} 0, & \text{falls } \sigma_1, \sigma_2 \text{ bekannt} \\ \hat{\sigma}_{1,2}^{(j)} \left(\frac{X_1^2}{\hat{\sigma}_1} - 1 \right) & \text{falls } \sigma_1, \sigma_2 \text{ unbekannt} \end{cases}$$

(3) M-Schritt:

$$\begin{aligned} \hat{\sigma}_{1,2}^{(j+1)} &:= \hat{\sigma}_{1,2} := \{s \in \mathbb{R} : f(s) = 0\} \\ &\quad \text{falls } \sigma_1, \sigma_2 \text{ bekannt} \\ \hat{\sigma}_{1,2}^{(j+1)} &:= \frac{1}{n} \sum_{i=1}^n (I_{\{V_i=1\}} X_{1,i} X_{2,i} + I_{\{V_i=2\}} E_{\Sigma^{(j)}}[\Lambda_{\Sigma^{(j)},1,2}(X_i)|X_{1,i}]) \\ &\quad \text{falls } \sigma_1, \sigma_2 \text{ unbekannt} \\ j &:= j+1 \end{aligned}$$

dabei ergibt sich die Funktion f bei bekannten σ_1, σ_2 mit $M = \sum_i I_{\{V_i=1\}}$ als

$$f(s) = s^3 - \frac{1}{M} \sum_{i:V_i=1} s^2 X_{1,i} X_{2,i} + s(\sigma_1\sigma_2 - X_{1,i}\sigma_2 - X_{2,i}\sigma_1) + \sigma_1\sigma_2 X_{1,i} X_{2,i}.$$

9.6.3 statistische Eigenschaften der Schätzung aus dem EM-Algorithmus

Wir notieren, daß bei unbekanntem σ_1, σ_2 für festen Stichprobenumfang n stets gilt: $\hat{\sigma}_{1,2}^{(j)} \rightarrow \hat{\sigma}_{1,2}$, denn es ergeben sich als Rekursionsformeln

$$\sigma_{1,2}^{(j+1)} = A + B\sigma_{1,2}^{(j)}$$

mit $A = \frac{1}{n} \sum_{i=1}^n \mathbf{I}_{\{V_i=1\}} X_{1,i} X_{2,i}$ und $B = (\sum_{i=1}^n \mathbf{I}_{\{V_i=2\}} X_{1,i}^2) / (\sum_{i=1}^n X_{1,i}^2)$. Falls $\mathbf{I}_{\{V_i=2\}}$ für alle i , ist die Folge der $\hat{\sigma}_{1,2}^{(j)}$ konstant B ; sonst ist B betragsmäßig kleiner als 1, und damit geht $\sigma_{1,2}^{(j)} \rightarrow \frac{A}{1-B}$.

Proposition 9.6.6 *Setzen wir $M = \sum_i \mathbf{I}_{\{V_i=1\}}$, so ergibt der EM-Algorithmus*

(a) *Beispiel 9.6.3:*

$$\sqrt{n}(\hat{\sigma}_{1,2} - \sigma_{1,2}) \xrightarrow{w} \mathcal{N}\left(0, \frac{1}{1-\alpha} \frac{\sigma_1\sigma_2 - \sigma_{1,2}^2}{\sigma_1\sigma_2 + \sigma_{1,2}^2}\right) \quad (9.49)$$

Für festes $n < \infty$ ist für $M = 0$ der Schätzer nicht definiert.

(b) *Beispiel 9.6.4:*

$$\sqrt{n}(\hat{\sigma}_{1,2} - \sigma_{1,2}) \xrightarrow{w} \mathcal{N}\left(0, \frac{1}{1-\alpha} (\sigma_{1,2}^2 + \sigma_1\sigma_2) - \frac{2\alpha}{1-\alpha} \sigma_{1,2}^2\right) \quad (9.50)$$

Für festes $n < \infty$ ist für $M = 0$ der Schätzer nicht definiert. Ist $M = m > 0$, so definieren wir $\tilde{\alpha} := 1 - \frac{m}{n}$ und erhalten, daß

$$\mathbb{E}[\hat{\sigma}_{1,2} | M = m] = \sigma_{1,2} \quad (9.51)$$

$$\begin{aligned} n \text{Var}[\hat{\sigma}_{1,2} | M = m] &= 2\sigma_{1,2}^2 + \frac{n-2}{m-2} (\sigma_{1,2}^2 + \sigma_1\sigma_2) = \\ &= \left[\frac{(\sigma_{1,2}^2 + \sigma_1\sigma_2)}{1-\tilde{\alpha}} - \frac{2\sigma_{1,2}^2\tilde{\alpha}}{1-\tilde{\alpha}} + \frac{2d_s\tilde{\alpha}}{n(1-\tilde{\alpha})-2} \right] \mathbf{I}_{m>2} + \infty \mathbf{I}_{m\leq 2} \end{aligned} \quad (9.52)$$

und somit für festes n

$$\mathbb{E}[\hat{\sigma}_{1,2} | M > 0] = \sigma_{1,2} \quad (9.53)$$

$$\text{Var}[\hat{\sigma}_{1,2} | M > 0] = \infty; \quad (9.54)$$

Bemerkung 9.6.7 Damit liefert der EM-Algorithmus, wie zu Beginn dieses Abschnittes definiert, in den ersten beiden Beispielen jeweils ein asymptotisch optimales Verfahren, was die schwache Konvergenz anlangt. Allerdings ist in beiden Beispielen der Algorithmus für jedes finite n auf einer Menge positiven Maßes — mindestens $\{\omega : \sum_i \mathbf{I}_{\{V_i=1\}} = 0\}$ — nicht definiert. Selbst wenn man dieses Ereignis ausschließt, konvergieren die Varianzen in Beispiel 9.6.4 nicht mit. Dies scheint in der Untersuchung von Little und Rubin (1987), Abschnitt 6.3.2 übersehen worden zu sein — jedenfalls ist dies nicht weiter analysiert worden.

BEWEIS ZU PROPOSITION 9.6.6:

Beispiel 9.6.3:

Wir verifizieren die Voraussetzungen zu Theorem 9.5.3.

(a) Die LA(M)N-Eigenschaft folgt nach Theorem 9.3.5 aus der L_2 -Differenzierbarkeit des bivariaten Skalenmodells.

(b) Für die Straffheitsbedingung setzen wir

$$Y_n := (Y_{n,1}, Y_{n,2}, Y_{n,3})^\tau = \left(\sum_{i=1}^n \mathbb{I}_{\{V_i=1\}} (X_{1,i}^2, X_{1,i} X_{2,i}, X_{2,i}^2)^\tau \right) / \left(\sum_{i=1}^n \mathbb{I}_{V_i=1} \right),$$

$$Y := (Y_1, Y_2, Y_3)^\tau = (\sigma_1, \sigma_{1,2}, \sigma_2)^\tau / (1 - \alpha).$$

Damit fassen wir $f(s)$ als Funktion $f(s) = \tilde{f}(s, u)|_{u=Y_n}$ auf und erhalten mit $f_s = \frac{\partial}{\partial s} \tilde{f}$ und $f_u = \frac{\partial}{\partial u} \tilde{f}$

$$f_s(\sigma_{1,2}, Y) = \sigma_{1,2}^2 + \sigma_1 \sigma_2 =: s_1 > 0, \quad f_u(\sigma_{1,2}, Y)^\tau = (\sigma_{1,2} \sigma_2; -s_1; \sigma_{1,2} \sigma_1) =: s_2,$$

so daß nach dem Satz über implizite Funktionen sich die implizite Funktion $f(s, u) = 0$ lokal schreiben läßt als $s(u) = -s_1(u - Y)/s_2 + o(|u - Y|)$ mit $s(Y) = \sigma_{1,2}$. Andererseits konvergiert $\sqrt{n}(Y_n - Y)$ nach Slutsky und dem Zentralen Grenzwertsatz gegen eine Normalverteilung, ist also straff. Mit der endlich-dimensionalen Δ -Methode folgt damit die asymptotische Normalität von $\sqrt{n}(\hat{\sigma}_{1,2} - \sigma_{1,2}) = \sqrt{n}(s(Y_n) - s(Y))$ und damit auch die Straffheit.

(c) Mit $d_s^{(i)} := \det \Sigma_i$, und $\Sigma_i = \begin{pmatrix} \sigma_1 & \sigma_{1,2+h_i} \\ \sigma_{1,2+h_i} & \sigma_2 \end{pmatrix}$, $|h_i| < \alpha$, $|h_1 - h_2| < \delta$ gilt

$$\begin{aligned} & \left| \log \frac{dP_{\theta_2, n}}{dP_{\theta_0, n}}(x) - \log \frac{dP_{\theta_1, n}}{dP_{\theta_0, n}}(x) \right| = \\ & = \frac{1}{2} \left| \log d_s^{(1)} - \log d_s^{(2)} + x^\tau \Sigma_2^{-1} (\Sigma_2 - \Sigma_1) \Sigma_1^{-1} x \right| \leq K \delta |x|^2 \end{aligned}$$

für ein $0 < K = K_\alpha < \infty$, und damit folgt (37).

Beispiel 9.6.4:

Es gilt nach Lemma A.2.1

$$X_{2,i} =: E[X_{2,i} | X_{1,i}] + U_i = \sigma_{1,2}/\sigma_1 X_{1,i} + U_i,$$

mit $U_i \sim \mathcal{N}(0, \sigma_2 - \sigma_{1,2}^2/\sigma_1)$ und U_i, X_j stochastisch unabhängig. Damit ist

$$\begin{aligned} \sqrt{n}(\hat{\sigma}_{1,2} - \sigma_{1,2}) &= \sqrt{n} \left(\frac{\frac{1}{n} \sum_{i: V_i=1} (\frac{\sigma_{1,2}}{\sigma_1} X_{1,i}^2 + X_{1,i} U_i)}{\sum_{i: V_i=1} X_{1,i}^2 / \sigma_1} - \sigma_{1,2} \right) = \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\sigma_{1,2} \left(\frac{X_{1,i}^2}{\sigma_1} - 1 \right) + \mathbb{I}_{V_i=1} X_{1,i} U_i \frac{\sum_{i=1}^n X_{1,i}^2 / \sigma_1}{\sum_{i: V_i=1} X_{1,i}^2 / \sigma_1} \right) \end{aligned}$$

Nun ist $\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{I}_{V_i=1} X_{1,i} U_i$ straff und

$$\frac{\sum_{i=1}^n X_{1,i}^2 / \sigma_1}{\sum_{i: V_i=1} X_{1,i}^2 / \sigma_1} = \frac{1}{1 - \alpha} + o_P(n^0);$$

daher ist

$$\sqrt{n}(\hat{\sigma}_{1,2} - \sigma_{1,2}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{\sigma_{1,2}}{\sigma_1} (X_{1,i}^2 / \sigma_1 - 1) + X_{1,i} U_i \frac{\mathbb{I}_{V_i=1}}{1 - \alpha} \right) + o_P(n^0)$$

Setzen wir $R_i := \frac{\sigma_{1,2}}{\sigma_1}(X_{1,i}^2/\sigma_1 - 1) + X_{1,i}U_i \frac{1_{V_i=1}}{1-\alpha}$, so sind R_i unabhängig identisch verteilt, $E R_i = 0$,

$$E R_i^2 = 2\sigma_{1,2}^2 + \frac{1}{1-\alpha}d_s = \frac{1}{1-\alpha}(\sigma_{1,2}^2 + \sigma_1\sigma_2) - \frac{2\alpha}{1-\alpha}\sigma_{1,2}^2,$$

so daß die behauptete schwache Konvergenz mit dem Zentralen Grenzwertsatz folgt. Verwenden wir Lemma A.2.1, so erhalten wir mit $X_1 := \sigma(\{X_{1,i}\}_i)$

$$\begin{aligned} E[\hat{\sigma}_{1,2}|M > 0] &= E\{E[\frac{\sum_{i: V_i=1} X_{1,i}X_{2,i}}{n \sum_{i: V_i=1} X_{1,i}^2} \sum_{i=1}^n X_{1,i}^2 | X_1 | M > 0]\} = \\ &= \frac{\sigma_{1,2}}{n\sigma_1} E[\sum_{i=1}^n X_{1,i}^2 | M > 0] = \frac{\sigma_{1,2}}{n\sigma_1} E[\sum_{i=1}^n X_{1,i}^2] = \sigma_{1,2}, \end{aligned}$$

wobei wir wegen der Unabhängigkeit von X_1 und $\sigma(M)$ die Bedingungen hintereinanderschalten dürfen.

Für den Varianzterm setzen wir $\underline{V} = \sigma(V_1, \dots, V_n)$ und erhalten mit

$$E[X_1^2 X_2^2 | X_1] = \frac{\sigma_{1,2}^2}{\sigma_1^2} X_1^4 + (\sigma_2 - \frac{\sigma_{1,2}^2}{\sigma_1}) X_1^2$$

$$\begin{aligned} E[\hat{\sigma}_{1,2}^2 | \underline{V}] &= E\{E[\frac{(\sum_{i: V_i=1} X_{1,i}X_{2,i})^2}{n^2 (\sum_{i: V_i=1} X_{1,i}^2)^2} (\sum_{i=1}^n X_{1,i}^2)^2 | X_1 | \underline{V}]\} = \\ &= E\{ \frac{(\frac{\sigma_{1,2}^2}{\sigma_1^2} (\sum_{i: V_i=1} X_{1,i}^2)^2 + (\sigma_2 - \frac{\sigma_{1,2}^2}{\sigma_1}) \sum_{i: V_i=1} X_{1,i}^2) (\sum_{i=1}^n X_{1,i}^2)^2}{n^2 (\sum_{i: V_i=1} X_{1,i}^2)^2} | \underline{V} \} = \\ &= E\{ (\frac{\sigma_{1,2}^2}{n^2 \sigma_1^2} + \frac{(\sigma_2 - \frac{\sigma_{1,2}^2}{\sigma_1})}{n^2 (\sum_{i: V_i=1} X_{1,i}^2)}) (\sum_{i=1}^n X_{1,i}^2)^2 | \underline{V} \} = \\ &= \frac{1}{n^2} E\{ (\frac{\sigma_{1,2}^2}{\sigma_1^2} + \frac{(\sigma_2 - \frac{\sigma_{1,2}^2}{\sigma_1})}{\sigma_1 S_m}) \sigma_1^2 (S_m + S_{n-m})^2 | M = m \} = \\ &= \frac{1}{n^2} [\sigma_{1,2}^2 E[S_n^2] + (\sigma_1\sigma_2 - \sigma_{1,2}^2)(E[S_m + 2S_{n-m}] + E[S_{n-m}^2] E[\frac{1}{S_m}])] \end{aligned}$$

mit S_m, S_{n-m} stochastisch unabhängigen χ^2 -Variablen mit m resp. $n-m$ Freiheitsgraden und S_n einer χ_n^2 -Variablen.

Da $\sigma(\{M > 0\}) \subset \underline{V}$ gilt

$$\begin{aligned} E[\hat{\sigma}_{1,2}^2 | M > 0] &= E[E[\hat{\sigma}_{1,2}^2 | \underline{V}] | M > 0] = \\ &= \sum_{m>0} \frac{P(M=m)}{n^2 P(M>0)} [\sigma_{1,2}^2 E[S_n^2] + (\sigma_1\sigma_2 - \sigma_{1,2}^2)(E[S_m + 2S_{n-m}] + E[S_{n-m}^2] E[\frac{1}{S_m}])] \end{aligned}$$

Benützt man jetzt⁷, daß für $i \in \mathbb{Z}$

$$E(\chi_m^2)^i = \frac{2^i \Gamma(\frac{m}{2} + i)}{\Gamma(\frac{m}{2})},$$

⁷vgl. Johnson und Kotz (1970), Formel (10) in Chap. 17

so erhalten wir

$$E[\hat{\sigma}_{1,2}^2 | M = m > 0] = \sigma_{1,2}^2 \left(1 + \frac{2}{n}\right) + (\sigma_1 \sigma_2 - \sigma_{1,2}^2) \left[2n - m + \frac{(n-m)(n-m+2)}{m-2}\right] / n^2$$

und damit

$$n \operatorname{Var}[\hat{\sigma}_{1,2}^2 | M = m > 0] = \frac{1 - \frac{2}{n}}{\frac{m}{n} - \frac{2}{n}} \sigma_1 \sigma_2 - \frac{1 + \frac{2}{n} - \frac{m}{n}}{\frac{m}{n} - \frac{2}{n}} \sigma_{1,2}^2.$$

Integration über $\mathcal{L}(M)|_{M>0}$ ergibt $E \hat{\sigma}_{1,2} = \sigma_{1,2}$ und, da $E \frac{1}{U_m} = \infty$ für $m \leq 2$, folgt (54). ////

Daß man in den Beispielen 9.6.3 und 9.6.4 wirklich auf die bedingten L_2 -Ableitungen zurückgreifen muß, also nicht einfach die bedingte Variante des im vollen Modell optimalen Verfahrens verwenden kann, und daß man sich nicht ohne Informationsverlust auf die vollständig vorliegenden Beobachtungen zurückziehen kann, zeigt

Proposition 9.6.8 (a) *Verwenden wir in Beispiel 9.6.3 oder 9.6.4 statt des EM-Algorithmus' in der Form wie oben dargestellt nur die vollständig vorliegenden Beobachtungen zur Schätzung, so gilt mit*

$$\tilde{\sigma}_{1,2} := \frac{1}{\sum_i \mathbb{I}_{V_i=1}} \sum_{i: V_i=1} X_{1,i} X_{2,i},$$

$$\sqrt{n}(\tilde{\sigma}_{1,2} - \sigma_{1,2}) \xrightarrow{w} \mathcal{N}\left(0, \frac{\sigma_{1,2}^2 + \sigma_1 \sigma_2}{1 - \alpha}\right) \quad (9.55)$$

(b) *Verwenden wir in Beispiel 9.6.3 statt des EM-Algorithmus' in der Form wie oben dargestellt die nachfolgend dargestellte Variante, Algorithmus 9.6.9, so gilt*

$$\sqrt{n}(\tilde{\sigma}_{1,2} - \sigma_{1,2}) \xrightarrow{w} \mathcal{N}\left(0, \frac{\sigma_{1,2}^2 + \sigma_1 \sigma_2}{1 - \alpha} + \frac{2\alpha \sigma_{1,2}^2}{(1 - \alpha)^2}\right) \quad (9.56)$$

Algorithmus 9.6.9

(0) Initialisierung:

$$j := 0$$

$$\tilde{\sigma}_{1,2}^{(j)} := \frac{1}{\sum_i \mathbb{I}_{V_i=1}} \sum_{i: V_i=1} X_{1,i} X_{2,i}$$

(1) E-Schritt:

$$E_{\Sigma^{(j)}}[X_2 | X_1] = \tilde{\sigma}_{1,2}^{(j)} \left(\frac{X_1^2}{\sigma_1} - 1\right)$$

(2) M-Schritt:

$$\tilde{\sigma}_{1,2}^{(j+1)} := \frac{1}{n} \left(\sum_{i: V_i=1} X_{1,i} X_{2,i} + \sum_{i: V_i=2} X_{1,i} E_{\Sigma^{(j)}}[X_{2,i} | X_{1,i}] \right)$$

$$j := j + 1$$

Bemerkung 9.6.10 (a) Die Limes–Varianz in (a), $\frac{\sigma_{1,2}^2 + \sigma_1 \sigma_2}{1 - \alpha}$ ist um $\frac{2\alpha}{1 - \alpha} \sigma_{1,2}^2$ größer als die optimale Limes–Varianz bei unbekanntem σ_1, σ_2 , und damit auch schlechter als die optimale Limes–Varianz bei bekannten σ_1, σ_2 .

(b) Verwenden wir bei bekannten Nebenparametern σ_1, σ_2 den Schätzer aus dem EM–Algorithmus für unbekanntem Nebenparameter, wobei wir den bekannten Nebenparameterwert anstelle der Schätzung einsetzen, so erhalten wir in (b) einen Schätzer, der sogar noch schlechter ist, als der “naive” in (a). Den gleichen Effekt können wir auch im Fall der Schätzung von σ_2 beobachten, der in McLachlan und Krishnan (1997), Example 2.1 behandelt wird:

Auch hier konvergiert der EM–Algorithmus bei unbekanntem Nebenparametern gegen eine Normalverteilung mit optimaler Varianz $\frac{2}{1 - \alpha} (\sigma_2^2 - \alpha \frac{\sigma_{1,2}^4}{\sigma_1^2})$, aber verwendet man diesen Algorithmus bei bekannten Nebenparametern und setzt die bekannten Nebenparameterwerte ein, so erhalten wir wieder mit $\frac{2}{1 - \alpha} (\sigma_2^2 + \frac{\alpha}{1 - \alpha} \frac{\sigma_{1,2}^4}{\sigma_1^2})$ eine schlechtere Limes–Varianz als bei “naiver” Vorgehensweise wie in (a). Die Rechnungen dazu verlaufen analog zu den bisher gemachten und werden hier nicht präsentiert.

BEWEIS ZU PROPOSITION 9.6.8:

(a) Definiere $R_i := (X_{1,i} X_{2,i} - \sigma_{1,2}) I_{V_i=1}$. Dann gilt $E R_i = 0$, $E R_i^2 = (1 - \alpha)(\sigma_1 \sigma_2 + \sigma_{1,2})$, und R_i unabhängig identisch verteilt; daher gilt mit dem Zentralen Grenzwertsatz

$$\frac{1}{\sqrt{n}(1 - \alpha)} \sum_{i=1}^n R_i \xrightarrow{w} \mathcal{N}\left(0, \frac{\sigma_{1,2}^2 + \sigma_1 \sigma_2}{1 - \alpha}\right).$$

Damit ist $\frac{1}{\sqrt{n}} \sum_{i=1}^n R_i$ straff und somit folgt (55) mit Slutsky.

(b) Wie im Beweis zu Proposition 9.6.6 schreiben wir $X_{2,i} = \sigma_{1,2}/\sigma_1 X_{1,i} + U_i$, mit $U_i \sim \mathcal{N}(0, \sigma_2 - \sigma_{1,2}^2/\sigma_1)$ und U_i, X_j stochastisch unabhängig. Damit aber ist

$$\begin{aligned} \sqrt{n}(\tilde{\sigma}_{1,2} - \sigma_{1,2}) &= \sqrt{n} \left(\frac{\frac{1}{n} \sum_{i: V_i=1} (\frac{\sigma_{1,2}}{\sigma_1} X_{1,i}^2 + X_{1,i} U_i)}{1 - \frac{1}{n} \sum_{i: V_i=2} X_{1,i}^2 / \sigma_1} - \sigma_{1,2} \right) = \\ &= \left(\frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n \sigma_{1,2} (\frac{X_{1,i}^2}{\sigma_1} - 1) + \sum_{i: V_i=1} X_{1,i} U_i}{1 - \frac{1}{n} \sum_{i: V_i=2} X_{1,i}^2 / \sigma_1} \right) \end{aligned}$$

Der Nenner dieses Bruchs geht stochastisch gegen $1 - \alpha$, und für den Zähler sehen wir ein, daß die Summanden $R_i := \sigma_{1,2} (\frac{X_{1,i}^2}{\sigma_1} - 1) + X_{1,i} U_i I_{V_i=1}$ unabhängig identisch verteilt sind, wobei $E R_i = 0$,

$$E R_i^2 = 2\sigma_{1,2}^2 + (1 - \alpha)d_s = (1 - \alpha)(\sigma_{1,2}^2 + \sigma_1 \sigma_2) + 2\alpha\sigma_{1,2}^2,$$

so daß mit Slutsky und Zentralem Grenzwertsatz

$$\sqrt{n}(\tilde{\sigma}_{1,2} - \sigma_{1,2}) \xrightarrow{w} \mathcal{N}\left(0, \frac{\sigma_{1,2}^2 + \sigma_1 \sigma_2}{1 - \alpha} + \frac{2\alpha\sigma_{1,2}^2}{(1 - \alpha)^2}\right)$$

////

9.6.4 One-Step-EM-Algorithmus

Analog zu Algorithmus 9.5.5 definieren wir für Beispiele 9.6.3 und 9.6.4

Algorithmus 9.6.11

(0) Initialisierung:

$$\hat{\sigma}_{1,2}^{(0)} := \frac{1}{\sum I_{V_i=1}} \sum_{i: V_i=1} X_{1,i} X_{2,i}$$

$$\hat{\sigma}_1 := \begin{cases} \sigma_1, & \text{falls } \sigma_1, \sigma_2 \text{ bekannt} \\ \frac{1}{n} \sum_{i=1}^n X_{1,i}^2 & \text{falls } \sigma_1, \sigma_2 \text{ unbekannt} \end{cases}$$

(1) E-Schritt:

$$E_{\Sigma^{(0)}}[\Lambda_{\Sigma^{(0)}} | X_1] = \begin{cases} 0, & \text{falls } \sigma_1, \sigma_2 \text{ bekannt} \\ \hat{\sigma}_{1,2}^{(0)} \left(\frac{X_1^2}{\hat{\sigma}_1} - 1 \right) & \text{falls } \sigma_1, \sigma_2 \text{ unbekannt} \end{cases}$$

(2) M-Schritt:

$$\hat{\sigma}_{1,2} := \sigma_{1,2}^{(0)} + \frac{1}{n} \sum_{i: V_i=1} \frac{\sigma_1 \sigma_2 - (\sigma_{1,2}^{(0)})^2}{\sigma_1 \sigma_2 + (\sigma_{1,2}^{(0)})^2} [X_{1,i} X_{2,i} + \sigma_{1,2}^{(0)} (1 - X_i^T (\Sigma^{(0)})^{-1} X_i)]$$

falls σ_1, σ_2 bekannt

$$\hat{\sigma}_{1,2} := \frac{1}{n} \left(\sum_{i=1}^n I_{V_i=1} \left\{ \frac{1}{1-\alpha} X_1 X_2 - \frac{\alpha \sigma_{1,2}^{(0)}}{(1-\alpha) \hat{\sigma}_1} X_1^2 \right\} + I_{V_i=2} \hat{\sigma}_{1,2}^{(0)} \left(\frac{X_1^2}{\hat{\sigma}_1} - 1 \right) \right)$$

falls σ_1, σ_2 unbekannt

$$\text{mit } \Sigma^{(0)} = \begin{pmatrix} \sigma_1 & \sigma_{1,2}^{(0)} \\ \sigma_{1,2}^{(0)} & \sigma_2 \end{pmatrix}.$$

Dies macht klar, daß man bei einem geringeren Aufwand — es sind keine Iterationen nötig — die gleiche asymptotische Effizienz wie beim klassischen EM-Algorithmus erreichen kann; die Pathologie für finites n , $M = 0$ scheint unumgänglich, da in diesem Fall kein sinnvoller Startschätzer definiert werden kann. Allerdings treten für $M > 0$ keine Schwierigkeiten auf, — insbesondere konvergieren auf $M > 0$ die Varianzen mit. Dies kann man zeigen, indem man die gleichgradige Quadrat-Integrierbarkeit der $\hat{\sigma}_{1,2}^{(n)}$ nachweist.

Kapitel 10

Glattheit des Zustandsraummodells

10.1 Motivation

In diesem Kapitel wollen wir die Begriffe der L_2 -Differenzierbarkeit — auch für vergrößerte Modelle wie in Kapitel 9 — auf Modelle mit Zeitabhängigkeit, insbesondere das Zustandsraummodell, übertragen.

Hierzu verallgemeinern wir zunächst in Abschnitt 10.2.1 das Wurzeldichtenkalkül aus dem vorigen Kapitel für Schemata von Martingaldifferenzen, definieren dann L_2 -Differenzierbarkeit in diesem Kontext und zeigen anschließend in Theorem 10.2.7, daß diese die etwas allgemeiner gefaßte LAMN-Eigenschaft (vgl. Definition 10.2.2) nach sich zieht.

In Theorem 10.2.10 weisen wir nach, daß sich L_2 -Differenzierbarkeit wie in Kapitel 9 auf geeignet vergrößerte Modelle überträgt. Unter diese Voraussetzungen fällt auch das lineare, zeitinvariante Zustandsraummodell, von dem wir die L_2 -Differenzierbarkeit in Theorem 10.3.2 nachweisen.

Um dabei nicht nur “Filter-Filtrierungen” zulassen zu können, also $\sigma((Y_s)_{s \leq t})$ als Subfiltrierung der natürlichen Filtrierung $\sigma((\beta_s, Y_s)_{s \leq t})$, werden in Abschnitt 10.3.3 Argumente dafür geliefert, auch Glättungs-Filtrierungen, also $\sigma((Y_s)_{s \leq t+m})$ als Subfiltrierung der natürlichen Filtrierung auffassen zu dürfen. Diese erweisen sich auch als wichtig, wenn es darum geht, die Identifizierbarkeit des Modells zu gewährleisten, in unserem Kontext die Frage, ob die Fisher-Information des vergrößerten Modells vollen Rang hat. Dieser Frage gehen wir in einem Exkurs in Abschnitt 10.4 gesondert nach. Dabei schließen wir unter gewissen Rangvoraussetzungen an die Hyper-Parameter auf die Regularität der Fisher-Information, indem wir nachweisen, daß dann ein (suboptimaler) asymptotisch linearer Schätzer in Gestalt eines Momentenschätzers existiert, der lokal gleichmäßig asymptotisch normalverteilt ist.

In Abschnitt 10.5 verifizieren wir die gefundenen Bedingungen für das normale Zustandsraummodell. Es wird ein allein auf Glätten bis maximal Ordnung 2 beruhender One-Step-EM-Algorithmus vorgestellt und dieser dem EM-Algorithmus von Shumway und Stoffer (1982) gegenübergestellt.

Abschließend wird im normalen Zustandsraummodell auch ein robuster One-Step-EM-Algorithmus präsentiert.

10.2 lokal–asymptotische Theorie für rekursive Modelle

Basierend auf einer Seminarsausarbeitung von H. Rieder verallgemeinern wir das Wurzeldichtenkalkül für die Situation unabhängiger Beobachtungen aus dem vorigen Kapitel für Schemata von Martingaldifferenzen.

10.2.1 Wurzeldichtenkalkül für Martingaldifferenzen

Sei von nun an der meßbare Raum (Ω, \mathcal{F}) versehen mit der Filtrierung

$$(\emptyset, \Omega) =: \mathcal{F}_{-1} \subset \mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_n \subset \mathcal{F} \quad (10.1)$$

Interpretiert man dann die zweite Komponente der Elemente aus $\mathcal{L}_2(\mathcal{F})$ nicht als Maß, sondern als Erwartungswert–Funktional, so kann man zu bedingten Erwartungswerten übergehen.

Damit ist es möglich, das Wurzeldichtenkalkül längs aufsteigenden Filtrierungen auf Maße aus $\mathcal{M}_b(\mathcal{F})$ zu erweitern, die keine Produktgestalt haben. Dieses Kalkül erfaßt auch einen endlichen Abschnitt der Filtrierung, und es sind auch nur partiell auf der Filtrierung definierte Maße zugelassen. Wir definieren also für $m > n$:

$$\mathcal{L}_2(\mathcal{F}_{m|n}) := \{(\xi, E_{P_m}[\cdot|\mathcal{F}_n]) \mid \xi \in L_2(P_m), P_m \in \mathcal{M}_b(\mathcal{F}_m)\} \quad (10.2)$$

Dabei gehen wir ohne Einschränkung davon aus, daß alle Maße P_m aus $\mathcal{M}_b(\mathcal{F}_m)$ eine Fortsetzung P auf \mathcal{F} besitzen. Verkürzt schreiben wir für die Paare $(\xi, E_{P_m}[\cdot|\mathcal{F}_n])$ auch $\xi\sqrt{dP_{m|n}}$. Auf diesen Paaren definieren wir eine Abbildung auf $\mathcal{L}_2(\mathcal{F}_{m|n}) \times \mathcal{L}_2(\mathcal{F}_{m|n})$ als:

$$\left\langle \xi\sqrt{dP_{m|n}} \mid \eta\sqrt{dQ_{m|n}} \right\rangle_{\mathcal{L}_2(\mathcal{F}_{m|n})} := E_\mu \left[\xi^\tau \eta \frac{\sqrt{p_m q_m}}{\sqrt{p_n q_n}} \mid \mathcal{F}_n \right] \quad (10.3)$$

Diese kann in natürlicher Weise bilinear¹ fortgesetzt werden, die Addition erklärt sich dementsprechend.

Für ein Maß $\lambda \in \mathcal{L}_2(\mathcal{F})$ bezeichne $\lambda_k := \lambda|_{\mathcal{F}_k}$ die Einschränkung auf \mathcal{F}_k . Die in der Gleichung für $k = n, m$ auftauchenden Größen p_k, q_k sind Radon–Nikodym–Ableitungen von P_k resp. Q_k bezüglich μ_k , wobei μ irgendein P und Q dominierendes Maß in $\mathcal{M}_b(\mathcal{F})$ ist, bezüglich dem auch der Erwartungswert gebildet worden ist. Der bedingte Erwartungswert in (3) ist wegen der Integrabilitätsvorgaben an ξ, η wohldefiniert. Da $p_k = 0$ μ -f.s. $p_{k+1} = 0$ impliziert, läßt sich der Ausdruck durch 0 in den Nullstellen des Nenners konsistent fortsetzen. Als Spezialfall erhält man μ -fast-sicher bis auf multiplikative Konstanten für $n = -1$ den $\mathcal{L}_2(\mathcal{F}_m)$ zurück, denn p_{-1}, q_{-1} müssen bezüglich der trivialen σ -Algebra meßbar sein, sie sind also μ -fast sicher Konstanten, und der mit der trivialen σ -Algebra ankonditionierte Erwartungswert stimmt μ -fast-sicher mit dem unbedingten überein. Läßt man nur normierte Maße zu und schließt die

¹Weil diese Abbildung im Fall, daß \mathcal{F}_n die triviale σ -Algebra ist, nach dem weiter unten durchgeführten Übergang zu Äquivalenzklassen ein Skalarprodukt ist, mit dem der Raum der (unbedingten) Wurzeldichten zum Hilbertraum wird, benützen wir hierfür die Skalarprodukt–Notation, und ebenso das weiter unten eingeführte Symbol $\|\cdot\|$.

Skalen in die Linearfaktoren ein, so ist die multiplikative Konstante 1. Weiter führen wir die Abbildung

$$\xi \sqrt{dP_{m|n}} \mapsto \|\xi \sqrt{dP_{m|n}}\| := \left\langle \xi \sqrt{dP_{m|n}} \mid \xi \sqrt{dP_{m|n}} \right\rangle_{\mathcal{L}_2(\mathcal{F}_{m|n})}^{\frac{1}{2}}$$

ein und identifizieren die Elemente $\xi \sqrt{dP_{m|n}}$, $\eta \sqrt{dQ_{m|n}}$,

$$\xi \sqrt{dP_{m|n}} \equiv \eta \sqrt{dQ_{m|n}} \quad \Leftrightarrow \quad \|\xi \sqrt{dP_{m|n}} - \eta \sqrt{dQ_{m|n}}\| = 0 \quad [\mu]. \quad (10.4)$$

Man beachte, daß wir keine Domination des gesamten Raums für die Existenz dominierender Maße μ benötigen, wenn wir jeweils nur abzählbar viele Elemente aus $\mathcal{L}_2(\mathcal{F}_{m|n})$ betrachten, und daß die Äquivalenz nicht von der Wahl des dominierenden Maßes abhängt.

Der Anlaß für diese Notation ist folgende Überlegung: Seien $B_m \in \mathcal{F}_m$ und $B_n \in \mathcal{F}_n$ beliebig, so gilt

$$\begin{aligned} P(B_m \cap B_n) &= \int_{B_n} \mathbf{I}_{B_m} dP_m = \int_{B_n} \mathbf{I}_{B_m} p_m d\mu = \int_{B_n} E_\mu[\mathbf{I}_{B_m} p_m \mid \mathcal{F}_n] d\mu = \\ &= \int_{B_n} E_\mu[\mathbf{I}_{B_m} \frac{p_m}{p_n} \mid \mathcal{F}_n] dP_n = \int_{B_n} E_\mu[\mathbf{I}_{B_m} \frac{p_m}{p_n} \mid \mathcal{F}_n] dP_m = \end{aligned} \quad (10.5)$$

$$= \int_{B_n} E_{P_m}[\mathbf{I}_{B_m} \mid \mathcal{F}_n] dP_m = \int_{B_n} E_{P_m}[\mathbf{I}_{B_m} \mid \mathcal{F}_n] dP_n \quad (10.6)$$

und damit

$$\int_{B_m} dP_{m|n} := E_{P_m}[\mathbf{I}_{B_m} \mid \mathcal{F}_n] = E_\mu[\mathbf{I}_{B_m} \frac{p_m}{p_n} \mid \mathcal{F}_n] \quad [P_n] \quad (10.7)$$

Im folgenden schreiben wir suggestiv $p_{m|n} := p_m/p_n$. Damit folgt insbesondere in der oben eingeführten Notation

$$dP_n = \prod_{j=0}^n dP_{j|j-1} \quad [P_n] \quad (10.8)$$

$$\sqrt{dP_n} = \prod_{j=0}^n \sqrt{dP_{j|j-1}} \quad [P_n] \quad (10.9)$$

$$\log(dP_n) = \sum_{j=0}^n \log(dP_{j|j-1}) \quad [P_n] \quad (10.10)$$

Wir bemerken an dieser Stelle, daß wir keine Existenz von regulären bedingten Wahrscheinlichkeiten (Markov-Kernen) voraussetzen. Haben wir jedoch diese zur Verfügung — wie zum Beispiel in polnischen Räumen —, so können wir $d\mu_m$ schreiben als $d\mu_{m|n}d\mu_n$ mit $d\mu_{m|n}$ einem Markov-Kern, wobei $p_{m|n}$ dann eine bedingte $\mu_{m|n}$ -Dichte ist.

Mit “algebraischer Induktion” können wir diese Definition auf Treppenfunktionen erweitern, dann auf monotone Limiten und gelangen schließlich zu allgemeinen $\xi \in L_2^k$.

Insgesamt lassen sich damit die Elemente $\xi \sqrt{dP_{m|n}}$ auffassen als Abbildungen von $L_2^k(P_m)$ nach $L_1(P_n)$, $\eta \mapsto E_{P_m}[\eta^\tau \xi \mid \mathcal{F}_n]$.

Bemerkung 10.2.1 In diesem Abschnitt haben wir von der diskreten Struktur der Beobachtungszeitpunkte j keinen Gebrauch gemacht. Dementsprechend gilt das Kalkül auch für stetige Filtrierungen \mathcal{F}_s . Nicht so klar ist die Übertragung der im folgenden diskutierten LA(M)N-Eigenschaft auf kontinuierliche Filtrierungen.

10.2.2 Log–Likelihood–Entwicklung bei Martingaldifferenzen

Die LAN–Eigenschaft (9.19) ist, wie wir im Abschnitt über die Konsequenzen aus Theorem 9.3.5 gesehen haben, sehr nützlich für die Bestimmung eines optimalen Verfahrens im Sinn des Faltungssatzes oder des asymptotischen Minimaxtheorems.

Für einige Modelle mit stochastischer Abhängigkeit können wir nicht die klassische LAN–Definition verwenden; stattdessen verwenden wir den Begriff LAMN:

Definition 10.2.2 Sei (Ω, \mathcal{F}) ein meßbarer Raum,

$$(\emptyset, \Omega) =: \mathcal{F}_{-1} \subset \mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F} \quad (10.11)$$

eine aufsteigende Filtrierung von \mathcal{F} ,

$$\mathcal{P}_n = \{P_{\theta,n} \in \mathcal{M}_1(\mathcal{F}_n) \mid \theta \in \Theta\} \quad (10.12)$$

ein parametrisches Modell längs dieser Filtrierung mit Parameterbereich Θ , einer offenen Teilmenge des \mathbb{R}^k , $k \geq 1$. Dann erfüllt die Folge von Familien von Maßen $(P_{\theta,n})_n$ in $\theta \in \Theta$ die LAMN–Bedingung (LAMN=local asymptotic mixing normal), wenn folgende Bedingungen erfüllt sind:

(A1) Es existiert eine Folge von \mathcal{F}_n –meßbaren, k –dimensionalen Zufallsvariablen $W_n = W_{\theta,n}$ und eine Folge von \mathcal{F}_n –meßbaren, $k \times k$ –dimensionalen Zufallsmatrizen $T_n = T_{\theta,n}$, die $P_{\theta,n}$ –f–s positiv definit sind, so daß für

$$\hat{\theta} = \hat{\theta}_n = \theta + s_n h \quad (10.13)$$

mit Standardisierungsfolge $s_n^\tau = s_n \succ 0$ und $h \in \mathbb{R}^k$ beliebig aber fest gilt:

$$\log \frac{dP_{\hat{\theta}_n}}{dP_{\theta,n}} = h^\tau T_n^{\frac{1}{2}} W_n - \frac{1}{2} h^\tau T_n h + o_{P_{\theta,n}}(n^0) \quad (10.14)$$

(A2) Es existiert eine fast sicher positiv definite, $k \times k$ –dimensionale Zufallsmatrix T , so daß

$$\begin{pmatrix} W_n \\ T_n \end{pmatrix} \circ P_{\theta,n} \xrightarrow{w} \mathcal{L} \begin{pmatrix} W \\ T \end{pmatrix} \quad (10.15)$$

wobei $W \sim \mathcal{N}_k(0, \mathbb{I}_k)$ und W unabhängig von T .

Die stärkere LAN–Eigenschaft aus Definition A.5.2 ist in Kraft, falls zusätzlich das Limeselement T eine Konstante ist.

Bemerkung 10.2.3 Im Unterschied zur Notation in Kapitel 9 schreiben wir nun den Parameter θ vor den Beobachtungs–/Zeitindex, weil wir so leichter mit Ausdrücken $s|t$ zurecht kommen, wie sie sich durch Bedingen in Abschnitt 10.2.1 ergeben.

Bemerkung 10.2.4 Für die spätere Untersuchung der LA(M)N–Eigenschaft des Zustandsraummodells werden wir im Kontext dieser Arbeit nur mit der LAN–Eigenschaft arbeiten; will man aber das allgemeine — auch möglicherweise nicht (asymptotisch) stationäre —, zeitinvariante Modelle untersuchen, also

bei der Schätzung der Hyper-Parameter auch $\rho(F) > 1$ zulassen, so ist zu diesem Zweck der LAMN-Setup der adäquate (vgl. Jeganathan (1988), part II). Dazu müßten wir dann aber auch Theorem 10.2.10 sowie Definition 10.2.11 entsprechend umformulieren, worauf wir hier verzichten.

10.2.3 L_2 -Differenzierbarkeit im Fall abhängiger Variablen

Eine hinreichende Bedingung für diese LAMN- bzw. LAN-Eigenschaft liefert, wie in diesem Kapitel gezeigt wird, folgende Version von L_2 -Differenzierbarkeit:

Definition 10.2.5 Sei \mathcal{P}_n ein parametrisches Modell gemäß (12), und sei jedes \mathcal{F}_n aus der Filtrierung (11) unterfiltriert in

$$(\emptyset, \Omega) =: \mathcal{F}_{n,-1} \subset \mathcal{F}_{n,0} \subset \mathcal{F}_{n,1} \subset \dots \subset \mathcal{F}_{n,n} = \mathcal{F}_n. \quad (10.16)$$

Weiter mögen $\sqrt{dP_{\hat{\theta},n;j|j-1}}$, $\sqrt{dP_{\theta,n;j|j-1}}$ die in Abschnitt 10.2.1 definierten Wurzel-dichten auf $(\mathcal{F}_{n,j})_{j \leq n}$ zu den Maßen $P_{\hat{\theta},n}$, $P_{\theta,n}$ bezeichnen, wobei $\hat{\theta} = \hat{\theta}_n$ eine Alternativenfolge gemäß (13) ist.

Dann ist das Modell in $\theta \in \Theta$ L_2 -differenzierbar, wenn es ein Dreiecksschema $\Lambda_{n,j} \in L_2^k(P_{\theta;n,j})$, ($j \leq n$), mit Standardisierungsfolge $s_n^\tau = s_n > 0$ und folgenden Eigenschaften gibt²:

(L2-1) “bedingte Zentriertheit”: Für alle $j \geq 0$ gilt

$$E[\Lambda_{n,j} | \mathcal{F}_{n,j-1}] = 0 \quad (10.17)$$

(L2-2) “endliche, unbedingte Fisher-Information”: Für alle $h \in \mathbb{R}^k$ gilt

$$\sup_{n \geq 0} \sum_{j=1}^n E \left[|h^\tau s_n \Lambda_{n,j}|^2 \right] < K < \infty \quad (10.18)$$

(L2-3) “unbedingte Lindebergbedingung”: Für alle $h \in \mathbb{R}^k$ und alle $\epsilon > 0$ gilt

$$\sum_{j=0}^n E \left[|h^\tau s_n \Lambda_{n,j}|^2 I_{|h^\tau s_n \Lambda_{n,j}| > \epsilon} \right] \xrightarrow{n \rightarrow \infty} 0 \quad (10.19)$$

(L2-4) “Wurzel-dichtenbedingung”: Für alle $h \in \mathbb{R}^k$ gilt

$$\sum_{j=0}^n E_\mu \left\| \sqrt{dP_{\hat{\theta},n;j|j-1}} - \sqrt{dP_{\theta,n;j|j-1}} \left(1 + \frac{h^\tau s_n \Lambda_{n,j}}{2} \right) \right\|^2 \xrightarrow{n \rightarrow \infty} 0 \quad (10.20)$$

(L2-5) “Konvergenz der bedingten Fisher-Information”: Es existiert eine fast sicher positive definite Zufallsmatrix T auf (Ω, \mathcal{F}) so daß

$$s_n \sum_{j=0}^n E \left[\Lambda_{n,j} \Lambda_{n,j}^\tau | \mathcal{F}_{n,j-1} \right] s_n = T + o_{P_n}(n^0) \quad (10.21)$$

²Die Erwartungswerte sind — wenn nicht explizit anders gekennzeichnet — im folgenden immer Erwartungswerte unter $P_{\theta,n}$.

Bemerkung 10.2.6 (a) Für die LAN-Eigenschaft ist die Approximierbarkeit von $\Lambda_{n,j}$ in $L_2(P_{\theta,n})$ durch eine ergodische Folge Λ_j hinreichend.

(b) Wir lassen von n abhängige Filtrierungen $\mathcal{F}_{n,j}$ zu, um die Begriffe auch auf Glättungsfiltrierungen wie in Abschnitt 10.3.3 anwenden zu können.

Mit diesen Eigenschaften kann man zeigen

Theorem 10.2.7 *Ist ein parametrisches Modell $\mathcal{P}_n = \{P_{\theta,n} \mid \theta \in \Theta\}$ mit Θ einer offenen Teilmenge des \mathbb{R}^k L_2 -differenzierbar im Sinn von Definition 10.2.5, so gilt die LAMN-Eigenschaft aus Definition 10.2.2; gilt zusätzlich noch Ergodizität der Folge Λ_j , so besitzt das Modell die LAN-Eigenschaft aus Definition A.5.2.*

Bemerkung 10.2.8 (a) Jeganathan (1982) gibt einen Beweis für die Situation, in der $\{P_{\theta,n}, \theta \in \Theta\}$ durch σ -finite Maße μ_n dominiert werden, die die Abhängigkeitsstruktur der Beobachtungen $\{X_j\}$ in dem Sinn respektieren, daß es reguläre bedingte Wahrscheinlichkeiten (Markov-Kerne)

$$P_{\theta,j}(\cdot \mid X_{j-1} = x_{j-1}, X_{j-2} = x_{j-2}, \dots)$$

gibt, die absolutstetig bezüglich μ_j sind.

Da wir mit dem oben angegebenen Wurzeldichtenkalkül diese Voraussetzungen nicht benötigen — für **(L2-4)** kann stets ein dominierendes Wahrscheinlichkeitsmaß μ auf \mathcal{F} in Abhängigkeit der betrachteten Alternativenfolge $P_{\hat{\theta}_{(n)},n}$ und $P_{\theta,n}$ gewählt werden, und Regularität wurde im Kalkül nirgends benötigt — geben wir hier einen Beweis, der weitestgehend die Notation und Argumentation von Rieder (1994), Thm. 2.3.5 [dort für die unabhängige Situation] übernimmt.

(b) Liegt globale Domination von $\mathcal{P}_n = \{P_{\theta,n}, \theta \in \Theta\}$ durch ein σ -finites Maß μ auf \mathcal{F} vor, das zugleich in Markov-Kerne bezüglich $\mathcal{F}_{n,\cdot}$, faktorisiert — symbolisch $d\mu_n = \prod_j d\mu_{n;j|j-1}$ — so folgt daraus bereits die Existenz regulärer bedingter Wahrscheinlichkeitsdichten $p_{\theta,n;j|j-1}$:

Aus (5) folgt für $m < n$, $B_n \in \mathcal{F}_n$, $B_m \in \mathcal{F}_{n,m}$ beliebig

$$P_\theta(B_m \cap B_n) = \int_{B_m} E_\mu[\mathbb{I}_{B_n} p_{\theta;n;n|m} \mid \mathcal{F}_{n,m}] dP_n.$$

Wegen der Faktorisierbarkeit von μ läßt sich der bedingte Erwartungswert gegeben $\mathcal{F}_{n,j}$ gewinnen als

$$E_\mu[\mathbb{I}_{B_n} p_{\theta;n;n|m} \mid \mathcal{F}_{n,m}] = \int \mathbb{I}_{B_n} p_{\theta;n;n|m} d\mu_{n;n|m},$$

insbesondere also $P_{\theta;n;n|m} \gg \mu_{n,m}$.

BEWEIS: Wir unterdrücken im folgenden die Abhängigkeit in θ und die der Filtrierungen in n und schreiben für die Alternativen mit $Q_n := P_{\hat{\theta},n}$

$$L^{(n)} := \log \frac{dQ_n}{dP_n} \stackrel{(10.10)}{=} \sum_{j=0}^n \log \frac{dQ_{n;j|j-1}}{dP_{j|j-1}} =: L_{n,j} \quad [Q_n + P_n],$$

wobei wir $\sum_j L_{n,j} = \infty$ setzen, falls ein $L_{n,j} = \infty$. Wir führen ein

$$L_{n,j} = 2 \log(1 + Z_{n,j}), \quad Z_{n,j} := \exp(L_{n,j}/2) - 1 = \sqrt{\frac{\mathbb{E}_\mu[dQ_{n,j}|\mathcal{F}_{j-1}]}{\mathbb{E}_\mu[dP_j|\mathcal{F}_{j-1}]} - 1} \quad (10.22)$$

und weiter analog zu Rieder (1994)

$$U_{n,j} := h^\tau s_n \Lambda_{n,j}, \quad (10.23)$$

$$D_{n,j} := \sqrt{dQ_{n,j|j-1}} - \sqrt{dP_{j|j-1}} \left(1 + \frac{U_{n,j}}{2}\right) \quad (10.24)$$

und außerdem

$$\mathbb{E}_n[\cdot] := \mathbb{E}_{P_n}[\cdot], \quad \text{Var}_n[\cdot] := \text{Var}_{P_n}[\cdot], \quad \mathbb{E}_{n,j}[\cdot] := \mathbb{E}_n[\cdot|\mathcal{F}_{j-1}], \quad (10.25)$$

$$T_n := s_n \sum_j \mathbb{E}_{n,j}[\Lambda_{n,j} \Lambda_{n,j}^\tau] s_n, \quad (10.26)$$

$$W_n := T_n^{-\frac{1}{2}} s_n \sum_j \Lambda_{n,j} \quad (10.27)$$

Damit erhalten wir mit $c(u) = (\log(1+u) - u)/u^2$

$$L^{(n)} = 2 \sum_j (Z_{n,j} + c(Z_{n,j}) Z_{n,j}^2) \quad [Q_n + P_n] \quad (10.28)$$

Wir behaupten, daß wie in Rieder (1994) gilt

$$\max_j |Z_{n,j}| = o_{P_n}(n^0) \quad (10.29)$$

$$2 \sum_j Z_{n,j} = \sum_j U_{n,j} - \frac{1}{4} \sum_j \mathbb{E}_{n,j} U_{n,j}^2 + o_{P_n}(n^0) \quad (10.30)$$

$$\sum_j Z_{n,j}^2 = \frac{1}{4} \sum_j \mathbb{E}_{n,j} U_{n,j}^2 + o_{P_n}(n^0) \quad (10.31)$$

$$|\sum_j (2c(Z_{n,j}) + 1) Z_{n,j}^2| = o_{P_n}(n^0) \quad (10.32)$$

Damit wäre (A1) bewiesen, denn

$$\begin{aligned} L^{(n)} &\stackrel{(10.30)}{=} \sum_j U_{n,j} - \frac{1}{4} \sum_j \mathbb{E}_{n,j} U_{n,j}^2 + \sum_j 2c(Z_{n,j}) Z_{n,j}^2 + o_{P_n}(n^0) = \\ &\stackrel{(10.31)}{=} \sum_j U_{n,j} - \frac{1}{2} \sum_j \mathbb{E}_{n,j} U_{n,j}^2 + \sum_j (2c(Z_{n,j}) - 1) Z_{n,j}^2 + o_{P_n}(n^0) = \\ &\stackrel{(10.32)}{=} \sum_j U_{n,j} - \frac{1}{2} \sum_j \mathbb{E}_{n,j} U_{n,j}^2 + o_{P_n}(n^0) = \\ &= \sum_j h^\tau T_n^{-\frac{1}{2}} W_n - \frac{1}{2} h^\tau T_n h + o_{P_n}(n^0) \end{aligned}$$

Zu den Behauptungen (29)–(32):

Sei $\varepsilon > 0$; es gilt:

$$\begin{aligned} P_n(\max_j |Z_{n,j}| > 2\varepsilon) &= P_n(\max_j |2Z_{n,j} - U_{n,j} + U_{n,j}| > 4\varepsilon) \leq \\ &\leq P_n(\max_j |U_{n,j}| > \varepsilon) + P_n(\max_j |Z_{n,j} - \frac{1}{2}U_{n,j}| > \varepsilon) \leq \\ &\leq \frac{1}{\varepsilon^2} \{ \sum_j \mathbb{E}_n[\mathbf{I}_{\{|U_{n,j}| > \varepsilon\}} U_{n,j}^2] + \mathbb{E}_n[|Z_{n,j} - \frac{1}{2}U_{n,j}|^2] \} = \\ &\stackrel{(L2-3)}{\stackrel{(L2-4)}{=}} o(n^0), \end{aligned}$$

und damit (29), denn im $\mathcal{L}_2(\mathcal{F}_{j|j-1})$, den wir nach Realisationen in \mathcal{F}_{j-1} faktorisieren, ist

$$\mathbb{E}_n[|Z_{n,j} - \frac{1}{2}U_{n,j}|^2] = \mathbb{E}_\mu \|D_{n,j}\|^2 - Q_n(\{L_{n,j} = \infty\})$$

Wir bemerken daß $Y_{n,j} := 2Z_{n,j} - 2\mathbf{E}_{n,j}[Z_{n,j}] - U_{n,j}$ wegen **(L2-1)** ein Martingaldifferenzschema ist; daher gilt

$$\mathrm{Var}_n[\sum_j Y_{n,j}] = \sum_j \mathrm{Var}_n[Y_{n,j}] \leq \sum_j \mathbf{E}_n[(2Z_{n,j} - U_{n,j})^2] = o(n^0) \quad (10.33)$$

wie oben, so daß insbesondere

$$2Z_{n,j} = 2\mathbf{E}_{n,j}[Z_{n,j}] - U_{n,j} + o_{P_n}(n^0) \quad (10.34)$$

Wie in Rieder (1994) gilt in $\mathcal{L}_2(\mathcal{F}_{n,j|j-1})$

$$\begin{aligned} 2[\mathbf{E}_{n,j}[Z_{n,j}]] &= 2\langle \sqrt{dQ_{n,j|j-1}} | \sqrt{dP_{j|j-1}} \rangle - 2 = \\ &= -\|\sqrt{dQ_{n,j|j-1}} - \sqrt{dP_{j|j-1}}\|^2 \end{aligned}$$

und mit $||a| - |b|| \leq |a - b|$ folgt mit **(L2-4)**

$$\begin{aligned} &\mathbf{E}_\mu \left(\left| \sum_j \|\sqrt{dQ_{n,j|j-1}} - \sqrt{dP_{j|j-1}}\|^2 \right|^{\frac{1}{2}} - \left| \frac{1}{4} \sum_j \mathbf{E}_{n,j}[U_{n,j}^2] \right|^{\frac{1}{2}} \right)^2 \leq \\ &\leq \mathbf{E}_\mu \sum_j \|D_{n,j}\|^2 \stackrel{(\mathbf{L2-4})}{=} o(n^0), \end{aligned}$$

also insbesondere

$$-\sum_j \mathbf{E}_{n,j}[Z_{n,j}] = \sum_j \|\sqrt{dQ_{n,j|j-1}} - \sqrt{dP_{j|j-1}}\|^2 = \frac{1}{4} \sum_j \mathbf{E}_{n,j}[U_{n,j}^2] + o_{P_n}(n^0)$$

und damit (30). Wieder aus der Dreiecksungleichung folgt

$$\left(\left| 4 \sum_j Z_{n,j}^2 \right|^{\frac{1}{2}} - \left| \sum_j U_{n,j}^2 \right|^{\frac{1}{2}} \right)^2 \leq \sum_j (2Z_{n,j} - U_{n,j})^2 \stackrel{(10.33)}{=} o_{P_n}(n^0)$$

Daher zeigen wir für (31)

$$\sum_j U_{n,j}^2 = \sum_j \mathbf{E}_{n,j} U_{n,j}^2 + o_{P_n}(n^0) \quad (10.35)$$

Sei $\eta \in (0, \infty)$, und setzen wir $V_{n,j} := U_{n,j} \mathbf{I}_{\{|U_{n,j}| \leq \eta\}}$, so gilt

$$\begin{aligned} P_n(\sum_j V_{n,j}^2 \neq \sum_j U_{n,j}^2) &\leq \sum_j P_n(|U_{n,j}| \geq \eta) \leq \\ &\leq \frac{1}{\eta^2} \sum_j \mathbf{E}_n[U_{n,j}^2 \mathbf{I}_{\{|U_{n,j}| \geq \eta\}}] \stackrel{(\mathbf{L2-3})}{=} o(n^0) \end{aligned} \quad (10.36)$$

$$\text{und } \mathbf{E}_n[\sum_j \mathbf{E}_{n,j}[V_{n,j}^2 - U_{n,j}^2]] = \sum_j \mathbf{E}_n[U_{n,j}^2 \mathbf{I}_{\{|U_{n,j}| \geq \eta\}}] \stackrel{(\mathbf{L2-3})}{=} o(n^0)$$

Zu $\varepsilon > 0$ hinreichend klein wähle man nun $\eta = \frac{\varepsilon^2}{\sigma^2}$, $\sigma^2 = \limsup_n \sum_j \mathbf{E}_n U_{n,j}^2$, wobei $0 < \sigma^2 < \infty$ wegen **(L2-5)** und **(L2-2)**. Wegen **(L2-3)** gilt schließlich n

$$(1 + 1/\eta^2) \sum_j \mathbf{E}_n[U_{n,j}^2 \mathbf{I}_{\{|U_{n,j}| \geq \eta\}}] < \varepsilon,$$

und dann

$$\begin{aligned} P_n(|\sum_j U_{n,j}^2 - \mathbf{E}_{n,j} U_{n,j}^2| > 2\varepsilon) &\leq \varepsilon + P_n(|\sum_j V_{n,j}^2 - \mathbf{E}_{n,j} V_{n,j}^2| > \varepsilon) \leq \\ &\leq \varepsilon + \frac{1}{\varepsilon^2} \sum_j \mathbf{E}_n V_{n,j}^4 \leq \frac{2}{\varepsilon^2} \sum_j \mathbf{E}_n V_{n,j}^2 \leq 2\varepsilon + o(n^0) \end{aligned}$$

Es gilt

$$|\sum_j (2c(Z_{n,j}^2) + 1)Z_{n,j}^2| \leq \max_j |2c(Z_{n,j}^2) + 1| \sum_j Z_{n,j}^2.$$

Wegen (31) und **(L2-2)** folgt die Straffheit von $\sum_j Z_{n,j}^2$ unter P_n , und wegen (29) und $\lim_{u \rightarrow 0} c(u) = -1/2$ folgt $\max_j |2c(Z_{n,j}^2) + 1| = o_{P_n}(n^0)$, also (32). Aussage (A2) folgt wie in Jeganathan (1982) aus Aldous und Eagleson (1978) und Hall (1977). ///

Für spätere Zwecke zeigen wir eine auf die abhängige Situation angepaßte Variante von Lemma 2.3.2 in Rieder (1994):

Lemma 10.2.9 *Falls ein Dreiecksschema $\Lambda_{n,j}$ Bedingungen **(L2-2)**–**(L2-5)** erfüllt, so gilt*

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} \sum_{j=0}^n \mathbb{E} \left[\left| \int h^\tau s_n \Lambda_{n,j} dP_{\theta,n;j|j-1} \right|^2 \right] = \\ &= \lim_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \sum_{j=0}^n \mathbb{E} \left[\left| \int h^\tau s_n \Lambda_{n,j} dP_{\theta,n;j|j-1} \right| \right] \end{aligned} \quad (10.37)$$

und das zentrierte Dreiecksschema

$$\tilde{\Lambda}_{n,j} := \Lambda_{n,j} - \mathbb{E}[\Lambda_{n,j} | \mathcal{F}_{n,j-1}] \quad (10.38)$$

erfüllt **(L2-1)**–**(L2-5)**.

BEWEIS: Wir verwenden die Notation vom Beweis zu Theorem 10.2.7 sowie

$$U_{n;j|j-1} := \mathbb{E}[U_{n,j} | \mathcal{F}_{n,j-1}] \quad (10.39)$$

und erhalten

$$U_{n;j|j-1} = - \int \left(\sqrt{dQ_{n;j|j-1}} - \sqrt{dP_{j|j-1}} \right)^2 - 2 \langle D_{n,j} | \sqrt{dP_{j|j-1}} \rangle \quad (10.40)$$

Mit der Dreiecksungleichung folgt

$$\begin{aligned} &\left(\left| \sum_j U_{n;j|j-1}^2 \right|^{\frac{1}{2}} - \left| \sum_j \left(\int [\sqrt{dQ_{n;j|j-1}} - \sqrt{dP_{j|j-1}}]^2 \right)^{\frac{1}{2}} \right|^{\frac{1}{2}} \right)^2 \leq \\ &\leq \sum_j \left| \int [\sqrt{dQ_{n;j|j-1}} - \sqrt{dP_{j|j-1}}]^2 - U_{n;j|j-1} \right|^2 = \\ &= 4 \sum_j \langle D_{n,j} | \sqrt{dP_{j|j-1}} \rangle^2 \leq 4 \sum_j \|D_{n,j}\|^2 \end{aligned}$$

Damit folgt nach **(L2-4)**

$$\mathbb{E}_\mu \left(\left| \sum_j U_{n;j|j-1}^2 \right|^{\frac{1}{2}} - \left| \sum_j \left(\int [\sqrt{dQ_{n;j|j-1}} - \sqrt{dP_{j|j-1}}]^2 \right)^{\frac{1}{2}} \right|^{\frac{1}{2}} \right)^2 = o(n^0),$$

und insbesondere

$$\sum_j \mathbb{E}_\mu U_{n;j|j-1}^2 = \sum_j \mathbb{E}_\mu \left(\int [\sqrt{dQ_{n;j|j-1}} - \sqrt{dP_{j|j-1}}]^2 \right) + o(n^0).$$

Aber nach **(L2-4)** und **(L2-2)** gilt

$$\sum_j \mathbb{E}_\mu \left(\int [\sqrt{dQ_{n;j|j-1}} - \sqrt{dP_{j|j-1}}]^2 \right) \leq \sum_j \mathbb{E}_\mu 2D_{n,j}^2 + \frac{1}{2} U_{n,j}^2 = o(n^0) + O(n^0)$$

und nach **(L2-4)** und **(L2-3)** gilt

$$\max_j \mathbb{E}_\mu \left(\int [\sqrt{dQ_{n;j|j-1}} - \sqrt{dP_{j|j-1}}]^2 \right) \leq 2 \sum_j \mathbb{E}_\mu D_{n,j}^2 + \frac{1}{2} \max_j \mathbb{E}_\mu U_{n,j}^2 = o(n^0),$$

da die Lindebergbedingung **(L2-3)** die Fellerbedingung $\gamma_n := \max_j \mathbb{E}_\mu U_{n,j}^2 = o(n^0)$ impliziert. Damit gilt die erste Hälfte von (37), denn

$$\sum_j A_j^2 \leq \max A_j \sum A_j;$$

die zweite folgt mit Cauchy–Schwartz aus der ersten.

Die Eigenschaften **(L2-2)**–**(L2-4)** für $\tilde{\Lambda}_{n,j}$ folgen nun wie bei Rieder (1994); für **(L2-5)** notieren wir, wobei wir wieder den Index n der Filtrierung weglassen, daß

$$\mathbb{E}[s_n \tilde{\Lambda}_{n,j} \tilde{\Lambda}_{n,j}^\tau | \mathcal{F}_{j-1}] = \mathbb{E}[s_n \Lambda_{n,j} \Lambda_{n,j}^\tau | \mathcal{F}_{j-1}] - \mathbb{E}[s_n \Lambda_{n,j} | \mathcal{F}_{j-1}] \mathbb{E}[\Lambda_{n,j}^\tau | \mathcal{F}_{j-1}].$$

Summieren wir über j , so verschwindet der Subtrahend asymptotisch in L_1 nach (37), also gilt insbesondere

$$s_n \sum_{j=1}^n \mathbb{E} \left[\tilde{\Lambda}_{n,j} \tilde{\Lambda}_{n,j}^\tau | \mathcal{F}_{n,j-1} \right] s_n = T + o_{P_n}(n^0)$$

////

Als nächstes wollen wir den Begriff der L_2 -Differenzierbarkeit für Modelle mit Abhängigkeit dieses Abschnitts in Verbindung mit dem der L_2 -Differenzierbarkeit bei größeren σ -Algebren aus Abschnitt 9.4.3 zusammenführen, um den Begriff der L_2 -Differenzierbarkeit längs Subfiltrierungen zu bekommen. Dazu notieren wir folgende Proposition, bei der wir wie üblich $\theta \in \Theta$ fixieren, und alle Erwartungswerte unter $P_{\theta,n}$ auswerten:

Theorem 10.2.10 *Sei $\mathcal{P}_n = \mathcal{P}_n^{\mathcal{F}}$ ein parametrisches Modell gemäß (12), das längs der Filtrierungsfolge $(\mathcal{F}_{n,j})_n$ L_2 -differenzierbar in $\theta \in \Theta$ mit Ableitung $\Lambda_{n,j}$ und Fisher-Information \mathcal{I} ist.*

Sei $(\mathcal{G}_{n,j})_n$ eine Subfiltrierungsfolge von $(\mathcal{F}_{n,j})_n$ in dem Sinn, daß für alle n und alle $j \leq n$ $\mathcal{G}_{n,j} \subset \mathcal{F}_{n,j}$. Gelten dann mit den Bezeichnungen aus Theorem 10.2.7

$$s_n = O\left(\frac{1}{\sqrt{n}}\right) \tag{10.41}$$

$$\Lambda_{n,j} \text{ sind unter } P_n \text{ gleichgradig quadratintegrierbar} \tag{10.42}$$

$$\sum_{j=1}^n \mathbb{E} s_n \left[\mathbb{E}[\Lambda_{n,j} | \mathcal{G}_{n,j}] \mathbb{E}[\Lambda_{n,j}^\tau | \mathcal{G}_{n,j}] \Big| \mathcal{G}_{n,j-1} \right] s_n = \tilde{T} + o_{P_n}(n^0) \succ 0 \tag{10.43}$$

$$\forall h \in \mathbb{R}^k : n \sup_j \mathbb{E}_\mu \int D_{n,j}^2 = o(n^0) \tag{10.44}$$

so ist auch das Modell $\mathcal{P}^{\mathcal{G}}$ L_2 -differenzierbar in θ mit Ableitung

$$\Lambda_{n,j}^{\mathcal{G}} := \mathbb{E}[\Lambda_{n,j} | \mathcal{G}_{n,j}] - \mathbb{E}[\Lambda_{n,j} | \mathcal{G}_{n,j-1}] \quad (10.45)$$

und Fisher-Information

$$\mathcal{I}^{\mathcal{G}} = \sum_j s_n \mathbb{E}[\Lambda_{n,j}^{\mathcal{G}} \Lambda_{n,j}^{\mathcal{G}\tau} s_n] s_n \succ 0. \quad (10.46)$$

BEWEIS: Wir zeigen zunächst **(L2-2)**–**(L2-5)** für $L_{n,j} := \mathbb{E}[\Lambda_{n,j} | \mathcal{G}_{n,j}]$ und schließen dann mit Lemma 10.2.9 auf die Behauptung.

Die gleichgradige quadratische Integrierbarkeit der $\Lambda_{n,j}$ überträgt sich auf die bedingten Erwartungswerte $L_{n,j} := \mathbb{E}[\Lambda_{n,j} | \mathcal{G}_{n,j}]$, wie man mit dem Vallée-Poussin-Kriterium (iii) aus Proposition A.4.1 zeigen kann:

Wegen Korollar A.4.2 genügt es, die gleichgradige Integrierbarkeit einer Majorante nachzuweisen. Diese findet sich in $M_{n,j} := \mathbb{E}[|\Lambda_{n,j}|^2 | \mathcal{G}_{n,j}]$, weil

$$|\mathbb{E}[\Lambda_{n,j} | \mathcal{G}_{n,j}]|^2 \leq \mathbb{E}[|\Lambda_{n,j}|^2 | \mathcal{G}_{n,j}].$$

Da $\Lambda_{n,j}$ gleichgradig quadratintegrierbar ist, existiert nach dem Kriterium von Vallée-Poussin eine konvexe Funktion q von \mathbb{R}_+ nach \mathbb{R}_+ mit $\lim_{t \rightarrow \infty} q(t)/t = \infty$ und $\sup \mathbb{E}_{\mu} q(|\Lambda_{n,j}|^2) < \infty$. Mit demselben q gilt wegen Jensen

$$\sup \mathbb{E}_{\mu}[q(M_{n,j})] \leq \sup \mathbb{E}_{\mu}[q(|\Lambda_{n,j}|^2)] < \infty$$

und somit die behauptete gleichgradige Integrierbarkeit.

Damit aber folgen **(L2-2)** und **(L2-3)**.

(L2-5) haben wir mit (43) explizit gefordert. **(L2-4)** schließlich ergibt sich aus Theorem 9.3.2 und der Gleichmäßigkeit (44). ////

10.2.4 Umsetzung der allgemeinen asymptotischen Theorie

Wie im Abschnitt 9.4.3 wollen wir nun die Begriffe *Influenzkurve* und *asymptotisch linearer Schätzer* auch für die Situation von Martingal-Differenzen-Schemata einführen.

Definition 10.2.11 Sei $\mathcal{P}_n = \mathcal{P}_n^{\mathcal{F}}$ ein parametrisches Modell gemäß (12), das längs der Filtrierungsfolge $(\mathcal{F}_{n,j})_n$ L_2 -differenzierbar in $\theta \in \Theta$ mit Ableitung $\Lambda_{n,j}$, und es gelte

$$\lim_n \sum_{j=0}^n s_n \mathbb{E}[\Lambda_{n,j} \Lambda_{n,j}^{\tau}] s_n = \mathcal{I} \quad (10.47)$$

mit \mathcal{I} der Fisher-Information.

Weiter sei $D \in \mathbb{R}^{p \times k}$, $p \leq k$, $\text{rk } D = p$. Dann nennen wir jedes Schema $\psi = (\psi_{n,j})$, $\psi_{n,j} \in L_2^p(P_{\theta;n,j})$ mit

$$\mathbf{(L2-1)} \quad \mathbb{E}[\psi_{n,j} | \mathcal{F}_{n,j-1}] = 0$$

$$\mathbf{(L2-2)} \quad \sum_{j=1}^n s_n \mathbb{E}[\psi_{n,j} \psi_{n,j}^{\tau}] s_n \xrightarrow{n \rightarrow \infty} \mathcal{C}(\psi) \succ 0$$

$$\mathbf{(L2-3)} \quad \sum_{j=1}^n \mathbb{E}[|h s_n \psi_{n,j}|^2 \mathbf{1}_{\{|h s_n \psi_{n,j}| > \varepsilon\}}] \rightarrow 0 \xrightarrow{n \rightarrow \infty} \quad \forall h \in \mathbb{R}^k$$

(corr) $\sum_{j=1}^n s_n \mathbb{E}[\psi_{n,j} \Lambda_{n,j}^\tau | s_n] \xrightarrow{n \rightarrow \infty} \mathbb{I}_k$ bzw. D

[partielle] Influenzkurve in θ . Die Menge aller [partiellen] Influenzkurven in θ heie $\Psi_2(\theta)$ resp. $\Psi_2^D(\theta)$.

Definition 10.2.12 Sei $\tau : \Theta \rightarrow \mathbb{R}^p$ eine in $\theta \in \Theta$ differenzierbare Abbildung mit $D = d\tau|_\theta$, $\text{rk } D = p$. Ein asymptotischer Schtzer $S = (S_n)$ fr $\tau(\theta)$, $S_n : (\Omega, \mathcal{F}_n) \rightarrow (\mathbb{R}^p, \mathbb{B}^p)$ heie asymptotisch linear in $P_{\theta,n}$, falls es eine partielle Influenzkurve $\psi_\theta \in \Psi_2^D(\theta)$ gibt, so da

$$s_n^{-1}(S_n - \tau(\theta)) = s_n \sum_{i=1}^n \psi_{\theta;n,j} + o_{P_{\theta,n}}(n^0) \quad (10.48)$$

Wie in Lemma 9.4.7 ergibt sich Bedingung **(corr)** in Definition 10.2.11 als Konsequenz der lokalen Fisher-Konsistenz bzw. der lokal uniformen asymptotischen Normalitt:

Lemma 10.2.13 Sei $S = (S_n)$ ein as. Schtzer fr $\tau(\theta)$ mit Entwicklung (48) fr ein Schema $\psi_{\theta;n,j}$ mit **(L2-1)**, **(L2-2)** und **(L2-3)**. Dann gilt **(corr)** genau dann, wenn fr alle $h \in \mathbb{R}^k$

$$s_n^{-1}(S_n - \tau(\theta + s_n h)) \circ P_{\theta,n} \xrightarrow{w} \mathcal{N}(0, \mathcal{C}_\theta(\psi_\theta)) \quad (10.49)$$

BEWEIS: Rieder (1994), Lemma 4.2.18., nur da wir nun in der Argumentation den Zentralen Grenzwertsatz fr Martingaldifferenzen von McLeish (1974) verwenden. ////

Bemerkung 10.2.14 In Lemma 10.2.13 ist die Martingaldifferenz-Eigenschaft **(L2-1)** aus Definition 10.2.11, unverzichtbar; sie lst den scheinbaren Widerspruch, da bei der Schtzung der Hyperparameter im Zustandsraummodell mit dem dort betrachteten Momentenschtzer (vgl. (143)–(146)) ein lokal gleichmig asymptotisch normalverteilter Schtzer existiert, der ‘‘asymptotisch linear’’ und adaptiert zur Filtrierung $\mathcal{F}_{n,j} = \sigma(Y_1, \dots, Y_j)$ ist, obwohl die entsprechende Modell-Fisher-Information singularr ist. Der Grund hierfr ist, da dieser Momentenschtzer bezglich der Filtrierung $\mathcal{F}_{n,j} = \sigma(Y_1, \dots, Y_j)$ kein Martingal-Differenzschema mehr ist!

Mit diesen Begriffen stehen starke Optimalittsaussagen zur Verfgung. Dazu wollen wir zwei Influenzkurven ψ , $\tilde{\psi}$ identifizieren — und in Zeichen $\psi = \tilde{\psi}$ schreiben, falls

$$\sum_i \mathbb{E} |s_n(\psi_{n,i} - \tilde{\psi}_{n,i})|^2 = o(n^0). \quad (10.50)$$

Zunchst zeigen wir folgende Verallgemeinerung von Rieder (1994), Proposition 4.2.20.

Proposition 10.2.15 Sei $\mathcal{P}_n = \mathcal{P}_n^{\mathcal{F}}$ ein parametrisches Modell gem (12), das lngs der Filtrierungsfolge $(\mathcal{F}_{n,j})_n$ L_2 -differenzierbar in $\theta \in \Theta$ mit Ableitung $\Lambda_{n,j}$, und es gelte (47) mit Fisher-Information \mathcal{I} .

Sei $S = (S_n)$ ein in $P_{\theta,n}$ asymptotisch linearer Schtzer mit Influenzkurve $\psi_\theta \in \Psi_2^D(\theta)$. Dann gilt — wobei wir den Index θ weglassen

(a)

$$\mathcal{C}(\psi) \succeq DID^\tau \quad (10.51)$$

mit “=” $\iff \psi = \hat{\psi}$ für

$$\hat{\psi}_{n,i} = DI^{-1}\Lambda_{n,i} \quad (10.52)$$

(b) $s_n^{-1}(S_n - \tau(\theta))$ ist regulär für $D = d\tau|_\theta$ mit normaler Grenzverteilung

$$\mathcal{N}(0, \mathcal{C}(\psi) = \mathcal{N}(0, \mathcal{C}(X)) * \mathcal{N}(0, \mathcal{C}(\hat{\psi})) \quad (10.53)$$

mit X unabhängig zu $\hat{\psi}$ und $\mathcal{L}(X) = \text{w-lim } \mathcal{L}(\sum_i s_n(\rho_{n,i} - \hat{\psi}_{n,i}))$. Die Grenzverteilung $\mathcal{N}(0, \mathcal{C}(\hat{\psi}))$ wird dann und nur dann angenommen, wenn $\psi = \hat{\psi}$.(c) Sei die zugrundeliegende Verlustfunktion³ $\ell \in \mathbf{L}$, ℓ stetig λ^p -f.s.. Dann ist

$$\begin{aligned} & \lim_{b \rightarrow \infty} \lim_{c \rightarrow \infty} \lim_{n \rightarrow \infty} \sup_{t \in \Theta, |t| \leq c} \int b \wedge \ell(s_n^{-1}[S_n - \tau(\theta + s_n t)]) dP_{\theta + s_n t, n} = \\ & = \int \ell d\mathcal{N}(0, \mathcal{C}_\theta(\psi_\theta)) \geq \int \ell d\mathcal{N}(0, \mathcal{C}_\theta(\hat{\psi}_\theta)) \end{aligned} \quad (10.54)$$

und die untere Schranke wird angenommen, wenn $\psi_\theta = \hat{\psi}_\theta$. Ist $\ell \in \mathbf{L}$ monoton quadratisch⁴, so kann die untere Schranke nur durch $\psi_\theta = \hat{\psi}_\theta$ angenommen werden.

BEWEIS: Zunächst gilt

$$\begin{aligned} & \text{Cov}\left[\sum_i s_n(\psi_{n,i} - \hat{\psi}_{n,i})\right] \stackrel{(\mathbf{L2-1})}{=} \sum_i s_n \text{Cov}[\psi_{n,i} - \hat{\psi}_{n,i}] s_n = \\ & = \sum_i s_n \{ \text{E}[\psi_{n,i} \psi_{n,i}^\tau] - \text{E}[\psi_{n,i} \hat{\psi}_{n,i}^\tau] - \text{E}[\hat{\psi}_{n,i} \psi_{n,i}^\tau] + \text{E}[\hat{\psi}_{n,i} \hat{\psi}_{n,i}^\tau] \} s_n = \\ & \stackrel{(\text{corr})}{=} o(n^0) + \mathcal{C}(\psi) - DI^{-1}D = \mathcal{C}(\psi) - \mathcal{C}(\hat{\psi}), \end{aligned}$$

also (a), (b) und (c) folgen wortwörtlich wie in Rieder (1994), Proposition 4.2.20.

////

Bemerkung 10.2.16 Auch ein Analogon zu Proposition 9.4.9 gilt im Kontext dieses Abschnitts.

10.3 LAN-Eigenschaft für lineare, zeitinvariante Zustandsraummodelle

10.3.1 Skizze der Vorgehensweise

Die im vorigen Abschnitt eingeführte L_2 -Differenzierbarkeit bei Abhängigkeitsstrukturen werden wir nun kombinieren mit der Übertragung der L_2 -Differenzierbarkeit auf gröbere σ -Algebren, um für das lineare Zustandsraummodell, parametrisiert durch die Hyper-Parameter F , Q , V , eine LAN-Eigenschaft herzuleiten. Dabei behandeln wir die unbeobachtbaren Zustände als Missings und zeigen nur die L_2 -Differenzierbarkeit des “vollen” Modells, bei dem

³vgl. Definition A.5.7⁴vgl. Proposition 9.4.8

auch die Zustände selbst zur Schätzung zur Verfügung stehen und schließen dann mit dem Übertragungsargument auf das Modell, bei dem wirklich nur die Y_s verfügbar sind.

10.3.2 Modell

Sei folgendes Modell gemäß (V1) gegeben

$$\beta_0 \sim G_0 \quad (10.55)$$

$$\beta_t = F\beta_{t-1} + \tilde{Q}v_t \quad v_t \sim G, \quad \mathbb{E}_G v_t v_t^\tau = \mathbb{I}_p \quad (10.56)$$

$$y_t = Z\beta_t + \tilde{V}\varepsilon_t \quad \varepsilon_t \sim H, \quad \mathbb{E}_H \varepsilon_t \varepsilon_t^\tau = \mathbb{I}_q \quad (10.57)$$

wobei die Verteilungen $G_0, G \in \mathcal{M}_1(\mathbb{B}^p)$ und $H \in \mathcal{M}_1(\mathbb{B}^q)$, sowie die Matrix $Z \in \mathbb{R}^{q \times p}$ bekannt sind, und die v_s unabhängig identisch verteilt sind, genauso die ε_t , und die Prozesse der v_s der ε_t sowie β_0 stochastisch unabhängig sind. Dabei sind die Matrizen $F \in \mathbb{R}^{p \times p}$, $0 \prec \tilde{Q} = \tilde{Q}^\tau \in \mathbb{R}^{p \times p}$, $0 \prec \tilde{V} = \tilde{V}^\tau \in \mathbb{R}^{q \times q}$ unbekannt und aus den Beobachtungen Y_t , $t \geq 0$ zu schätzen. Dazu fassen wir diese Größen mit den in (B.1) eingeführten Operatoren $[\text{s}] \text{vec}[\cdot]$ zusammen zu $\theta = (\text{vec}[F]^\tau, \text{svec}[\tilde{Q}]^\tau, \text{svec}[\tilde{V}]^\tau)^\tau \in \mathbb{R}^{p^2 + \binom{p}{2} + \binom{q}{2}}$.

Bemerkung 10.3.1 Daß wir den Hyper-Parameter Z nicht mit in den zu schätzenden Parameter θ integrieren hat folgenden Grund: Wenn wir nur die Beobachtungen Y_t zur Verfügung haben, können wir nicht unterscheiden zwischen (Z, \tilde{Q}) und $(\alpha Z, (\tilde{Q}/\alpha^2))$ mit $0 \neq \alpha \in \mathbb{R}$.

10.3.3 Glättungs-Filtrierungen

Im Unterschied zum Fall unabhängiger Beobachtungen kann es unter Umständen erforderlich sein, “Glättungs-Filtrierungen” zu verwenden. Bei einer “Glättungs-Filtrierung” wollen wir uns zu einem gewissen Zeitpunkt s zur Inferenz auf den unbekanntem Zustand β_s nicht nur auf die Information der Beobachtungen bis einschließlich s beschränken, sondern auch zulassen, daß Beobachtungen Y_t für $t > s$ verwendet werden.

Setzen wir T als den Beobachtungshorizont der Stichprobe, so ergeben sich als wichtige Typen die “fixed-lag”-Glättung mit $t = (s + m) \wedge T$, und mit $s = T$ die feinst-mögliche Glättungsfiltrierung.

Uns soll hier der Fall eines Zustandsraummodells mit Anfangszustand $\beta_0 \equiv 0$ interessieren. Dazu betrachten wir für jeden Beobachtungshorizont T eine Folge $k_{T,s}$ mit $s \leq k_{T,s} \leq T$ und $k_{T,s} \leq k_{T,s+1}$, anhand deren wir die Filtrierung \mathcal{F}_T definieren als

$$\mathcal{F}_{T,s,k_{T,s}} := \sigma(\beta_{0:s}, Y_{1:k_{T,s}}), \quad \mathcal{F}_T = (\mathcal{F}_{T;s,k_{T,s}})_{s \leq T}. \quad (10.58)$$

Wenn aus dem Zusammenhang klar, lassen wir den Index T von $k_{T,s}$ weg. Dies ergibt in der Situation, in der wir die Zustände β_s nicht kennen, als Filtrierungsfolge $\mathcal{F}_{T,0,k_{T,s}}$ mit $k_{T,s} = s$, für eine nur auf gefilterten Werten basierende Filtrierung, $k_{T,s} = (s + m) \wedge T$ für eine “fixed-lag”-Glättung und $k_{T,s} = T$ für die feinst-mögliche Glättungsfiltrierung.

Dabei besteht zunächst die konzeptionelle Schwierigkeit, einzusehen, daß es sich bei $(\mathcal{F}_{T,0,k_{T,s}})_s$ um Unter- σ -Algebren von $(\mathcal{F}_{T,s,s})_s$ handelt, obwohl unter

Umständen $k_{T,s} > s$ ist. Dies machen wir uns an einer Folge unabhängig identisch verteilter, reellwertiger Beobachtungen $Y_i \sim P_\theta$ aus einem parametrischen Modell $\mathcal{P} = \{P_\theta \mid \theta \in \mathbb{R}\}$ klar:

Sei $\nu \in M_\sigma(\mathbb{B})$, $\nu \gg \mathcal{P}$ und p_θ eine ν -Dichte von P_θ . Wir nutzen aus, daß unabhängig von einer möglicherweise schon fortgeschritteneren Kenntnis der Y_t , $t > s$, die σ -Algebra $\sigma(\{Y_r\}_{r \leq s})$ bereits suffizient für die Bestimmung der Likelihood im Zeitpunkt s ist.

Faßt man die Likelihood $p_\theta(y)$ in der Beobachtung $Y = y$ auf als faktorisierten, bedingten Erwartungswert $E_\nu[p_\theta(Y) \mid Y = y]$, so ist klar, daß wir dies im unabhängigen Fall mit n Beobachtungen auch ν^n -fast-sicher schreiben können als

$$p_{\theta,n}(y_{1:n}) = \prod_{j=1}^n p_\theta(y_j) = \prod_{j=1}^n E_\nu[p_\theta(Y_j) \mid Y_{1:n} = y_{1:n}],$$

wobei wir den Vektor (Y_1, \dots, Y_n) mit $Y_{1:n}$ abgekürzt haben. Entsprechendes gilt im Zustandsraum.

Sind nun aber nur die Y_t beobachtbar, kommt es auf diese unterschiedlichen Filtrierungen an, was wir im folgenden untersuchen wollen. Dazu führen wir zusätzlich die faktorisierte Unter- σ -Algebra $\bar{\mathcal{F}}_{0,T} = \sigma(Y_{1:T} = y_{1:T})$ ein und nehmen der Einfachheit halber an, das Zustandsraummodell sei durch eine Folge von Produktmaßen $\nu_T = (\mu_1 \otimes \mu_2)^T$ dominiert. Dann ergibt sich jeweils ν_T -fast sicher

$$\begin{aligned} p_{\theta;T}^y(y_{1:T}) &= E_{\nu_T}[p_{\theta;T}^y(\beta_{0:T}, Y_{1:T}) \mid \bar{\mathcal{F}}_{0,T}] = \\ &= E_{\nu_T}\left[\prod_{s=1}^T \frac{p_{\theta;T}^{\beta,y}(\beta_{0:s}, Y_{1:k_s})}{p_{\theta;T}^{\beta,y}(\beta_{1:s-1}, Y_{1:k_{s-1}})} \mid \bar{\mathcal{F}}_{0,T}\right] = \\ &= E_{\nu_T}\left[\prod_{s=1}^T p_{\theta;T}^{\beta,y}(\beta_{0:s}, Y_{1:k_s} \mid \beta_{0:s-1}, Y_{1:k_{s-1}}) \mid \bar{\mathcal{F}}_{0,T}\right] = \\ &=: E_{\nu_T}\left[\prod_{s=1}^T p_{s|s-1;k_s|k_{s-1}}^{\beta,y} \mid \bar{\mathcal{F}}_{0,T}\right] = \tag{10.59} \\ &= \int \prod_{s=1}^T p_{s|s-1;k_s|k_{s-1}}^{\beta,y}(\beta_{0:s}, y_{1:k_s}) d\mu_1^T(d\beta_{1:T}) \end{aligned}$$

Es ist unmittelbar aus der Definition von $p_{\theta;T}^y(y_1, \dots, y_T)$ klar, daß sich die Wahl der Filtrierung nicht auf die Funktion der Likelihood auswirkt. Andererseits werden wir sehen, daß sich bei der Herleitung der lokalen Asymptotik der Likelihood durchaus Unterschiede ergeben können, und zwar weil zum Beispiel $p_{\theta;s|s-1;s|s-1}^{\beta,y}$ nicht "genügend" Information über θ trägt, in dem Sinn, daß die entsprechende Fisher-Information singular wird, wohingegen zum Beispiel für $p_{\theta;s|s-1;s+2|s+1}^{\beta,y}$ kein Rangdefekt mehr auftritt — vgl. Proposition 10.5.1.

Ist die Fisherinformation des Modells längs einer Filtrierung \mathcal{F} regulär, so fallen wegen der dann bestehenden Eindeutigkeit der as. suffizienten Statistik — vgl. Bemerkung A.5.3 — asymptotisch alle hergeleiteten Loglikelihood-Entwicklungen längs feineren Filtrierungen $\mathcal{F}'_T \supset \mathcal{F}_T$ zusammen.

Unter noch zu spezifizierenden Rangvoraussetzungen an die Matrizen F , Z , Q und V sowie einer Voraussetzung an $\rho(F)$, den Betrag des betragsmäßig

größten Eigenvektors, wird sich ergeben, daß dann bereits eine Filtrierung mit $k_{T,s} = (s+2) \wedge T$ eine reguläre Fisher-Information liefert.

Daher wird es nicht notwendig sein, bis zum Beobachtungshorizont T zu glätten, wie dies zum Beispiel in der Arbeit von Shumway und Stoffer (1982) geschieht. Wir weisen an dieser Stelle darauf hin, daß die Domination bei obiger Erläuterung nur der vereinfachten Schreibweise diene; sie ist nicht notwendig und wird dementsprechend auch nicht gefordert; im allgemeinen werden wir daher eine entsprechend angepaßte Notation analog zum Wurzeldichtenkalkül für Martingaldifferenzen aus Abschnitt 10.2.1 verwenden.

10.3.4 die LAN-Eigenschaft

Theorem 10.3.2 *Sei der Parameter $\theta = (F, \tilde{Q}, \tilde{V})$ aus dem Parameterbereich*

$$\Theta := \{(A, B, C) \in \mathbb{R}^{p \times p} \times \mathbb{R}^{p \times p} \times \mathbb{R}^{q \times q} \mid \rho(A) < 1, B = B^\tau \succ 0, C = C^\tau \succ 0\},$$

mit $\rho(M)$ dem Betrag des betragsmaximalen Eigenwertes von M .

Besitzt dann G endliche [multivariate] Lokations-Fisher-Information und besitzen G und H endliche [multivariate] Skalen-Fisher-Information und besitzt die Fisher-Information des Modells vollen Rang, wenn wir eine Filtrierungsfolge $\mathcal{G}_T = (\mathcal{F}_{T, k_{T,s}, 0})$ gemäß Abschnitt 10.3.3, also mit einem $k_{T,s}$ mit den dort eingeführten Eigenschaften, verwenden, so ist das Modell (55) bis (57), L_2 -differenzierbar im Sinn von Definition 10.2.5 und besitzt demzufolge die LAN-Eigenschaft.

Bemerkung 10.3.3 (a) Wie im Zeitreihenkontext üblich, ersetzen wir die Variable n durch T , um dadurch den Beobachtungshorizont anzugeben.

(b) Im Lichte von Theorem B.2.4 ist die Endlichkeit der Fisher-Informationen gleichbedeutend damit, daß G und H p - bzw. q -dimensionale Lebesgue-Dichten g und h , und $g, v; g(v)$, $i = 1, \dots, p$ und $\varepsilon_j h(\varepsilon)$, $j = 1, \dots, q$ Distributions-Gradienten ∇g , $(g_r \mathbf{1}_{r=i} + (\nabla g)_r)_r$, $(h_s \mathbf{1}_{s=j} + (\nabla h)_s)_s$ besitzen. Die multivariate Lokations-Fisher-Information lautet dann

$$\mathcal{I}_g := E_G[\nabla g \nabla g^\tau / g^2], \quad (10.60)$$

die multivariaten Skalen-Fisher-Informationen

$$\mathcal{J}_g := E_G \left[\text{svec}[\mathbb{I}_p + v \nabla g^\tau / g] \text{svec}[\dots]^\tau \right], \quad (10.61)$$

$$\mathcal{J}_h := E_H \left[\text{svec}[\mathbb{I}_q + \varepsilon \nabla h^\tau / h] \text{svec}[\dots]^\tau \right], \quad (10.62)$$

mit $\text{svec}[\cdot]$, dem in (B.1) eingeführten Operator.

(c) Garel und Hallin (1995) geben Bedingungen für die LAN-Eigenschaft von multivariaten ARMA-Prozessen an; dies ist relevant, da es sich bei der Zustandsgleichung (56) um ein multivariates AR(1)-Modell handelt. Allerdings ist ihre Bedingung (C3), nämlich daß $\int |\nabla g|^4 dG$ endlich ist, in unserer Darstellung abgeschwächt auf die Endlichkeit von $\int |\nabla g|^2 dG$.

Ganz generell läßt sich die LAN-Eigenschaft im ARMA-Modell bereits unter Endlichkeit der multivariaten Lokations-Fisher-Information und den üblichen Stationaritätsbedingungen (vgl. Brockwell und Davis (1991), Theorem 11.3.1 und Theorem 11.3.2.) an die Parameter zeigen.

(d) Wang (1994) gibt ebenfalls Bedingungen für die LAN-Eigenschaft von multivariaten ARMA-Prozessen an; er startet jedoch bereits mit der L_2 -Differenzierbarkeits-Eigenschaft **(L2-4)**.

(e) Bickel et al. (1998) beschäftigen sich mit der LAN-Eigenschaft für allgemeiner parametrisierte Hidden-Markov-Models (HMM's), gehen also von einem endlichen (oder zumindest diskreten und kompakten) Zustandsraum aus, und Jensen und Petersen (1999) verallgemeinern die dort getroffenen Aussagen auf allgemeine, aber kompakte Zustandsräume.

Bei uns hingegen ist der Zustandsraum i.a. nicht kompakt, und wir stellen hier auf eine spezielle Parametrisierung ab, so daß wir deren Annahmen über die Übergangsdichten erheblich abschwächen können.

(f) Die Bedingung, die Fisherinformation möge vollen Rang haben, ist im eindimensionalen, normalen Zustandsraummodell automatisch erfüllbar, vgl. Proposition 10.4.6, sowie Proposition 10.5.1.

(g) Wie schon erwähnt, fallen wegen der Eindeutigkeit der as. suffizienten Statistik Loglikelihood-Entwicklungen längs jeder feineren Filtrierung $\mathcal{G}'_T \supset \mathcal{G}_T$ mit der unter \mathcal{G}_T zusammen, so daß die LAN-Eigenschaft nur noch über die größte Subfiltrierung $\tilde{\mathcal{G}}_T \subset \mathcal{G}_T$, längs der die Fisher-Information vollen Rang besitzt, von der Filtrierung abhängt, wohingegen die L_2 -Ableitungen sehr wohl von der Filtrierung abhängen.

BEWEIS VON THEOREM 10.3.2: Wir beweisen zunächst die L_2 -Differenzierbarkeit des Modells bei Beobachtung von (β_s, y_t) . Dabei ersetzen wir, wie bereits in Bemerkung 10.3.3 angedeutet, die Variable n durch T ; dementsprechend sind die betrachteten Alternativenfolgen von der Gestalt $\theta_T = \theta_0 + h/\sqrt{T}$. Ebenso verwenden wir s anstelle von j als Laufindex.

Der Prozeß der $\{\beta_t\}$, und somit auch der der Paare $\{\beta_t, y_t\}$ ist asymptotisch stationär. Genauer gibt es wegen $\rho(F) < 1$ einen strikt stationären und ergodischen Prozeß $\{\tilde{\beta}_t\}$ mit $\|\tilde{\beta}_t - \beta_t\|_{L_2(P_\theta)} = O(\rho(F)^t)$ und $\|\tilde{\beta}_t - \beta_t\| = O(\rho(F)^t)$ G^∞ -f.s., vgl. Brockwell und Davis (1991), Proposition 3.1.1 und 3.1.2 bzw. Theorem 11.3.1 für die Konvergenz in L_2 und fast sicher und Lemma A.3.2 für die Stationarität und Ergodizität des Limes-Elementes $\{\tilde{\beta}_t\}$.

Wie im Eingangsabschnitt dieses Kapitels kürzen wir den Vektor der Zufallsgrößen $(\beta_0, \beta_1, \dots, \beta_t)$ als $\beta_{0:t}$ und analog den der (y_1, \dots, y_t) als $y_{1:t}$ ab. Weiterhin verwenden wir die Filtrierungsnotation (58). Damit schreibt sich die Loglikelihood als

$$\begin{aligned} \log \frac{dP_{\theta+\Delta;T}}{dP_{\theta;T}}(\beta_{0:T}, y_{1:T}) &= \sum_{s=1}^T \log g(\tilde{Q}^{-1}(\beta_s - F\beta_{s-1})) - \log \det \tilde{Q} + \\ &\quad + \log h(\tilde{V}^{-1}(y_s - Z\beta_s)) - \log \det \tilde{V} \end{aligned}$$

Als Kandidat für die L_2 -Ableitung ergibt sich:

$$\Lambda_{F,t} := \tilde{Q}^{-1} \Lambda_g(\tilde{Q}^{-1}(\beta_t - F\beta_{t-1})) \beta_{t-1}^\tau \quad (10.63)$$

$$\Lambda_{Q,t} := -\tilde{Q}^{-1} + \tilde{Q}^{-1} \Lambda_g(\tilde{Q}^{-1}(\beta_t - F\beta_{t-1})) (\tilde{Q}^{-1}(\beta_t - F\beta_{t-1}))^\tau \quad (10.64)$$

$$\Lambda_{V,t} := -\tilde{V}^{-1} + \tilde{V}^{-1} \Lambda_h(\tilde{V}^{-1}(y_t - Z\beta_t)) (\tilde{V}^{-1}(y_t - Z\beta_t))^\tau \quad (10.65)$$

Offenbar sind alle drei Komponenten für jedes t Martingaldifferenzen bezüglich der Filtrierungsfolge $\mathcal{F}_T^0 = (\mathcal{F}_{T,s,s})$ und damit, wie eingangs in Abschnitt 10.3.3

beschrieben, auch bezüglich jeder Filtrierungsfolge $\mathcal{F}_T = (\mathcal{F}_{T,s,k_{T,s}})$ mit einem $k_{T,s}$ mit den dort eingeführten Eigenschaften. Somit gilt **(L2-1)**.

Die asymptotische Stationarität der $\{\beta_t, y_t\}$ überträgt sich auf die der L_2 -Ableitungen und somit folgen die Bedingungen **(L2-2)** und **(L2-3)**, die im wesentlichen auf gleichgradige Integrierbarkeit hinauslaufen, wie im eindimensionalen Fall, vgl. Kreiss (1987).

Aus der asymptotischen Ergodizität aller drei Komponenten schließlich folgt **(L2-5)**.

Für **(L2-4)** ist die Endlichkeit von $\bar{\mathcal{I}}_g, \bar{\mathcal{J}}_g, \bar{\mathcal{J}}_h$ hinreichend: Wegen Theorem B.3.4 ist damit das Lokations-/Skalenmodell mit G als Zentralverteilung bzw. das Skalenmodell mit H als Zentralverteilung L_2 -differenzierbar, was wie im Eindimensionalen **(L2-4)** für den Parameter $\theta = (F, Q, V)$ impliziert — vgl. hierzu die Mitschätzung eines Skalenparameters im [univariaten] AR(p)-Modell in Swensen (1980).

Nun zum Fall, daß nur die y_t zur Schätzung zur Verfügung stehen. Wir fixieren zum Beobachtungshorizont T eine Filtrierung $\mathcal{F}_T = (\mathcal{F}_{T,s,k_{T,s}})$ mit $k_{T,s} = (s+m) \wedge T$ für ein $m \in \mathbb{N}_0$. Zu dieser betrachten wir die Subfiltrierung $\mathcal{G}_T = (\mathcal{F}_{T,0,k_{T,s}})$ und zeigen die Voraussetzungen von Theorem 10.2.10.

(41) ist nach Definition $s_n = 1/\sqrt{n}$ erfüllt. Die gleichgradige Quadratintegrierbarkeit (42) folgt aus der L_2 -Approximierbarkeit von $\Lambda_{n,t}$ durch eine stationäre und ergodische Funktion Λ_t .

Wegen der Ergodizität der Erzeugendenfolge (v_t, ε_t) gilt dies auch für die (asymptotisch) stationäre Folge der unbedingten L_2 -Ableitungen, und, weil auch die bedingende Folge der $\{y_t\}$ unter den Voraussetzungen des Theorems (asymptotisch) stationär und ergodisch ist, gilt dies nach Lemma A.3.3 auch für die Folge der bedingten L_2 -Ableitungen, und somit bleibt auch **(L2-5)** gültig, da wir die Regularität der Fisher-Information vorausgesetzt haben.

(44) schließlich ergibt sich wieder aus der (asymptotischen) Stationarität der L_2 -Ableitungen im feiner filtrierten Modell, wobei wir $\hat{\theta}$ für die Alternativenfolge und \bar{s} für $s+m$ schreiben

$$n \sup_s E_\mu \int D_{T,s}^2 = n \max \left\{ \max_{T-m+1 \leq s \leq T} \int D_{T,s}^2, \int \left(\sqrt{p_{\hat{\theta}, \bar{s}|\bar{s}-1; s|s-1}} - \sqrt{p_{\theta, \bar{s}|\bar{s}-1; s|s-1}} \left(1 + \frac{h^\tau}{2\sqrt{T}} \Lambda_\theta \right) \right)^2 d\lambda^k \right\} = o(n^0),$$

und weil die letzten m Terme $D_{T,s}$ jeder für sich $o(1/T)$ sind.

Aus der soeben gezeigten L_2 -Differenzierbarkeit folgt unmittelbar aus Theorem 10.2.7 die LAN-Eigenschaft. ////

Bemerkung 10.3.4 Sind für das Modell (55) bis (57) die Voraussetzungen zu Theorem 10.3.2 erfüllt, und gibt es zusätzlich eine Version der Lebesgue-dichten g , die symmetrisch ist im Sinn, daß $g(-x) = g(x)$, so ist das Modell unter Mitbeobachtung der Zustände robust adaptiv:

Die Bestandteile der Scoresfunktion, die zur Dichte h gehören, sind im idealen Modell sowieso unabhängig von denen zu g . Interessant ist also nur die Interaktion Λ_F und Λ_Q . Hier zeigt sich die Behauptung aber wie im univariaten Lokations-/Skalenmodell.

Diese Adaptivität geht i.a. beim Bedingen auf \mathcal{G}_t verloren, wie wir in Abschnitt 10.5 sehen werden; allerdings herrscht eine "Adaptivität" zwischen der

Schätzung der Zustände und der der Hyper-Parameter, so daß wir in diesem Fall die Hyper-Parameter nur \sqrt{T} -konsistent zu schätzen brauchen.

10.4 Exkurs: lokale Identifizierbarkeit der Hyper-Parameter

Hier wollen wir die Frage beantworten, wann die Fisher-Information des Modells \mathcal{I}_θ vollen Rang hat. Hierzu wollen wir den Begriff der lokalen Identifizierbarkeit verwenden, der auf eine Arbeit von Rothenberg (1971) zurückgeht:

Definition 10.4.1 Sei $Y(\theta) := \{Y_l(\theta); l = 1, \dots, k\}$ ein k -dimensionaler Zufallsvektor, dessen Verteilung durch $\theta \in \Theta \subset \mathbb{R}^\eta$ η -dimensional parametrisiert ist, und bezeichnen wir mit $\mu(\theta)$ und $\Sigma(\theta)$ seine ersten beiden Momente. Dann heißt das Modell $\mathcal{P} := \{\mathcal{L}(Y(\theta))\}$ in $\theta_0 \in \Theta$ identifizierbar zweiter Ordnung (Θ), falls

$$\forall \theta \in \Theta \quad \left. \begin{array}{l} \mu(\theta) = \mu(\theta_0) \\ \Sigma(\theta) = \Sigma(\theta_0) \end{array} \right\} \implies \theta = \theta_0,$$

d.h. falls θ_0 der eindeutige Parameter in Θ mit Bild (μ, Σ) ist. Das Modell heißt in $\theta_0 \in \Theta$ lokal identifizierbar zweiter Ordnung, wenn es eine Umgebung N_{θ_0} von θ_0 gibt, auf der θ identifizierbar zweiter Ordnung ist.

Wir wollen für unsere Untersuchungen eine Charakterisierung dieser Eigenschaft ähnlich der von Goodrich und Caines (1979) herleiten:

Theorem 10.4.2 (Goodrich/Caines) Seien $\{Y_s\}_{1 \leq s \leq T}$ stationäre, ergodische k -dimensionale Beobachtungen aus einer Folge parametrischer Modelle $\mathcal{P}_T = \{P_{\theta,T} | \theta \in \Theta\}$, deren erste beide Momente existieren; sei

$$\Phi : \Theta \rightarrow \mathbb{R} \times \mathbb{R}^{\binom{k}{2}}, \quad \theta \mapsto \Phi(\theta) = \left(E_{P_\theta}[Y_1], \text{svec}(\text{Cov}_{P_\theta}[Y_1]) \right) \quad (10.66)$$

Sei $\theta_0 \in \Theta$ so, daß lokal um θ_0 $\text{rk } D\Phi = \text{const}$ ist und sei Φ stetig differenzierbar. Dann ist die Modellfolge \mathcal{P}_t lokal identifizierbar (zweiter Ordnung), genau dann, wenn die Matrix M ,

$$M = (M_{i,j})_{i,j} = \left[\frac{1}{2} \text{tr} \left(\Sigma^{-\frac{1}{2}} \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-\frac{1}{2}} \right) \left(\Sigma^{-\frac{1}{2}} \frac{\partial \Sigma}{\partial \theta_j} \Sigma^{-\frac{1}{2}} \right) + \frac{\partial \mu}{\partial \theta_i}{}^\tau \Sigma^{-1} \frac{\partial \mu}{\partial \theta_j} \right] \Bigg|_{\theta=\theta_0}, \quad (10.67)$$

$1 \leq i, j \leq \eta$

regulär ist.

Dies kann man verwenden, um auf die Invertierbarkeit der Fisher-Information in einem parametrischen Modell zu schließen. Dabei betrachten wir Beobachtungen $\{Y_s\}_{1 \leq s \leq T}$ wie in Theorem 10.4.2. Im Unterschied hierzu betrachten wir aber auch noch die Autokovarianzen bis zur Ordnung m . Als Preis dafür müssen wir zu Glättungsfiltrierungen bis maximal Ordnung m übergehen. Diese sind ursprünglich bei einer zu erklärenden Variable definiert, die nicht zu beobachten ist; geht man gleich zum durch die Beobachtungen erzeugten, größeren Modell über, äußert sich dies in einer nicht uniformen Beobachtungsfiltrierung \mathcal{G}_T mit $\mathcal{G}_T = \{\sigma(Y_r)_{r \leq k_s}\}_{s \leq T}$ mit $s, k_{s-1} \leq k_s \leq T$, bzw. in einer fixed-lag-Glättung $\mathcal{G}_T^m := \{\sigma(Y_r)_{r \leq (s+m) \wedge T}\}_{s \leq T}$.

Theorem 10.4.3 Seien $\theta_0 \in \Theta$, $\{Y_s\}_{1 \leq s \leq T}$ wie in Theorem 10.4.2, und sei die Autokovarianzfunktion exponentiell abklingend. Sei zusätzlich das Modell L_2 -differenzierbar längs der Filtrierung \mathcal{G}_T^m , wobei wir zulassen, daß die Fisher-Information singular ist. Die Fisher-Information sei aber stetig in θ_0 . Dann ist die Fisher-Information \mathcal{I}_{θ_0} regulär, wenn $\{\mathcal{L}([Y_1^\tau(\theta), \dots, Y_{m+1}^\tau(\theta)]^\tau) \mid \theta \in \Theta\}$ lokal identifizierbar (zweiter Ordnung) ist, also wenn wir in die Abbildung Φ in Theorem 10.4.2 nicht nur die Kovarianz $\text{Cov}_{P_\theta}[Y_1]$, sondern auch noch die Autokovarianzen $\text{Cov}_{P_\theta}[Y_1, Y_{1+i}]$, $i = 1, \dots, m$ einschließen.

BEWEIS ZU THEOREM 10.4.3: Wir bemerken zunächst, daß im stationären, ergodischen Fall die Regularität der Fisher-Information für den Beweis des LAN-Theorems 10.2.7 nicht benötigt wird.

Daher dürfen wir Lemma 10.2.13 verwenden. Von der Schätzung $\hat{\Phi}(\theta)$ benötigen wir nur, daß

$$\sqrt{n}(\hat{\Phi}(\theta) - \Phi(\theta)) \circ P_\theta \xrightarrow{w} \mathcal{N}(0, \mathcal{C}_\theta) \quad (10.68)$$

mit einem in θ stetigen \mathcal{C}_θ . Es gilt

$$\Phi(\theta_0 + h/\sqrt{n}) - \Phi(\theta_0) = \partial\Phi(\theta_0)h/\sqrt{n} + o(1/\sqrt{n}). \quad (10.69)$$

Für jedes θ können wir andererseits $\Phi(\theta)$ erwartungstreu durch die empirischen Momente schätzen⁵, wofür wir schlicht $\hat{\Phi}(\theta)$ schreiben; dabei haben wir durch die Wahl der Filtrierung erreicht, daß die entsprechenden Summanden als Martingaldifferenzen bezüglich dieser Filtrierung aufgefaßt werden können.

Nach Hannan (1970) aber gilt (68), und mit der Stetigkeit von \mathcal{C}_θ und (69), sowie der lokal in θ gleichmäßigen Normalapproximation, abzulesen in der exponentiellen Mixing-Eigenschaft der Komponenten von Φ , folgt

$$\sqrt{n}(\hat{\Phi}(\theta) - \Phi(\theta)) \circ P_{\theta+h/\sqrt{n}} \xrightarrow{w} \mathcal{N}(\partial\Phi(\theta_0)h, \mathcal{C}_\theta)$$

Mit anderen Worten, $\hat{\Phi}(\theta)$ ist regulär⁶ im Sinn von Definition A.5.4. Nun ist $\hat{\Phi}(\theta)$ linear in $(Y_s, [Y_s Y_{s+i}^\tau]_{0 \leq i \leq m})$, für jedes θ erwartungstreu und adaptiert zur Filtrierung \mathcal{G}_T^m . Daher gilt nach Lemma 10.2.13 $E_\theta \rho_\theta \Lambda_\theta^\tau = \partial\Phi(\theta)$. Also hat nach Cauchy-Schwartz \mathcal{I}_θ mindestens Rang $\text{rk } \partial\Phi(\theta)$, und damit ist \mathcal{I}_θ regulär, falls $\partial\Phi(\theta)$ vollen Rang hat. ////

Bemerkung 10.4.4 In der Originalarbeit Goodrich und Caines (1979) wie auch in Caines (1988), Abschnitt 9.6 verwenden die Autoren die stärkere Voraussetzung der zweimal stetigen Differenzierbarkeit der Loglikelihood des Modells nach θ . Dies benötigen wir nicht, sondern wir werden mit der Δ -Methode auf die Invertierbarkeit von \mathcal{I}_θ schließen, so daß nur die partielle Integration erlaubt sein muß.

10.4.1 Skizze des Vorgehens

Wir werden zeigen, daß es unter geeigneten Rangbedingungen an F , Z , Q und V einen (suboptimalen) asymptotisch linearen Schätzer gibt, der sich aus

⁵Genauer wird $\mu(\theta)$ durch $\hat{\mu}(\theta) := \frac{1}{n} \sum_j X_j$ und $\Sigma(\theta)$ durch $\hat{\Sigma}(\theta) := \frac{1}{n-1} \sum_j (X_j - \hat{\mu}(\theta))(X_j - \hat{\mu}(\theta))^\tau$ geschätzt

⁶Zusammen mit Proposition A.5.6 ist damit der Nachweis der Rangeigenschaft von T resp. \tilde{T} in Abschnitten 10.4.3 und 10.4.4 nicht mehr erforderlich. Er wurde aber trotzdem durchgeführt, weil man im Spezialfall des Zustandsraummodells noch explizitere Darstellungen erhält und anhand dieser (im Prinzip) auf die Güte der Startschätzung schließen kann.

der empirischen Autokovarianzfunktion (ACF) $\gamma^{(Y)}$ des y -Prozesses gewinnen läßt, indem man die ersten drei Terme $\gamma_i^{(Y)}$, $i = 0, 1, 2$ nach den gesuchten Hyper-Parametern F , Q , V auflöst. Mithilfe eines Zentralen Grenzwertsatzes für $\hat{\gamma}^{(Y)}$ und der endlich-dimensionalen Δ -Methode schließen wir auf eine asymptotische Normalität dieser Momentenschätzer mit einer Limes-Kovarianz von gewünschtem Rang.

Andererseits kann man dann für die in der endlich-dimensionalen Δ -Methode auftauchende Jakobische partielle Integration anwenden und erhält so die Invertierbarkeit von \mathcal{I}_θ längs einer fixed-lag-Glättungsfiltrierung mit $m = 2$.

Bemerkung 10.4.5 (a) Einen ähnlichen Ansatz, wenngleich “nur” für das Gaußsche Modell, verfolgt Pagan (1980). Allerdings sind bei ihm die Z_t zeitvariabel und $q = 1$. In diesem Fall sind die Eigenschaften (UCC / UCO)⁷ hinreichend für die Identifizierbarkeit.

(b) In Theorem 10.4.6 könnten wir im Gaußschen Fall auf die Bedingung $\text{rk } V = q$ verzichten, was allerdings die Rechnungen unnötig weiter verkomplizieren würde; alle anderen Bedingungen sind unter $\rho(F) < 1$ im Gaußschen Fall auch hinreichend: Ohne Einschränkung sei $p = q$, $Z = \mathbb{I}_p$; man stelle sich F , Q und V als Diagonalmatrizen vor. Dann gilt für die suffiziente Autokovarianz von y_t mit $C = (\mathbb{I}_p - F)^{-1}Q$

$$\gamma^{(Y)}(h) = F^h C + \mathbb{I}_{\{h=0\}} V$$

Ist nun FQ nicht invertierbar, so ist nicht festzustellen, ob F oder Q nicht invertierbar ist, da im einzigen Term ohne Vorfaktor F , $\gamma^{(Y)}(0)$, nicht entschieden werden kann, was zu V und was zu C gehört.

(c) Der Nachweis hinreichender Bedingungen in Caines (1988) Abschnitt 9.4 auf Basis der vollen Rang-Bedingung der Beobachtungsmatrix

$$\mathcal{O} = [Z, ZF, \dots, ZF^p]$$

und der Erreichbarkeitsmatrix

$$\mathcal{C} := [Q^{\frac{1}{2}}, F^T Q^{\frac{1}{2}}, \dots, (F^T)^p Q^{\frac{1}{2}}]$$

ist nur scheinbar allgemeiner als unsere Bedingungen:

- Anders als behauptet, kann man nicht jedes beobachtbare lineare Zustandsraummodell mit einem Basiswechsel auf eine Übergangsmatrix \tilde{F} von Frobeniusgestalt⁸

$$\tilde{F} = \begin{pmatrix} a_1 & a_2 & \dots & a_{p-1} & a_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ & & \ddots & \vdots & \vdots \\ \dots & & & \ddots & 1 & 0 \end{pmatrix} \quad (10.70)$$

bringen, wie folgendes Gegenbeispiel zeigt:

Die Matrix $F = \mathbb{I}_p$ ist bereits in Jordan-Normalform und läßt sich daher nicht durch Konjugation mit einer Matrix $T \in \text{GL}(2)$ auf Gestalt

⁷vgl. Bemerkung 7.2.22

⁸in englischen Büchern oft auch als “Companion-Matrix” bezeichnet, vgl. Searle (1997)

$F' = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$ bringen, da diese die Jordan–Normalform $J_{F'} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ besitzt. Andererseits stimmen die beiden charakteristischen Polynome der beiden Matrizen überein, $\varphi_F(\lambda) = \varphi_{F'}(\lambda) = (\lambda - 1)^2$.

- Für den Beweis von Lemma 4.3 in Abschnitt 9.4 benötigt Caines (1988) dann doch invertierbares F , und daß Q invertierbar, sowie $\text{rk } Z = p$ ist, also bleiben im wesentlichen unsere Bedingungen unabgeschwächt.

(d) Ein anderer Begriff der Identifizierbarkeit findet sich ebenfalls in Caines (1988): Es geht nun nicht mehr darum in vorgegebenen Dimensionen p, q ein spezielles Zustandsraummodell (F, Z, Q, V) zu identifizieren, sondern ein Zustandsraummodell $(\tilde{F}, \tilde{Z}, \tilde{Q}, \tilde{V})$ in geeigneten, frei wählbaren Dimensionen \tilde{p}, \tilde{q} anzugeben, das geeignet ist, das Verhalten des y_t -Prozesses möglichst gut (im Sinne des mittleren quadratischen Vorhersagefehlers) nachzubilden.

Dieser schwächere Identifizierbarkeitsbegriff liegt auch Aoki (1990) zugrunde, und man kommt hier mit Beobachtbarkeit und Kontrollierbarkeit aus; man gelangt so zu sogenannten “Minimum Prediction Error Methods”, auf die wir im folgenden aber nicht eingehen. Stattdessen wollen wir uns allein auf die Identifizierbarkeit im Sinne der vollen Modellidentifikation konzentrieren.

10.4.2 hinreichende Bedingungen für Regularität von \mathcal{I}

Als zentrales Resultat im Kontext des linearen, zeitinvarianten Zustandsraummodells werden wir zeigen:

Theorem 10.4.6 *Ist im linearen, zeitinvarianten Zustandsraummodell mit*

$$\rho(F) < 1 \quad (10.71)$$

$$\text{rk } Z \geq \text{rk } F = \text{rk } Q = p, \quad \text{rk } V = q \quad (10.72)$$

$$\text{E} |\varepsilon_t|^4 < \infty, \quad \text{E} |v_t|^4 < \infty \quad (10.73)$$

$$\varepsilon \sim \mathcal{N}_q(0, V) \quad \text{oder} \quad \text{rk Cov}[\text{svec}(\varepsilon\varepsilon^\tau)] = \begin{pmatrix} q \\ 2 \end{pmatrix} \quad (10.74)$$

$$v \sim \mathcal{N}_p(0, Q) \quad \text{oder} \quad \text{rk Cov}[\text{svec}(vv^\tau)] = \begin{pmatrix} p \\ 2 \end{pmatrix}, \quad (10.75)$$

so ist $\det \mathcal{I}_\theta > 0$, wenn wir die Zweischritt–Glättungsfiltrierung \mathcal{G}_T^2 verwenden — in der Notation von (58) $\mathcal{G}_T^2 = (\mathcal{F}_{T,0,(s+2)\wedge T})$.

Bemerkung 10.4.7 (a) Im Fall normalverteilter ε_t, v_t ist die Fisher–Information bereits regulär, wenn wir $\mathcal{G}_T^1 = (\mathcal{F}_{T,0,(s+1)\wedge T})$, die Einschnitt–Glättungsfiltrierung verwenden, wie wir in Proposition 10.5.1 sehen werden.

(b) Die Bedingungen des Theorems sind nicht notwendig; so müssen für die L_2 -Differenzierbarkeit noch nicht einmal zweite Momente von ε, v existieren. Falls aber diese nicht existieren, ist auch der später verwendete Startschätzer auf Basis der Autokovarianzfunktion des Y -Prozesses nicht mehr zulässig.

10.4.3 asymptotische Kovarianz der emp. ACF $\hat{\gamma}^{(Y)}$

Zunächst betrachten wir die ACF $\gamma^{(Y)}$ und deren kanonischen Schätzer⁹ $\hat{\gamma}_i^{(Y)} := \frac{1}{T} \sum_{t=1}^{T-i} y_t y_{t+i}^\tau$. Es gilt

Proposition 10.4.8 *Sei im linearen, zeitinvarianten Zustandsraummodell*

$$\rho(F) < 1, \quad \mathbb{E} |\varepsilon_t|^4 < \infty, \quad \mathbb{E} |v_t|^4 < \infty.$$

Dann gilt

$$\sqrt{T}(\hat{\gamma}_i^{(Y)} - \gamma_i^{(Y)})_{i=0, \dots, m} \xrightarrow{w} \mathcal{N}_{(m+1)q^2}(0, T) \quad (10.76)$$

Dabei ist

$$T = (\mathcal{T}_{i,j; i_1, i_2, j_1, j_2})_{\substack{1 \leq i, j \leq m \\ 1 \leq i_1, i_2, j_1, j_2 \leq q}} := \text{asCov}(\hat{\gamma}_i^{(Y)}, \hat{\gamma}_j^{(Y)}),$$

$$\tilde{T} = (\tilde{\mathcal{T}}_{i,j; i_1, i_2, j_1, j_2})_{\substack{1 \leq i, j \leq m \\ 1 \leq i_1, i_2, j_1, j_2 \leq p}} := \text{asCov}(\hat{\gamma}_i^{(\beta)}, \hat{\gamma}_j^{(\beta)}),$$

mit

$$C = \sum_{t=0}^{\infty} F^t Q (F^\tau)^t, \quad (10.77)$$

$$\mathcal{T}_{i,j} = Z^{\otimes 2} \tilde{\mathcal{T}}_{i,j} (Z^{\otimes 2})^\tau + \mathcal{R}(j-i) + K_{q,q} \mathcal{R}(j+i) + [(\mathbb{I}_{q^2} + K_{q,q}) V^{\otimes 2} + \kappa^{(\varepsilon)}] \mathbf{I}_{\{i=j=0\}} \quad (10.78)$$

$$\begin{aligned} \tilde{\mathcal{T}}_{i,j} = & \mathcal{S}(j-i) + K_{p,p} \mathcal{S}(j+i) + \\ & + \{[(\mathbb{I}_{p^2} - F^{\otimes 2})^{-1} \otimes \mathbb{I}_{p^2}] + [\mathbb{I}_{p^2} \otimes (\mathbb{I}_{p^2} - F^{\otimes 2})^{-1} F^{\otimes 2}]\} \times \\ & \times (\mathbb{I}_{p^4} - F^{\otimes 4})^{-1} [(F^{-i^-} \otimes F^{i^+} \otimes F^{-j^-} \otimes F^{j^+})] \text{vec}(\kappa^{(v)}), \end{aligned} \quad (10.79)$$

$$\begin{aligned} \mathcal{S}(l) = & \mathbf{I}_{l < 0} \{ (\mathbb{I}_p \otimes (F^\tau)^{|l|}) (\mathbb{I}_{p^2} - F^{\otimes 2})^{-1} C^{\otimes 2} + \\ & + (C F^\tau)^{\otimes 2} (F^{|l|} \otimes \mathbb{I}_p) (\mathbb{I}_{p^2} - F^{\otimes 2})^{-\tau} + \\ & + \sum_{\nu=0}^{|l|-1} (C (F^\tau)^\nu \otimes F^{|l|-1-\nu} C) \} \\ & + \mathbf{I}_{l \geq 0} \{ ((F^\tau)^l \otimes \mathbb{I}_p) (\mathbb{I}_{p^2} - F^{\otimes 2})^{-1} C^{\otimes 2} + \\ & + (C F^\tau)^{\otimes 2} (\mathbb{I}_p \otimes (F^\tau)^l) (\mathbb{I}_{p^2} - F^{\otimes 2})^{-\tau} + \\ & + \sum_{\nu=0}^{l-1} (F^{l-1-\nu} C \otimes C (F^\tau)^\nu) \}, \end{aligned} \quad (10.80)$$

$$\mathcal{R}(l) = [(Z \gamma_l^{(\beta)} Z^\tau) \otimes V + V \otimes (Z \gamma_{-l}^{(\beta)} Z^\tau)] \quad (10.81)$$

und $\kappa^{(\cdot)}$ der vierten Kumulante von ε bzw v .

Zum Beweis benötigen wir

Lemma 10.4.9 *Sei im VAR(1)-Modell*

$$\beta_t = F \beta_{t-1} + v_t$$

$\rho(F) < 1$, $\mathbb{E} |v_t|^4 < \infty$. Dann gilt

$$\sqrt{T}(\hat{\gamma}_i^{(\beta)} - \gamma_i^{(\beta)})_{i=1, \dots, m} \xrightarrow{w} \mathcal{N}_{mp^2}(0, \tilde{T}) \quad (10.82)$$

⁹In der Modellspezifikation hatten wir bereits die Zentriertheit $\mathbb{E} v = 0$, $\mathbb{E} \varepsilon = 0$ vorausgesetzt.

BEWEIS ZU LEMMA 10.4.9: Wir verwenden Hannan (1970) und spezialisieren Formel (3.3) in Kapitel IV für $n \hat{=} j$ $m \hat{=} i$:

$$\begin{aligned} \text{Cov}(\hat{\gamma}_i^{(\beta)}, \hat{\gamma}_j^{(\beta)}) &= \frac{1}{T} \sum_{\nu=-T+1}^{T-1} \left(1 - \frac{|\nu|}{T}\right) \left\{ \kappa_{i_1, i_2, j_1, j_2}^{(\beta)}(0, i, \nu, \nu + j) + \right. \\ &\quad \left. + \{[\gamma_\nu^{(\beta)} \otimes \gamma_{\nu+j-i}^{(\beta)}]_{i_1, i_2, j_1, j_2} + [\gamma_{\nu-j}^{(\beta)} \otimes \gamma_{\nu+i}^{(\beta)}]_{i_1, i_2, j_2, j_1}\} \right\}_{i_1, i_2, j_1, j_2} \end{aligned} \quad (10.83)$$

Für gemischte Terme des β -Prozesses¹⁰ schreiben wir

$$\begin{aligned} \kappa_{i_1, i_2, i_3, i_4}^{(\beta)}(t_1, t_2, t_3, t_4) &:= \mathbb{E}[\beta_{t_1, i_1} \beta_{t_2, i_2} \beta_{t_3, i_3} \beta_{t_4, i_4}] - \mathbb{E}[\beta_{t_1, i_1} \beta_{t_2, i_2}] \mathbb{E}[\beta_{t_3, i_3} \beta_{t_4, i_4}] - \\ &\quad - \mathbb{E}[\beta_{t_1, i_1} \beta_{t_3, i_3}] \mathbb{E}[\beta_{t_2, i_2} \beta_{t_4, i_4}] - \mathbb{E}[\beta_{t_1, i_1} \beta_{t_4, i_4}] \mathbb{E}[\beta_{t_2, i_2} \beta_{t_3, i_3}], \end{aligned}$$

wobei β_{t_i, i_i} die i_i -te Koordinate der t_i -ten Variable β_{t_i} ist. Nun ist (asymptotisch) für $i \geq 0$

$$(\gamma_i^{(\beta)})^\tau = \gamma_{-i}^{(\beta)} = \mathbb{E}[\beta_t \beta_{t-i}^\tau] = F^i C.$$

Wir zerlegen die Summe (83) über die Summanden jeweils getrennt in Bereiche, wo der Index negativ oder positiv ist. Gehen wir zunächst vom normalverteilten Fall aus; dann verschwindet die vierte Kumulante, und wir erhalten die \mathcal{S} -Terme, falls wir zunächst den Faktor $\left(1 - \frac{|\nu|}{T}\right)$ ignorieren; aber dieser geht für $T \rightarrow \infty$ gegen 0. Die Konvergenz erfolgt nach dem Satz von der dominierten Konvergenz, wenn wir noch berücksichtigen, daß jeder der Summanden als Kovarianzmatrix eine positiv semidefinite $(p^2 \times p^2)$ -Matrix ist. Wie man den Kumulanten-Term berechnet, ist klar.

Die behauptete Normalität folgt mit einem Zentralen Grenzwertsatz wie Hannan (1970), Thm. IV.14. ////

BEWEIS ZU PROPOSITION 10.4.8: Mit $y_t = Z\beta_t + \varepsilon_t$ folgt wegen der Unabhängigkeit des β und ε -Prozesses

$$\begin{aligned} \mathbb{E}[Y_{t_1, i_1} Y_{t_1+i, i_2} Y_{t_2, j_1} Y_{t_2+j, j_2}] &= \mathbb{E}[(Z\beta)_{t_1, i_1} (Z\beta)_{t_1+i, i_2} (Z\beta)_{t_2, j_1} (Z\beta)_{t_2+j, j_2}] + \\ &\quad + \mathbb{E}[\varepsilon_{t_1, i_1} (Z\beta)_{t_1+i, i_2} \varepsilon_{t_2, j_1} (Z\beta)_{t_2+j, j_2}] + \mathbb{E}[\varepsilon_{t_1, i_1} (Z\beta)_{t_1+i, i_2} (Z\beta)_{t_2, j_1} \varepsilon_{t_2+j, j_2}] + \\ &\quad + \mathbb{E}[(Z\beta)_{t_1, i_1} \varepsilon_{t_1+i, i_2} \varepsilon_{t_2, j_1} (Z\beta)_{t_2+j, j_2}] + \mathbb{E}[(Z\beta)_{t_1, i_1} \varepsilon_{t_1+i, i_2} (Z\beta)_{t_2, j_1} \varepsilon_{t_2+j, j_2}] + \\ &\quad + \mathbb{E}[\varepsilon_{t_1, i_1} \varepsilon_{t_1+i, i_2} \varepsilon_{t_2, j_1} \varepsilon_{t_2+j, j_2}] = \\ &= \{Z^{\otimes 2} \tilde{\mathcal{T}}_{i, j}; (Z^{\otimes 2})^\tau\}_{i_1, i_2, j_1, j_2} + \mathcal{R}(j-i)_{i_1, i_2, j_1, j_2} + \mathcal{R}(j+i)_{i_1, i_2, j_2, j_1} + \\ &\quad + \mathbb{I}_{\{i=j=0\}} \left[\kappa_{i_1, i_2, j_1, j_2}^{(\varepsilon)} + (V \otimes V)_{i_1, i_2, j_1, j_2} + (V \otimes V)_{i_1, i_2, j_2, j_1} \right] + \\ &\quad + \mathbb{E}[Y_{t_1, i_1} Y_{t_1+i, i_2}] \mathbb{E}[Y_{t_2, j_1} Y_{t_2+j, j_2}] \end{aligned} \quad (10.84)$$

die behauptete Darstellung von \mathcal{T} , und die asymptotische Normalität folgt wieder mit einem Zentralen Grenzwertsatz wie Hannan (1970), Thm. IV.14. ////

Für das weitere Vorgehen spezialisieren wir Proposition 10.4.8 auf $q = 1$.

$$\begin{aligned} \text{as. Cov}(\hat{\gamma}_i^{(Y)}, \hat{\gamma}_j^{(Y)}) &= \frac{Z^4 Q^2}{1 - F^2} \left\{ F^{|i+j|} \left(\frac{1 + F^2}{1 - F^2} + |i + j| \right) + \right. \\ &\quad \left. + F^{|i-j|} \left(\frac{1 + F^2}{1 - F^2} + |i - j| \right) + \frac{2Z^2 QV}{1 - F^2} \left(F^{|i+j|} + F^{|i-j|} \right) + 2V^2 \mathbb{I}_{i=j} \right\} \succ 0. \end{aligned}$$

¹⁰Hier verwenden wir, daß $\mathbb{E}\beta_t = 0$ gilt.

10.4.4 Rangbetrachtung von \mathcal{T}

Wir werden zeigen, daß

Proposition 10.4.10

$$\text{rk } \mathcal{T} \geq \binom{q}{2} + mp^2 =: d_1 \quad (10.85)$$

Hierzu benötigen wir folgende

Proposition 10.4.11 *Für schließlich jedes n gilt*

$$\tilde{\mathcal{T}}_n := n \text{Cov}\{[\text{svec}(\hat{\gamma}_0^{(\beta):n.m})]^\tau, [\text{vec}(\hat{\gamma}_1^{(\beta):n.m})]^\tau, \dots, [\text{vec}(\hat{\gamma}_m^{(\beta):n.m})]^\tau\}^\tau$$

mit $\hat{\gamma}_i^{(\beta):n.m} := \frac{1}{n} \sum_{t=1}^{n-m} \beta_t \beta_{t+i}^\tau$ ist regulär mit Rang $d_3 := d_2 + mp^2$, wobei $d_2 = \text{rk Cov}[\text{svec}(\beta_t \beta_t^\tau)] = \binom{p}{2}$.

BEWEIS ZU PROPOSITION 10.4.10: Betrachten wir nun die Matrix C_0

$$C_0 := \text{Cov}\{[\text{svec}(\varepsilon \varepsilon^\tau)]^\tau, [\text{vec}(\hat{\gamma}_1^{(\beta):n.m})]^\tau, \dots, [\text{vec}(\hat{\gamma}_m^{(\beta):n.m})]^\tau\}^\tau$$

Diese hat Blockgestalt

$$C_0 = \begin{pmatrix} C_1 & 0 \\ 0 & C_2 \end{pmatrix}$$

mit

$$\begin{aligned} C_1 &= \text{svec}^{-1}\{\text{vec}[\kappa^{(\varepsilon)} + (\mathbb{I}_{q^2} + K_{q,q})V^{\otimes 2}]\} \\ C_2 &= (\tilde{\mathcal{T}}_{i,j})_{i,j=1,\dots,m} \end{aligned}$$

Aus Proposition B.5.1 bzw. Korollar B.5.2 folgt, daß $\text{rk } C_1 = \binom{q}{2}$, und aus Proposition 10.4.11, daß $\text{rk } C_2 = mp^2$.

Wegen $\text{rk } Z = p$ besitzt daher die Matrix

$$C_3 := \text{diag}[\mathbb{I}_{\binom{q}{2}}, Z^{\otimes 2}, \dots, Z^{\otimes 2}] C_0 \text{diag}[\dots]^\tau$$

Rang d_1 . Andererseits läßt sich \mathcal{T} darstellen als $\mathcal{T} = C_3 + C_4 + C_5$ mit $C_4, C_5 \succeq 0$ und

$$C_4 = \begin{pmatrix} Z^{\otimes 2} \tilde{\mathcal{T}}_{0,0} [Z^{\otimes 2}]^\tau & 0 \\ 0 & 0 \end{pmatrix}, \quad C_5 = [\mathcal{R}(i-j) + K_{q,q} \mathcal{R}(i+j)]_{i,j=0,\dots,m}.$$

Zu zeigen ist nun $C_5 \succeq 0$; sei dazu $U_t := Z\beta_t$, $\gamma^{(U)}$ die dazugehörige ACF. Dann ist $C_5 = C_6 + C_6^\tau$ mit

$$C_6 := \begin{aligned} & (\text{as.}) \text{Cov}[(U_{t+i}\varepsilon_t)_{i_1, i_2} (U_{t+j}\varepsilon_t)_{j_1, j_2}]_{i,j} \\ & + (\text{as.}) \text{Cov}[(U_{t+i}\varepsilon_t)_{i_1, i_2} (U_{t-j}\varepsilon_t)_{j_2, j_1}]_{i,j} \end{aligned}$$

Zu überprüfen ist, ob mit $U_{+\varepsilon} := \text{vec}[(U_{t+i}\varepsilon_t^\tau)_i]$ und $U_{-\varepsilon} := \text{vec}[(U_{t-i}\varepsilon_t^\tau)_i]$

$$\mathbb{E} \left\{ \begin{pmatrix} A \\ A \end{pmatrix}^\tau \begin{pmatrix} U_{+\varepsilon} \\ U_{-\varepsilon} \end{pmatrix} \begin{pmatrix} U_{+\varepsilon} \\ U_{-\varepsilon} \end{pmatrix}^\tau \begin{pmatrix} A \\ A \end{pmatrix} \right\} \succeq 0.$$

Dies aber ist positiv semidefinit nach Konstruktion und wegen Stationarität hat (as.) $\text{Cov}[(\cdot), (\cdot)]$ die Struktur $\begin{pmatrix} D & B \\ B & D \end{pmatrix}$, also $D^2 - B^2 \geq 0$, in unserem Fall

$$\text{Cov}[(A^\tau U_{+\varepsilon})(A^\tau U_{+\varepsilon})] > \text{Cov}[(A^\tau U_{+\varepsilon})(A^\tau U_{-\varepsilon})],$$

und damit

$$C_6 = \text{Cov}[(U_{+\varepsilon})(U_{+\varepsilon})] + \text{Cov}[(U_{+\varepsilon})(U_{-\varepsilon})] \succeq 0.$$

Damit folgt die Behauptung. ////

Zum Beweis von Proposition 10.4.11 verwenden wir folgende Resultate:

Proposition 10.4.12 *Der Prozeß $\{R_t\}$,*

$$R_t := ([\text{svec}(\beta_t \beta_t^\tau)]^\tau, [\text{vec}(\beta_t \beta_{t+1}^\tau)]^\tau, \dots, [\text{vec}(\beta_t \beta_{t+m}^\tau)]^\tau)^\tau$$

ist strikt stationär und besitzt eine Kovarianz $\gamma^{(R)}(0) := C^{(R)}$ von vollem Rang d_3 .

Proposition 10.4.13 *Sei $\{X_t\}$ ein \mathbb{R}^p -wertiger, schwach stationärer Prozeß zweiter Ordnung. Falls $\gamma_0^{(X)} > 0$ und $|\gamma_m^{(X)}| \rightarrow 0$ für $h \rightarrow \infty$, dann besitzt die Kovarianzmatrix $\Gamma_n := (\gamma_{i-j}^{(X)})_{i,j=1,\dots,n}$ vollen Rang.*

BEWEIS ZU PROPOSITION 10.4.11: Wegen Proposition 10.4.12 ist Proposition 10.4.13 mit $X_t \hat{=} R_t$ in Kraft, denn $\gamma_m^{(R)} = O(\rho(F)^{2m})$. Damit gilt dann für $0 \neq a \in \mathbb{R}^{d_3}$

$$a^\tau \tilde{\mathcal{T}}_n a = \frac{1}{n} \sum_{s,t=1}^{n-m} a^\tau \text{Cov}[R_t, R_s] a = \frac{1}{n} \sum_{s,t=1}^{n-m} a^\tau \gamma_{t-s}^{(R)} a > 0,$$

also die behauptete positive Definitheit. ////

Nun noch die Beweise zu Propositionen 10.4.12 und 10.4.13

BEWEIS ZU PROPOSITION 10.4.13: [vgl. Proposition 5.1.1 in Brockwell und Davis (1991)] Nehmen wir an, die Behauptung sei falsch; sei dazu $r+1$ der erste Index, so daß Γ_{r+1} singularär wird. Dann gilt mit einem $0 \neq a_{r+1} \in \mathbb{R}^p$ und $a_j \in \mathbb{R}^p$

$$a_{r+1}^\tau X_{r+1} = \sum_{j=1}^r a_j^\tau X_j \quad \text{fast sicher.}$$

Wegen der Stationarität von X folgt damit

$$a_{r+1+m}^\tau X_{r+1+m} = \sum_{j=1}^r a_j^\tau X_{j+m} = \sum_{j=1}^r a_j^{(m)\tau} X_j \quad \text{fast sicher}$$

mit gewissen $a_j^{(m)} \in \mathbb{R}^p$. Diese fassen wir zu einem Vektor $a^{(m)} \in \mathbb{R}^{r \cdot p}$ zusammen und schreiben \underline{X} für $(X_1^\tau, \dots, X_r^\tau)^\tau$. Nach Voraussetzung ist Γ_r regulär; sei daher $\lambda_{\min} > 0$ der minimale Eigenwert von Γ_r . Nun gilt

$$b := a_{r+1}^\tau \gamma_0^{(X)} a_{r+1} = a^{(h)\tau} \Gamma_r a^{(m)} \geq \lambda_{\min} |a^{(m)}|^2,$$

also sind die $a^{(m)}$ gleichmäßig beschränkt durch b/λ_{\min} . Andererseits ist auch

$$0 < b = a^{(m)\tau} \text{Cov}[\underline{X}X_{r+1+m}]a_{r+1} \leq |a_{r+1}| \frac{b}{\lambda_{\min}} \max_{l=1\dots r} |\gamma_{m+l}^{(X)}| \rightarrow 0,$$

ein Widerspruch. ////

BEWEIS ZU PROPOSITION 10.4.12: Zunächst gilt

$$C_{0,0}^{(R)} = \text{svec}^{-1} \left[(\mathbb{I}_{p^4} - F^{\otimes 4})^{-1} \text{vec}[\kappa^{(v)}] + \text{vec}[(\mathbb{I}_{p^2} + K_{p,p})C^{\otimes 2}] \right].$$

Rekursiv erhalten wir für $i \geq 1$

$$C_{i,i}^{(R)} = (F \otimes \mathbb{I}_p) C_{i-1,i-1}^{(R)} (F \otimes \mathbb{I}_p)^\tau + Q \otimes C$$

und

$$C_{i,j}^{(R)} = \begin{cases} (\mathbb{I}_p \otimes F^{i-j}) C_{j,j}^{(R)} & \text{für } i \geq j \\ C_{i,i}^{(R)} (\mathbb{I}_p \otimes F^{j-i})^\tau & \text{für } i < j \end{cases}$$

Damit erhalten wir induktiv durch Abzug des $(F \otimes I)^{j-i}$ -fachen der “ i -ten” Zeile von der “ j -ten”, wobei dies aufsteigend in i geschieht, eine obere Dreiecksmatrix $\tilde{C}^{(R)}$ mit demselben Rang wie $C^{(R)}$ und Diagonale $\text{diag}[\tilde{C}_{0,0}^{(R)}, Q \otimes C, \dots, Q \otimes C]$. Damit folgt die Behauptung. ////

10.4.5 Berechnung der Hyper-Parameter aus $\hat{\gamma}^{(Y)}$

Es gilt wegen $Z^-Z = \mathbb{I}_p$

$$\underline{\gamma} = (\text{svec}[ZCZ^\tau + V]^\tau, \text{vec}[FC]^\tau, \text{vec}[F^2C]^\tau)^\tau + o_P(T^0) \quad (10.86)$$

Lösen wir dies nach F, Q, V auf, so erhalten wir:

$$\hat{F} = \text{vec}^{-1}[\underline{\gamma}_2] \{ \text{vec}^{-1}[\underline{\gamma}_1] \}^{-1} \quad (10.87)$$

$$\hat{C} = \hat{F}^{-1} \text{vec}^{-1}[\underline{\gamma}_1] \quad (10.88)$$

$$\hat{Q} = \hat{C} - \hat{F} \hat{C} \hat{F}^\tau \quad (10.89)$$

$$\hat{V} = \text{svec}^{-1}[\underline{\gamma}_0] - Z \hat{C} Z^\tau \quad (10.90)$$

Gilt nun $\text{rk} F = \text{rk} Z = p$, so ist F eindeutig bestimmt, da dann $\{Z^- \gamma_2^{(Y)} (Z^-)^\tau\}$ invertierbar ist. Dann aber sind auch C, Q und V eindeutig bestimmt; genauer ist $Z^-Z = \mathbb{I}_p$, und $C = \text{vec}^{-1}[(\mathbb{I}_{p^2} - (F \otimes F^\tau))^{-1} \text{vec}(Q)]$, so daß dann tatsächlich $\hat{F} = F, \hat{Q} = Q, \hat{C} = C, \hat{V} = V$.

10.4.6 Differenzierbarkeit von G und Ableitung

Diese Auflösung nach F, Q, V ist glatt in folgendem Sinn.

Proposition 10.4.14 Für $\rho(F) \leq 1$ ist die Abbildung

$$G : \mathbb{R}^{p \times p} \times \mathbb{R}^{p \times p} \times \mathbb{R}^{q \times q} \rightarrow \mathbb{R}^{(q^2+q)/2} \times \mathbb{R}^{p^2} \times \mathbb{R}^{p^2}$$

$$(u, v, w) \mapsto G(u, v, w) := \begin{pmatrix} \text{svec}[Z \sum_{t=0}^{\infty} u^t v (u^t)^\tau Z^\tau + w] \\ \text{vec}[\sum_{t=0}^{\infty} u^{t+1} v (u^t)^\tau] \\ \text{vec}[\sum_{t=0}^{\infty} u^{t+2} v (u^t)^\tau] \end{pmatrix}$$

an der Stelle (F, Q, V) stetig differenzierbar mit Jakobi-Matrix

$$\partial G := \begin{pmatrix} \frac{\partial}{\partial F_{k,l}} C_{i,j} & (\mathbb{I}_{p^2} - (F \otimes F^\tau))^{-1} & \mathbb{I}_{q^2} \\ \frac{\partial}{\partial F_{k,l}} [FC]_{i,j} & (F \otimes \mathbb{I}_p)(\mathbb{I}_{p^2} - (F \otimes F^\tau))^{-1} & 0 \\ \frac{\partial}{\partial F_{k,l}} [F^2 C]_{i,j} & (F^2 \otimes \mathbb{I}_p)(\mathbb{I}_{p^2} - (F \otimes F^\tau))^{-1} & 0 \end{pmatrix} \quad (10.91)$$

Ist $\text{rk } Q = \text{rk } F = \text{rk } Z = p$ und $\text{rk } V = q$, so ist ∂G invertierbar.

Wir halten wieder den eindimensionalen Fall fest:

$$\partial G = \begin{pmatrix} \frac{2FZ^2Q}{(1-F^2)^2} & \frac{Z^2}{1-F^2} & 1 \\ \frac{Z^2Q}{1-F^2} + \frac{2F^2Z^2Q}{(1-F^2)^2} & \frac{FZ^2}{1-F^2} & 0 \\ \frac{2FZ^2Q}{1-F^2} + \frac{2F^3Z^2Q}{(1-F^2)^2} & \frac{F^2Z^2}{1-F^2} & 0 \end{pmatrix}, \quad \det \partial G = \frac{Z^2F^4Q}{(1-F^2)^2}.$$

Zum Beweis brauchen wir folgendes

Lemma 10.4.15

$$\begin{aligned} \frac{\partial}{\partial F_{k,l}} [F^\nu C]_{i,j} = & (\mathbb{I}_p \otimes C) \sum_{s=0}^{\nu-1} F^{\nu-1-s} \otimes (F^s)^\tau + \\ & + (\mathbb{I}_{p^2} - (F \otimes F^\tau))^{-1} (F^\nu \otimes F^\tau C) + \\ & + (F^{\nu+1} C \otimes \mathbb{I}_p) (\mathbb{I}_{p^2} - (F \otimes F^\tau))^{-1} \end{aligned} \quad (10.92)$$

BEWEIS VON LEMMA 10.4.15: Produktregel. ////

BEWEIS VON PROPOSITION 10.4.14: Zieht man von den “ γ_3 ”-Zeilen das $F \otimes \mathbb{I}_p$ -fache der “ γ_2 ”-Zeilen ab, so erhalten wir

$$\det \partial G = \det \begin{pmatrix} Z^{\otimes 2} \frac{\partial}{\partial F_{k,l}} C_{i,j} & Z^{\otimes 2} (\mathbb{I}_{p^2} - (F \otimes F^\tau))^{-1} & \mathbb{I}_{q^2} \\ \frac{\partial}{\partial F_{k,l}} [FC]_{i,j} & (F \otimes \mathbb{I}_p) (\mathbb{I}_{p^2} - (F \otimes F^\tau))^{-1} & 0 \\ \mathbb{I}_p \otimes CF^\tau & 0 & 0 \end{pmatrix} \neq 0$$

mit der Konjugation mit D_S erreichen wir dann die gewünschten Dimensionen. ////

Da Q und V per Definition symmetrisch sind, müssen wir in der Parametrisierung — wie auch bei der Berechnung der Fisher-Information — die Dimension mit der linearen Abbildung D_S definiert als

$$\begin{aligned} D_S : \quad \mathbb{R}^{p^2+p^2+q^2} & \rightarrow \mathbb{R}^{p^2+p^2/2+p+q^2/2+q} \\ X = \begin{pmatrix} A_{i,j=1,\dots,p} \\ B_{i,j=1,\dots,p} \\ C_{i,j=1,\dots,q} \end{pmatrix} & \mapsto D_S(X) = \begin{pmatrix} A'_{i,j=1,\dots,p} \\ B'_{i \leq j=1,\dots,p} \\ C'_{i \leq j=1,\dots,q} \end{pmatrix} = \\ & = \begin{pmatrix} A_{i,j=1,\dots,p} \\ \left[\frac{(B_{i,j}+B_{j,i})}{2} \right]_{i \leq j=1,\dots,p} \\ \left[\frac{(C_{i,j}+C_{j,i})}{2} \right]_{i \leq j=1,\dots,q} \end{pmatrix}, \end{aligned} \quad (10.93)$$

noch auf Dimension $d := (3p^2 + p + q^2 + q)/2$ reduzieren und Q und V mit $\text{svec}(Q)$, $\text{svec}(V)$ parametrisieren.

10.4.7 Beweis von Theorem 10.4.6

BEWEIS ZU THEOREM 10.4.6: Wir wenden Theorem 10.4.3 auf die stationären, ergodischen Beobachtungen $R_t := (Y_t^\tau, Y_{t-1}^\tau, Y_{t-2}^\tau)^\tau$ an, wobei wir wegen der Zentriertheit die Abhängigkeit von μ in θ nicht berücksichtigen. Zu zeigen ist nach dem Kriterium aus Theorem 10.4.2

$$\det[\text{tr} \left(\Sigma^{-\frac{1}{2}} \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-\frac{1}{2}} \right) \left(\Sigma^{-\frac{1}{2}} \frac{\partial \Sigma}{\partial \theta_j} \Sigma^{-\frac{1}{2}} \right)]_{1 \leq i, j \leq d} > 0 \quad (10.94)$$

Dies ist sicher dann der Fall, wenn $\text{rk} \Sigma = d$ und $\text{rk} \frac{\partial \Sigma}{\partial \theta_i} \frac{\partial \Sigma}{\partial \theta_j} = d$, wobei wir in unserer Notation $\Sigma \rightsquigarrow \mathcal{T}$ zu $m = 2$ und $\frac{\partial \Sigma}{\partial \theta_i} \frac{\partial \Sigma}{\partial \theta_j} \rightsquigarrow [D_S \partial G|_\theta D_S^\tau]^{-1}$ identifizieren. Dies haben wir in Proposition 10.4.10 und in Proposition 10.4.14 gezeigt.

////

Bemerkung 10.4.16 Auf Basis der ACF arbeitet auch Aoki (1990), um die Hyper-Parameter eines normalen, linearen, zeitinvarianten, und stationären Zustandsraummodells aus den Beobachtungen zu schätzen; er verwendet allerdings die Innovationsdarstellung des Zustandsraummodells, in unserer Notation

$$\tilde{\beta}_{t+1} = \beta_{t+1|t} = F\beta_{t|t-1} + \Gamma\zeta_t \quad (10.95)$$

$$y_t = Z\tilde{\beta}_t + \zeta_t \quad (10.96)$$

mit $\Gamma = F \lim_t M_t$, $\mathcal{L}(\zeta_t) = \lim_t \mathcal{L}(\Delta y_t)$. Aoki schätzt dann die Hyper-Parameter F , G , Z und $\Delta := \lim_t \text{Cov}(\Delta y_t)$.

10.5 Beispiel: zeitinvariantes, lineares, normales Zustandsraummodell

10.5.1 L_2 -Differenzierbarkeit des normalen Modells

Als wichtigen Spezialfall für Modell (55)–(57) betrachten wir nun den Fall, daß $G = \mathcal{N}_p(0, \mathbb{I}_p)$, $H = \mathcal{N}_q(0, \mathbb{I}_q)$, und $\tilde{Q}_0 = \tilde{Q}$. Offenbar erfüllen diese Verteilungen die Endlichkeitsvoraussetzungen (B.9) und (B.10). Im folgenden nehmen wir zusätzlich die Regularität der Fisher-Information des Modells mit einer Filtrierung \mathcal{G}_T gemäß (58) an, wofür wir in Theorem 10.4.6 hinreichende Bedingungen gegeben haben. Damit ist Theorem 10.3.2 in Kraft.

Da uns im normalen Zustandsraummodell eher die Varianzen $Q = \tilde{Q}^2$ und $V = \tilde{V}^2$ interessieren, parametrisieren wir das Modell schlicht mit der stetig differenzierbaren, injektiven Parameter-Transformation $\tau(F, \tilde{Q}, \tilde{V}) = (F, Q, V)$ um. Da die Jakobische vollen Rang hat, ändert dies nichts an der L_2 -Differenzierbarkeit des Modells.

10.5.2 L_2 -Ableitung

Ableiten der Loglikelihood im vollen Modell liefert mit Ableitung mit den Abkürzungen $\tilde{v}_s = Q^{-1}(\beta_s - F\beta_{s-1})$, $\tilde{\varepsilon}_s = V^{-1}\varepsilon_s$

$$\Lambda_{\theta, s} = (\Lambda_{\text{vec}(F), s}, \Lambda_{\text{vec}(Q), s}, \Lambda_{\text{vec}(V), s})^\tau, \quad (10.97)$$

$$\Lambda_{\text{vec}(F),s}(\tilde{v}_s, \beta_{s-1}) = -\tilde{v}_s \beta_{s-1}^\tau \quad (10.98)$$

$$2\Lambda_{\text{svec}(Q),s}(\tilde{v}_s) = \tilde{v}_s \tilde{v}_s^\tau - Q^{-1} \quad (10.99)$$

$$2\Lambda_{\text{svec}(V),s}(\varepsilon_s) = \tilde{\varepsilon}_s \tilde{\varepsilon}_s^\tau - V^{-1} \quad (10.100)$$

Beim Übergang zur größeren Filtrierung, bei der nur noch die Beobachtungen y_s zur Verfügung stehen, müßten wir eigentlich laut Theorem 10.4.6 die Zweischritt-Glätter $E[\Lambda_{[s]\text{vec}(\cdot),s|s+2}]$ verwenden. Es gilt im Fall der Normalverteilung jedoch

Proposition 10.5.1 *Sind im normalen Zustandsraummodell folgende Voraussetzungen an die Hyper-Parameter F , Q , V erfüllt*

$$\rho(F) < 1 \quad (10.101)$$

$$\text{rk } Z \geq \text{rk } F = \text{rk } Q = p, \quad \text{rk } V = q \quad (10.102)$$

so ist bei Verwendung der Einschnitt-Glättungsfiltrierung \mathcal{G}_T^1 , in der Notation von (58), geschrieben als $\mathcal{G}_T^1 = (\mathcal{F}_{T,0,\min\{s+1,T\}})$, die Fisher-Information \mathcal{I}_θ regulär.

Vor dem Beweis zeigen wir noch ein Lemma, das die Terme $\beta_{s|s+k}$, $v_{s|s+k}$, $\varepsilon_{s|s+k}$ zu beliebigem Lag k in Abhängigkeit des Δy_s -Prozesses darstellt.

Lemma 10.5.2 *Seien $\Sigma_{s|s[-1]}$ die aus dem klassischen Kalman-Filter bekannten Matrizen und M_s die entsprechenden Kalman-Gains.*

Definieren wir

$$\Delta_s = \text{Cov}(\Delta y_s) = Z \Sigma_{s|s-1} Z^\tau + V \quad (10.103)$$

und

$$J_{s:s-1} := \mathbb{I}_p, \quad J_{s:s} := J_s := \Sigma_{s|s} F^\tau \Sigma_{s+1|s}^{-1}, \quad J_{r:s} = J_{r:s-1} J_s, \quad (10.104)$$

und davon ausgehend für $k \geq 1 - s$

$$A_{s,k} = (\mathbf{I}_{(k \geq 0)} J_{s-1:s+k-1} + \mathbf{I}_{(k < 0)} F^{-1-k}) M_{s+k}, \quad (10.105)$$

$$B_{s,k} = \mathbf{I}_{(k=0)} Z^\tau \Delta_s^{-1} + \mathbf{I}_{(k > 0)} \Sigma_{s|s-1}^{-1} J_{s:s+k} M_{s+k}, \quad (10.106)$$

$$C_{s,k} = \mathbf{I}_{(k=0)} \Delta_s^{-1} - \mathbf{I}_{(k > 0)} V^{-1} Z J_{s:s+k} M_{s+k}, \quad (10.107)$$

so erhalten wir

$$\beta_{s-1|s+k} = \sum_{l=1-s}^k A_{s,l} \Delta y_{s+l}, \quad (10.108)$$

$$\tilde{v}_{s|s+k} = \sum_{l=0}^k B_{s,l} \Delta y_{s+l}, \quad (10.109)$$

$$\tilde{\varepsilon}_{s|s+k} = \sum_{l=0}^k C_{s,l} \Delta y_{s+l}, \quad (10.110)$$

wobei wir $\sum_{l=0}^k [\dots]$ für $k < 0$ als 0 lesen.

BEWEIS ZU LEMMA 10.5.2: Für $k < 0$ gilt $\beta_{s-1|s+k} = F^{-1-k} \beta_{s+k|s+k}$. Nun ist $\beta_{1|1} = M_1 \Delta y_1$, und damit per Induktion

$$\beta_{t|t} = \beta_{t|t-1} + M_t \Delta y_t = F \beta_{t-1|t-1} + M_t \Delta y_t \stackrel{\text{I.V.}}{=} \sum_{r=1}^t F^{t-r} M_r \Delta y_r,$$

so daß

$$\begin{aligned}\beta_{s-1|s+k} &= F^{-1-k} \sum_{r=1}^{s+k} F^{s+k-r} M_r \Delta y_r = \\ &= \sum_{r=1-s}^k F^{-1-r} M_{s+r} \Delta y_{s+r},\end{aligned}$$

also die Behauptung für $A_{s,k}$. Für $k > 0$ setzen wir mit $\tilde{s} = s - r$ und zeigen zunächst mit Induktion, daß

$$\Delta\beta_{s|s} = \beta_s - \beta_{s|s} = a_{r,s} \Delta\beta_{\tilde{s}|\tilde{s}} + \sum_{l=1}^{r-1} b_{l,s} v_{s-l} + c_{l,s} \varepsilon_{s-l}$$

mit

$$a_{r,s} = \Sigma_{s|s} J_{\tilde{s}:s-1}^T \Sigma_{\tilde{s}|\tilde{s}}^{-1}, \quad b_{r,s} = a_{r,s} \Sigma_{\tilde{s}|\tilde{s}} \Sigma_{\tilde{s}|\tilde{s}-1}^{-1}, \quad c_{r,s} = -a_{r,s} M_{\tilde{s}}.$$

Für $r = 0$ ist nichts zu zeigen. Sei also die Behauptung schon für r gezeigt. Dann ist

$$\begin{aligned}\Delta\beta_{s|s} &\stackrel{\text{I.V.}}{=} a_{r,s} \Delta\beta_{\tilde{s}|\tilde{s}} + \sum_{l=1}^{r-1} b_{l,s} v_{s-l} + c_{l,s} \varepsilon_{s-l} = \\ &= a_{r,s} (\mathbb{I}_p - M_{\tilde{s}} Z) \Delta\beta_{\tilde{s}|\tilde{s}-1} - a_{r,s} M_{\tilde{s}} \varepsilon_{\tilde{s}} + \sum_{l=1}^{r-1} b_{l,s} v_{s-l} + c_{l,s} \varepsilon_{s-l} = \\ &= a_{r,s} (\mathbb{I}_p - M_{\tilde{s}} Z) F \Delta\beta_{\tilde{s}-1|\tilde{s}-1} + a_{r,s} (\mathbb{I}_p - M_{\tilde{s}} Z) v_{\tilde{s}} - c_{r,s} \varepsilon_{\tilde{s}} + \\ &\quad + \sum_{l=1}^{r-1} b_{l,s} v_{s-l} + c_{l,s} \varepsilon_{s-l} = \\ &= a_{r,s} \Sigma_{\tilde{s}|\tilde{s}} \Sigma_{\tilde{s}|\tilde{s}-1}^{-1} F \Delta\beta_{\tilde{s}-1|\tilde{s}-1} + a_{r,s} \Sigma_{\tilde{s}|\tilde{s}} \Sigma_{\tilde{s}|\tilde{s}-1}^{-1} v_{\tilde{s}} \\ &\quad - a_{r,s} M_{\tilde{s}} \varepsilon_{\tilde{s}-1} + \sum_{l=1}^{r-1} b_{l,s} v_{s-l} + c_{l,s} \varepsilon_{s-l} = \\ &= a_{r,s} \Sigma_{\tilde{s}|\tilde{s}} J_{\tilde{s}-1}^T \Sigma_{\tilde{s}-1|\tilde{s}-1}^{-1} \Delta\beta_{\tilde{s}-1|\tilde{s}-1} + \sum_{l=1}^r b_{l,s} v_{s-l} + c_{l,s} \varepsilon_{s-l} = \\ &= a_{r+1,s} \Delta\beta_{\tilde{s}-1|\tilde{s}-1} + \sum_{l=1}^r b_{l,s} v_{s-l} + c_{l,s} \varepsilon_{s-l}\end{aligned}$$

Setzen wir nun

$$\begin{aligned}\tilde{a}_{r,s} &= \mathbb{I}_{(r>0)} Z F a_{r-1,s-1}, \\ \tilde{b}_{r,s} &= \mathbb{I}_{(r=0)} Z + \mathbb{I}_{(r>0)} (Z F b_{r-1,s-1}), \\ \tilde{c}_{r,s} &= \mathbb{I}_{(r=0)} \mathbb{I}_q + \mathbb{I}_{(r>0)} (Z F c_{r-1,s-1}),\end{aligned}$$

so erhalten wir für $r \geq 1$

$$\Delta y_s = \tilde{a}_{r,s} \Delta\beta_{\tilde{s}|\tilde{s}} + \sum_{l=0}^{r-1} \tilde{b}_{l,s} v_{s-l} + \tilde{c}_{l,s} \varepsilon_{s-l},$$

denn

$$\Delta y_s = Z F \Delta\beta_{s-1|s-1} + \varepsilon_s + Z v_s.$$

Damit gilt für die Kovarianz der gemeinsamen Normalverteilung von $(\Delta\beta_{s-1|s-1}^T, v_s^T, \varepsilon_s^T, \Delta y_{s+k}^T)^T$

$$\text{Cov}\left(\begin{pmatrix} \Delta\beta_{s-1|s-1} \\ v_s \\ \varepsilon_s \\ \Delta y_{s+k} \end{pmatrix}\right) = \begin{pmatrix} \Sigma_{s-1|s-1} & 0 & 0 & \Sigma_{s-1|s-1} \tilde{a}_{k+1}^T \\ 0 & Q & 0 & Q \tilde{b}_k^T \\ 0 & 0 & V & V \tilde{c}_k^T \\ \tilde{a}_{k+1} \Sigma_{s-1|s-1} & \tilde{b}_k Q & \tilde{c}_k V & \Delta_{s+k} \end{pmatrix},$$

wobei der Index $s+k$ in den Termen \tilde{a} , \tilde{b} , \tilde{c} weggelassen wurde. Somit gilt nach Lemma A.2.1 für

$$\begin{aligned}
\mathbb{E}[\Delta\beta_{s-1|s-1}|\Delta y_{s+k}] &= \Sigma_{s-1|s-1}\tilde{a}_{k+1}^{\tau}\Delta_{s+k}^{-1}\Delta y_{s+k} = \\
&= J_{s-1:s+k-1}M_{s+k}\Delta y_{s+k} = A_{s,k}\Delta y_{s+k} \\
\mathbb{E}[v_s|\Delta y_{s+k}] &= Q\{\mathbb{I}_{(k=0)}Z + \mathbb{I}_{(k>0)}\tilde{b}_k^{\tau}\}\Delta_{s+k}^{-1}\Delta y_{s+k} = \\
&= Q\{\mathbb{I}_{(k=0)}Z + \mathbb{I}_{(k>0)}\Sigma_{s|s-1}^{-1}J_{s:s+k}M_{s+k}\}\Delta y_{s+k} \\
\mathbb{E}[\tilde{v}_s|\Delta y_{s+k}] &= \{\mathbb{I}_{(k=0)}Z + \mathbb{I}_{(k>0)}\Sigma_{s|s-1}^{-1}J_{s:s+k}M_{s+k}\}\Delta y_{s+k} = \\
&= B_{s,k}\Delta y_{s+k} \\
\mathbb{E}[\varepsilon_s|\Delta y_{s+k}] &= V\{\mathbb{I}_{(k=0)}\mathbb{I}_q + \mathbb{I}_{(k>0)}\tilde{c}_k^{\tau}\}\Delta_{s+k}^{-1}\Delta y_{s+k} = \\
&= V\{\mathbb{I}_{(k=0)}\mathbb{I}_q - \mathbb{I}_{(k>0)}M_s^{\tau}\Sigma_{s|s}^{-1}J_{s:s+k}M_{s+k}\}\Delta y_{s+k} = \\
&= V\{\mathbb{I}_{(k=0)}\mathbb{I}_q - \mathbb{I}_{(k>0)}V^{-1}ZJ_{s:s+k}M_{s+k}\}\Delta y_{s+k} \\
\mathbb{E}[\tilde{\varepsilon}_s|\Delta y_{s+k}] &= \{\mathbb{I}_{(k=0)}\mathbb{I}_q - \mathbb{I}_{(k>0)}V^{-1}ZJ_{s:s+k}M_{s+k}\}\Delta y_{s+k} = \\
&= C_{s,k}\Delta y_{s+k}
\end{aligned}$$

und weil $\mathbb{E}[\beta_{s-1|s-1}|\Delta y_{s+k}] = 0$, folgt $\mathbb{E}[\beta_{s-1}|\Delta y_{s+k}] = A_{s,k}\Delta y_{s+k}$, und, weil die Δy_s orthogonal sind, ist die Behauptung zu $A_{s,l}$ für $l = (1-s), \dots, k$ gezeigt. Für $B_{s,l}$ und $C_{s,l}$ ist für $s < 0$ nichts zu zeigen, denn v_s und ε_s sind unabhängig von $\sigma(\Delta y_1, \dots, \Delta y_{s-1})$. ////

Damit ergeben sich¹¹

$$\beta_{s-1|s+1} = \beta_{s-1|s-1} + \Sigma_{s-1|s-1}F^{\tau}Z^{\tau}\Delta_s^{-1}\Delta y_s + [\Sigma_{s-1|s-1}F^{\tau}\Sigma_{s|s-1}^{-1}\Sigma_{s|s}F^{\tau}Z^{\tau}]\Delta_{s+1}^{-1}\Delta y_{s+1} \quad (10.111)$$

$$\tilde{v}_{s|s+1} = Q^{-1}(\beta_{s|s+1} - F\beta_{s-1|s+1}) = \quad (10.112)$$

$$= Z^{\tau}\Delta_s^{-1}\Delta y_s + [\Sigma_{s|s-1}^{-1}\Sigma_{s|s}F^{\tau}Z^{\tau}]\Delta_{s+1}^{-1}\Delta y_{s+1} \quad (10.113)$$

$$\tilde{\varepsilon}_{s|s+1} = V^{-1}(y_s - Z\beta_{s|s+1}) = \quad (10.114)$$

$$= \Delta_s^{-1}\Delta y_s + [V^{-1}Z\Sigma_{s|s}F^{\tau}Z^{\tau}]\Delta_{s+1}^{-1}\Delta y_{s+1}, \quad (10.115)$$

und wir erhalten die uns interessierenden L_2 -Ableitungen.

$$\Lambda_{F,s|s+1}(y_{1:(s+1)}) = -\tilde{v}_{s|s+1}\beta_{s-1|s+1}^{\tau} + R_{s|s+1} \quad (10.116)$$

$$2\Lambda_{Q,s|s+1}(y_{1:(s+1)}) = \tilde{v}_{s|s+1}\tilde{v}_{s|s+1}^{\tau} - Q_{s|s+1} \quad (10.117)$$

$$2\Lambda_{V,s|s+1}(y_{1:(s+1)}) = \tilde{\varepsilon}_{s|s+1}\tilde{\varepsilon}_{s|s+1}^{\tau} - V_{s|s+1} \quad (10.118)$$

$$R_{s|s+1} = \Sigma_{s-1|s-1} + \Sigma_{s-1|s-1}F^{\tau}Z^{\tau}\Delta_s^{-1}ZF\Sigma_{s-1|s-1} + [\Sigma_{s-1|s-1}F^{\tau}\Sigma_{s|s-1}^{-1}\Sigma_{s|s}F^{\tau}Z^{\tau}]\Delta_{s+1}^{-1}[\dots]^{\tau} \quad (10.119)$$

$$Q_{s|s+1} = Z^{\tau}\Delta_s^{-1}Z + [\Sigma_{s|s-1}^{-1}\Sigma_{s|s}F^{\tau}Z^{\tau}]\Delta_{s+1}^{-1}[\dots]^{\tau} \quad (10.120)$$

$$V_{s|s+1} = \Delta_s^{-1} + [V^{-1}Z\Sigma_{s|s}F^{\tau}Z^{\tau}]\Delta_{s+1}^{-1}[\dots]^{\tau} \quad (10.121)$$

wobei $\Sigma_{s|s[-1]}$ die aus dem Kalman-Filter bekannten Matrizen sind. Eine Zentrierung dieser Terme wie in Theorem 10.2.10 ist hier nicht notwendig:

¹¹den Randfall $s = T$ betrachten wir nicht gesondert, sondern notieren nur, daß jeweils anstelle von $T+1$ der Index T zu setzen ist.

Weil, wie in Lemma 6.3.4 gezeigt, $\sigma(\{\Delta y_r\}_{r \leq s}) = \sigma(\{y_r\}_{r \leq s})$ gilt, und weil Δy_s stochastisch unabhängig sind, sind die nach $y_{1:s-1} = y_{1:s-1}^{\natural}$ bedingten Erwartungswerte von $\Lambda_{Q,s|s+1}$ und $\Lambda_{V,s|s+1}$ gerade die unbedingten, also 0; bei $\Lambda_{F,s|s+1}$ können wir $\beta_{s-1|s+1}$ schreiben als $\beta_{s-1|s-1} + g(\Delta y_s, \Delta y_{s+1})$ mit einer gewissen bilinearen Abbildung g . Nur dieses g trägt zu $R_{s|s+1}$ bei und es gilt mit $\bar{\mathcal{F}}_{s-1} = \sigma(y_{1:s-1} = y_{1:s-1}^{\natural})$

$$\begin{aligned} \mathbb{E}[\Lambda_{F,s|s+1} | \bar{\mathcal{F}}_{s-1}] &= \mathbb{E}[v_{s|s+1}(\beta_{s-1|s-1} + g(\Delta y_s, \Delta y_{s+1}))^\tau | \bar{\mathcal{F}}_{s-1}] - R_{s|s+1} = \\ &= \mathbb{E}[v_{s|s+1} | \bar{\mathcal{F}}_{s-1}] \beta_{s-1|s-1}^\tau + \mathbb{E}[v_{s|s+1} g(\Delta y_s, \Delta y_{s+1})^\tau | \bar{\mathcal{F}}_{s-1}] - R_{s|s+1} = \\ &= \mathbb{E}[v_{s|s+1}] \beta_{s-1|s-1}^\tau + \mathbb{E}[v_{s|s+1} g(\Delta y_s, \Delta y_{s+1})^\tau] - R_{s|s+1} = 0 \end{aligned}$$

BEWEIS ZU PROPOSITION 10.5.1: Zerlegen wir jede der Komponenten von $\Lambda_{s|s+1}$ in die Projektionen auf die orthogonalen Komponenten $\beta_{s-1|s-1}$, Δy_s und Δy_{s+1} , so können wir zunächst schreiben

$$\begin{aligned} (\beta_{s-1|s+1}^\tau, v_{s|s+1}^\tau, \varepsilon_{s|s+1}^\tau)^\tau &= K(\tilde{Y}_1^\tau, \tilde{Y}_2^\tau, \tilde{Y}_3^\tau)^\tau, \\ K &= \begin{pmatrix} \Sigma_{s-1|s-1}^{-\frac{1}{2}}, & \Sigma_{s-1|s-1} F^\tau Z^\tau \Delta_s^{-\frac{1}{2}}, & [\Sigma_{s-1|s-1} F^\tau \Sigma_{s|s-1}^{-1} \Sigma_{s|s} F^\tau Z^\tau] \Delta_{s+1}^{-\frac{1}{2}} \\ 0, & Z^\tau \Delta_s^{-\frac{1}{2}}, & [\Sigma_{s|s-1}^{-1} \Sigma_{s|s} F^\tau Z^\tau] \Delta_{s+1}^{-\frac{1}{2}} \\ 0, & \Delta_s^{-\frac{1}{2}}, & [V^{-1} Z \Sigma_{s|s} F^\tau Z^\tau] \Delta_{s+1}^{-\frac{1}{2}} \end{pmatrix} \end{aligned}$$

mit $\tilde{Y} = (\tilde{Y}_1^\tau, \tilde{Y}_2^\tau, \tilde{Y}_3^\tau)^\tau \sim \mathcal{N}_d(0, \mathbb{I}_d)$, wobei $d = p^2 + \binom{p}{2} + \binom{q}{2}$.

Die Matrix K dieser Lineartransformation hat vollen Rang, wie wir leicht feststellen, indem wir von der zweiten Zeile das Z^τ -fache der letzten abziehen und notieren, daß $(\Sigma_{s|s-1}^{-1} - Z^\tau V^{-1} Z) \Sigma_{s|s} = \mathbb{I}_p$ ist. Mit Korollar B.5.2 folgt, daß die Kovarianz von $\text{svec}[K\tilde{Y}(K\tilde{Y})^\tau]$ vollen Rang hat. Schreiben wir nun $\text{svec}[K\tilde{Y}(K\tilde{Y})^\tau]$ aus zu

$$\begin{aligned} \text{svec}[K\tilde{Y}(K\tilde{Y})^\tau]^\tau &= \left(\text{svec}[\beta\beta^\tau]^\tau, \text{vec}[\beta v^\tau]^\tau, \text{svec}[\beta\varepsilon^\tau]^\tau, \right. \\ &\quad \left. \text{svec}[v v^\tau]^\tau, \text{vec}[v\varepsilon^\tau]^\tau, \text{svec}[\varepsilon\varepsilon^\tau]^\tau \right), \end{aligned}$$

wobei wir die Indizes $s|s+1$ bzw. $s-1|s+1$ weggelassen haben, so folgt die Behauptung mit der Projektion auf die zweite, vierte und sechste ‘‘Zeile’’. ///

Bemerkung 10.5.3 Wie unmittelbar im Beweis zu Proposition 10.5.1 klar wird, genügt die Filtrierung mit ‘‘reinen Filter’’-Termen $\Lambda_{s|s}$ nicht für eine Fisher-Information mit vollem Rang:

Eine solche Filtrierung hätte zur Folge, daß die letzte ‘‘Spalte’’ von K wegfiel, und da $v_{s|s} = Z^\tau \varepsilon_{s|s}$ könnte man Q und V nicht simultan identifizieren.

Dies ist auch anschaulich klar, denn erst Beobachtungen Y_r , $r \geq s$ tragen Information über v_s und ε_s ; anhand der gerade gemachten Beobachtung Y_s aber kann man nicht unterscheiden, ob eine große Auslenkung ($|\Delta Y_s|$ groß) auf ein großes Q oder ein großes V zurückzuführen ist.

10.5.3 Fisher-Information

In diesem Abschnitt leiten wir — gleich in allgemeiner Form — die Fisher-Information bei Verwendung der m -Schritt-Glättungsfiltrierung¹² \mathcal{G}_T^m her. Da-

¹²in der Notation von (58) geschrieben als $\mathcal{G}_T^m = (\mathcal{F}_{T,0,\min\{s+m,T\}})$

zu verwenden wir folgendes Lemma, dessen Beweis klar ist und hier nicht dargestellt wird.

Lemma 10.5.4 Seien $U^{(s)} \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}_q(0, \mathbb{I}_q)$ für $s = 1, \dots, T$ und weitere Matrizen $A^{(l,i)} \in \mathbb{R}^{r_i \times q}$, für $i = 1, 2, 3, 4$ und $l = 1, \dots, T$, sowie abgeleitete normalverteilte Variablen $X^{(i)} := \sum_{l=1}^T A^{(l,i)} U^{(l)}$ und $X^{(1,2)} := X^{(1)} X^{(2)\tau}$, $X^{(3,4)} := X^{(3)} X^{(4)\tau}$. Dann ist

$$\begin{aligned} \text{Cov}(X_{i,j}^{(1,2)} X_{k,l}^{(3,4)}) &= \left[\sum_{l=1}^T A^{(l,1)} A^{(l,3)\tau} \right]_{i,k} \left[\sum_{l=1}^T A^{(l,2)} A^{(l,4)\tau} \right]_{j,l} + \\ &+ \left[\sum_{l=1}^T A^{(l,1)} A^{(l,4)\tau} \right]_{i,l} \left[\sum_{l=1}^T A^{(l,2)} A^{(l,3)\tau} \right]_{j,k} \end{aligned} \quad (10.122)$$

Proposition 10.5.5 Die Fisher-Information des linearen, stationären, normalen, zeitinvarianten Zustandsraummodells \mathcal{I}_θ schreibt sich bei Verwendung von \mathcal{G}_T^m als

$$\mathcal{I}_\theta = D_S \tilde{\mathcal{I}}_\theta D_S^\tau \quad (10.123)$$

mit der linearen Abbildung D_S aus Definition (93) und

$$\tilde{\mathcal{I}}_\theta = \begin{pmatrix} \tilde{\mathcal{I}}_{F,F} & \tilde{\mathcal{I}}_{F,Q} & \tilde{\mathcal{I}}_{F,V} \\ \tilde{\mathcal{I}}_{F,Q}^\tau & \tilde{\mathcal{I}}_{Q,Q} & \tilde{\mathcal{I}}_{Q,V} \\ \tilde{\mathcal{I}}_{F,V}^\tau & \tilde{\mathcal{I}}_{Q,V}^\tau & \tilde{\mathcal{I}}_{V,V} \end{pmatrix}. \quad (10.124)$$

Dabei ist

$$\begin{aligned} \tilde{\mathcal{I}}_{F,F,i,j,k,l} &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{s=1}^T [F_{i,k}^{(s)} G_{j,l}^{(s)} + H_{i,l}^{(s)\tau} H_{j,k}^{(s)}] \\ \tilde{\mathcal{I}}_{F,Q,i,j,k,l} &= \lim_{T \rightarrow \infty} \frac{1}{2T} \sum_{s=1}^T [F_{i,k}^{(s)} H_{j,l}^{(s)} + F_{i,l}^{(s)} H_{j,k}^{(s)}] \\ \tilde{\mathcal{I}}_{F,V,i,j,k,l} &= \lim_{T \rightarrow \infty} \frac{1}{2T} \sum_{s=1}^T [I_{i,k}^{(s)} J_{j,l}^{(s)} + I_{i,l}^{(s)} J_{j,k}^{(s)}] \\ \tilde{\mathcal{I}}_{Q,Q,i,j,k,l} &= \lim_{T \rightarrow \infty} \frac{1}{4T} \sum_{s=1}^T [F_{i,k}^{(s)} F_{j,l}^{(s)} + F_{i,l}^{(s)} F_{j,k}^{(s)}] \\ \tilde{\mathcal{I}}_{Q,V,i,j,k,l} &= \lim_{T \rightarrow \infty} \frac{1}{4T} \sum_{s=1}^T [I_{i,k}^{(s)} I_{j,l}^{(s)} + I_{i,l}^{(s)} I_{j,k}^{(s)}] \\ \tilde{\mathcal{I}}_{V,V,i,j,k,l} &= \lim_{T \rightarrow \infty} \frac{1}{4T} \sum_{s=1}^T [K_{i,k}^{(s)} K_{j,l}^{(s)} + K_{i,l}^{(s)} K_{j,k}^{(s)}] \end{aligned}$$

und

$$\begin{aligned} F^{(s)} &= \sum_{l=0}^{m \wedge (T-s)} B_{s,l} \Delta_s B_{s,l}^\tau, & G^{(s)} &= \sum_{l=1-s}^{m \wedge (T-s)} A_{s,l} \Delta_s A_{s,l}^\tau, \\ H^{(s)} &= \sum_{l=0}^{m \wedge (T-s)} A_{s,l} \Delta_s B_{s,l}^\tau, & I^{(s)} &= \sum_{l=0}^{m \wedge (T-s)} B_{s,l} \Delta_s C_{s,l}^\tau \end{aligned}$$

$$J^{(s)} = \sum_{l=0}^{m \wedge (T-s)} A_{s,l} \Delta_s C_{s,l}^T, \quad K^{(s)} = \sum_{l=0}^{m \wedge (T-s)} C_{s,l} \Delta_s C_{s,l}^T$$

mit den Ausdrücken aus (105), (106) und (107).

BEWEIS: Unter Normalverteilungsannahmen sind die Δy_t stochastisch unabhängig normalverteilt. Wir wenden daher Lemma 10.5.4 an mit den Identifikationen

$$\begin{aligned} A^{l,1} &\rightsquigarrow B_{s,l} \Delta_{s+l}^{\frac{1}{2}} \mathbf{I}_{\{0 \leq l \leq (s+m) \wedge T\}}, & A^{l,2} &\rightsquigarrow A_{s,l} \Delta_{s+l}^{\frac{1}{2}} \mathbf{I}_{\{1-s \leq l \leq (s+m) \wedge T\}} \\ A^{l,3} &\rightsquigarrow B_{s,l} \Delta_{s+l}^{\frac{1}{2}} \mathbf{I}_{\{0 \leq l \leq (s+m) \wedge T\}}, & A^{l,4} &\rightsquigarrow A_{s,l} \Delta_{s+l}^{\frac{1}{2}} \mathbf{I}_{\{1-s \leq l \leq (s+m) \wedge T\}} \end{aligned}$$

für den Summanden Nummer s in $\tilde{\mathcal{I}}_{F,F}$, und entsprechend für die weiteren Blöcke. ////

10.5.4 bedingter ML-Schätzer

Berechnet man nun den (bedingten) ML-Schätzer, so erhält man, indem man $\tilde{s} = (s+1) \wedge T$ setzt und die Terme

$$\Sigma_{s|\tilde{s}} = \text{Cov}(\beta_{s|\tilde{s}}|\tilde{s}) = \Sigma_{s|s} + J_s(\Sigma_{s+1|\tilde{s}} - \Sigma_{s+1|s})J_s^T \quad (10.125)$$

$$\Sigma_{s,s-1|s} = \text{Cov}(\beta_{s|s}, \beta_{s-1|s}|s) = \Sigma_{s|s} J_{s-1}^T \quad (10.126)$$

$$\begin{aligned} \Sigma_{s,s-1|\tilde{s}} &= \text{Cov}(\beta_{s|\tilde{s}}, \beta_{s-1|\tilde{s}}|\tilde{s}) = \\ &= \Sigma_{s|s} J_{s-1}^T + J_s(\Sigma_{s+1|\tilde{s}} J_s^T - F \Sigma_{s|s}) J_{s-1}^T \end{aligned} \quad (10.127)$$

verwendet,

$$\hat{F}_T := \left(\sum_{s=1}^T \beta_{s|\tilde{s}} \beta_{s-1|\tilde{s}}^T + R_{s|\tilde{s}} \right) \left(\sum_{s=1}^T \beta_{s-1|\tilde{s}} \beta_{s-1|\tilde{s}}^T + \Sigma_{s-1|\tilde{s}} \right)^{-1} \quad (10.128)$$

$$\hat{Q}_T := \frac{1}{T} \left(\sum_{s=1}^T v_{s|\tilde{s}} \beta_{s|\tilde{s}}^T + \Sigma_{s|\tilde{s}} - \hat{F}_T \Sigma_{s,s-1|\tilde{s}} \right) \quad (10.129)$$

$$\hat{V}_T := \frac{1}{T} \left(\sum_{s=1}^T Z \Sigma_{s|\tilde{s}} Z^T + \varepsilon_{s|\tilde{s}} \varepsilon_{s|\tilde{s}}^T \right) \quad (10.130)$$

Man erkennt in \hat{F}_t die Form des klassischen LS-Autokorrelationsschätzers und in \hat{V}_t , \hat{Q}_t die Form der klassischen Momentenschätzung für Kovarianzen.

10.5.5 Algorithmus von Shumway und Stoffer

Shumway und Stoffer (1982) gehen in ihrer Arbeit davon aus, daß man zum Zeitpunkt T alle bis zu diesem Moment aufgelaufene Information verwenden sollte und benutzen daher den Kalman-Glätter.

Für diesen kann man wie in Abschnitt 1.2.3 (vgl. auch Anderson und Moore (1979), Kapitel 7) mithilfe der Orthogonalität der Δy_s folgende Rekursionen herleiten

$$\Sigma_{T,T-1|T} = \text{Cov}(\beta_{T|T}, \beta_{T-1|T}|T) = (\mathbb{I} - M_T Z) F \Sigma_{T|T} \quad (10.131)$$

$$\beta_{s|T} = \beta_{s|s} + J_s(\beta_{s+1|T} - \beta_{s+1|s}) \quad (10.132)$$

$$\Sigma_{s|T} = \text{Cov}(\beta_{s|T}|T) = \Sigma_{s|s} + J_s(\Sigma_{s+1|T} - \Sigma_{s+1|s})J_s^T \quad (10.133)$$

$$\begin{aligned} \Sigma_{s,s-1|T} &= \text{Cov}(\beta_{s|T}, \beta_{s-1|T}|T) \\ &= \Sigma_{s|s}J_{s-1}^T + J_s(\Sigma_{s+1,s|T} - F\Sigma_{s|s})J_{s-1}^T \end{aligned} \quad (10.134)$$

Dementsprechend erhalten sie in unserer Notation

$$\tilde{v}_{s|T} = Q^{-1}(\beta_{s|T} - F\beta_{s-1|T}) \quad (10.135)$$

$$\tilde{\varepsilon}_{s|T} = V^{-1}(y_s - Z\beta_{s|T}) \quad (10.136)$$

$$\Lambda_{F|T}(\tilde{v}_{s|T}, \beta_{s-1|T}) = -\tilde{v}_{s|T}\beta_{s-1|T}^T + Q^{-1}(F\Sigma_{s-1|T} - \Sigma_{s,s-1|T}) \quad (10.137)$$

$$\begin{aligned} 2\Lambda_{Q|T}(\tilde{v}_{s|T}) &= \tilde{v}_{s|T}\tilde{v}_{s|T}^T + Q^{-1}(F\Sigma_{s,s-1|T}^T + \Sigma_{s,s-1|T}F^T + \\ &\quad + \Sigma_{s|T} + F\Sigma_{s-1|T}F^T)Q^{-1} - Q^{-1} \end{aligned} \quad (10.138)$$

$$2\Lambda_{V|T}(\tilde{\varepsilon}_{s|T}) = \tilde{\varepsilon}_{s|T}\tilde{\varepsilon}_{s|T}^T + V^{-1}Z\Sigma_{s|T}Z^T V^{-1} - V^{-1} \quad (10.139)$$

und somit als Schätzer

$$\hat{F}_T := \left(\sum_{s=1}^T \beta_{s|T}\beta_{s-1|T}^T + \Sigma_{s,s-1|T} \right) \left(\sum_{s=1}^T \beta_{s-1|T}\beta_{s-1|T}^T + \Sigma_{s-1|T} \right)^{-1} \quad (10.140)$$

$$\hat{Q}_T := \frac{1}{T} \left(\sum_{s=1}^T (\beta_{s|T} - \hat{F}_T\beta_{s-1|T})\beta_{s|T}^T + \Sigma_{s|T} - \hat{F}_T\Sigma_{s,s-1|T} \right) \quad (10.141)$$

$$\hat{V}_T := \frac{1}{T} \left(\sum_{s=1}^T Z\Sigma_{s|T}Z^T + (y_s - Z\beta_{s|T})(y_s - Z\beta_{s|T})^T \right). \quad (10.142)$$

Dieser Schätzer kann aber, wie zu Beginn von Abschnitt 10.3 bemerkt, asymptotisch nicht besser sein als der nur mit den “One-Step-Ahead”-Glättern und Filtern ausgeführte Algorithmus, rechtfertigt also nicht den zusätzlichen Aufwand der Berechnung der über Lag 1 hinaus geglätteten Größen.

10.5.6 Identifizierbarkeit und ein Startschätzer

Wie in Abschnitt 10.4 gezeigt, gilt für $\text{rk } Q = \text{rk } F = \text{rk } Z = p$ und $\text{rk } V = q$ lokale Identifizierbarkeit, in dem Sinn, daß die Fisher-Information des Modells \mathcal{I}_θ vollen Rang hat.

Beim Beweis von Theorem 10.4.6 haben wir dabei den Startschätzer

$$F^0 = (Z^{-}\hat{\gamma}_2^{(Y)}(Z^{-})^\tau)(Z^{-}\hat{\gamma}_1^{(Y)}(Z^{-})^\tau)^{-} \quad (10.143)$$

$$C^0 = (F^0)^{-}Z^{-}\hat{\gamma}_2^{(Y)}(Z^{-})^\tau \quad (10.144)$$

$$Q^0 = C^0 - F^0C^0F^{0\tau} \quad (10.145)$$

$$V^0 = \hat{\gamma}_0^{(Y)} - ZC^0Z^\tau \quad (10.146)$$

eingeführt; diesen wollen wir im folgenden als Startschätzer für ein One-Step-Verfahren verwenden.

Bemerkung 10.5.6 Wir behaupten nicht, daß unser Startschätzer (F^0, Q^0, V^0) bereits effizient ist. Vielmehr ist schon die Schätzung für betragsmäßig kleine F

und Q beliebig schlecht: Es gilt für $q = p = 1$

$$\text{as.Var}(F^0) = \frac{1}{F^2} [(1 - F^2)(2V^2 + \frac{1}{Q^2}) + 2Z^2V]$$

Daß dies für kleines $|F|$ nicht optimal sein kann, ist klar, da wir dann den Quotienten sehr kleiner Autokovarianzen berechnen. Allerdings sollte dieses schlechte Effizienzverhalten durch den One-Step wieder wettgemacht werden.

10.5.7 One-Step-EM-Algorithmus

Die Identifizierbarkeit des Modells sei gegeben. Wir verwenden die auf $\hat{\gamma}$ basierenden Schätzer (87)–(90) als Startschätzer für F , Q und V und erhalten

Algorithmus 10.5.7

(0) Initialisierung:

Berechnung von $\hat{\gamma}_i := \frac{1}{T} \sum_{s=1}^{T-i} y_s^{\text{h}} y_{s+i}^{\text{h},\tau}$ für $i = 0, 1, 2$ mit y_i^{h} den tatsächlichen Beobachtungen; damit Ermittlung von

$$\begin{aligned} F^0 &= (Z^- \hat{\gamma}_2^{(Y)} (Z^-)^\tau) (Z^- \hat{\gamma}_1^{(Y)} (Z^-)^\tau)^{-1} \\ C &= (F^0)^{-1} Z^- \hat{\gamma}_1^{(Y)} (Z^-)^\tau \\ Q^0 &= C - F^0 C F^{0\tau} \\ V^0 &= \hat{\gamma}_0^{(Y)} - Z C^0 Z^\tau \end{aligned}$$

(1) E-Schritt:

Berechnung der L_2 -Ableitungsterme $\Lambda_{\cdot; s|\bar{s}}$ mit $\bar{s} = (s+1) \wedge T$ gemäß (111), (113) und (115) sowie der Fisher-Information \mathcal{I}_θ gemäß Proposition 10.5.5 auf Basis der Hyperparameter Z , F^0 , Q^0 , V^0 .

(2) M-Schritt:

$$\hat{\theta} := \theta^0 + \frac{1}{T} \mathcal{I}_{\theta^0}^{-1} \sum_{s=1}^T \Lambda_{\theta^0, s|\bar{s}}(y_{1:\bar{s}}^0), \quad (10.147)$$

Bemerkung 10.5.8 Mehrere Autoren haben Ansätze zur ML-Schätzung der Hyper-Parameter im “ \mathcal{G}_t ”-Modell dargestellt. Wir nennen hier nur Gupta und Mehra (1974), Ledolter (1979), Pagan (1980). Für eine Übersicht verweisen wir auf Schneider (1988).

Alle diese Autoren arbeiten allerdings unter der Hypothese Gaußscher Prozesse (v_s, ε_t) , während der hier dargestellte Ansatz, wenn auch mit größerem Rechenaufwand bei der Berechnung der bedingten Erwartungswerte, auch im generellen Setup von Abschnitt 10.3 funktioniert.

Schneider (1988) schreibt den ML-Verfahren zwar eine höhere Genauigkeit und eine lokal bessere Konvergenz als dem EM-Algorithmus von Shumway und Stoffer (1982) zu, allerdings sei das globale Konvergenzverhalten des EM-Algorithmus’ stabiler und dieser leichter zu berechnen. In all diesen Kriterien sollte der hier dargestellte One-Step-Algorithmus jeweils die “guten” Eigenschaften auf sich vereinen.

10.5.8 robuster EM–Algorithmus

Wieder sei die Identifizierbarkeit des Modells gegeben. Da wir die Hyper–Parameter als Nebenparameter bei der Prädiktion / beim Filtern / beim Glätten ansehen, genügt hier ein \sqrt{T} –konsistenter Schätzer. Daher brauchen wir uns nicht die Mühe machen, einen robust–optimalen Schätzer zu konstruieren, sondern gehen wie in Bemerkung 9.5.8 skizziert vor. Wir verwenden einen auf einer robusten Schätzung der ACF basierenden Schätzer als Startschätzer:

Algorithmus 10.5.9

(-1) Vorbereitung:

robuste, \sqrt{T} –konsistente Schätzung $\hat{\gamma}^{(Y)}$ der ACF¹³

(0) Initialisierung:

Berechnen von

$$\begin{aligned} F^0 &= (Z^{-\hat{\gamma}_2^{(Y)}}(Z^-)^\tau)(Z^{-\hat{\gamma}_1^{(Y)}}(Z^-)^\tau)^- \\ C &= (F^0)^- Z^{-\hat{\gamma}_1^{(Y)}}(Z^-)^\tau \\ Q^0 &= C - F^0 C F^{0\tau} \\ V^0 &= \hat{\gamma}_0^{(Y)} - Z C Z^\tau \end{aligned}$$

(1) E–Schritt:

Berechnung der L_2 –Ableitungsterme $\Lambda_{\cdot, s|\tilde{s}}$ mit $\tilde{s} = (s+1) \wedge T$ gemäß (111), (113) und (115) sowie der Fisher–Information \mathcal{I}_θ gemäß Proposition 10.5.5 auf Basis der Hyperparameter Z , F^0 , Q^0 , V^0 . Dabei ermittle man die Terme

$$\begin{aligned} \bar{a}_{s,0} &= J_{s-1} M_s \Delta_s^{-\frac{1}{2}}, & \bar{a}_{s,1} &= J_{s-1:s} M_{s+1} \Delta_{s+1}^{-\frac{1}{2}}, \\ \bar{b}_{s,0} &= Z^\tau \Delta_s^{-\frac{1}{2}}, & \bar{b}_{s,1} &= \Sigma_{s|s-1}^{-1} J_s M_{s+1} \Delta_{s+1}^{-\frac{1}{2}}, \\ \bar{c}_{s,0} &= \Delta_s^{-\frac{1}{2}}, & \bar{c}_{s,1} &= V^{-1} Z J_s M_{s+1} \Delta_{s+1}^{-\frac{1}{2}} \end{aligned}$$

und aus diesen $\bar{S}_s = \sum_{k=1}^{s-1} [F^{s-k-1} M_k] \Delta_k[\dots]^\tau$,

$$\begin{aligned} \bar{A}_s &= \bar{a}_{s,0} \bar{a}_{s,0}^\tau + \bar{a}_{s,1} \bar{a}_{s,1}^\tau, & \bar{B}_s &= \bar{b}_{s,0} \bar{b}_{s,0}^\tau + \bar{b}_{s,1} \bar{b}_{s,1}^\tau, \\ \bar{C}_s &= \bar{c}_{s,0} \bar{c}_{s,0}^\tau + \bar{c}_{s,1} \bar{c}_{s,1}^\tau, & \bar{A}\bar{B}_s &= \bar{a}_{s,0} \bar{b}_{s,0}^\tau + \bar{a}_{s,1} \bar{b}_{s,1}^\tau, \\ \bar{B}\bar{C}_s &= \bar{b}_{s,0} \bar{c}_{s,0}^\tau + \bar{b}_{s,1} \bar{c}_{s,1}^\tau, & \bar{A}\bar{C}_s &= \bar{a}_{s,0} \bar{c}_{s,0}^\tau + \bar{a}_{s,1} \bar{c}_{s,1}^\tau \end{aligned}$$

(2) M–Schritt 1:

Per Monte–Carlo berechnen von

$$(a) \ c_s := c_s(y_{1:s-1}) = E_{\theta^0} [R_s \min\{1, \frac{b}{|R_s|}\} | \mathcal{G}_{s-1}]$$

$$(b) \ A := \sum_{s=1}^T E_{\theta^0} [(R_s - c_s) \min\{1, \frac{b}{|R_s|}\} R_s^\tau]$$

mit $R_s = E_{\theta^0} [\Lambda_{\theta^0, s} | Y_{1:\tilde{s}}]$

¹³z.B auf der Basis von Ma und Genton (1998)

(3) M-Schritt 2:

$$\hat{\theta} := \theta^0 + \frac{1}{T} A^{-1} \sum_{s=1}^T (R_s(y_{1:\bar{s}}) - c_s(y_{1:s-1})) \min\left\{1, \frac{b}{|R_s|}\right\} \quad (10.148)$$

mit y_i^{\natural} den tatsächlichen Beobachtungen.

Bemerkung 10.5.10 MC-Simulation Beim Simulieren in (2) kommt uns die Markovsche Struktur von R_s im idealen Modell zugute: R_s hängt außer über deterministische Matrizen von der Vergangenheit $y_{1:s-1}$ nur über $\beta_{s-1|s-1}$ ab, und im idealen Modell gilt¹⁴

$$(\beta_{s-1|s+1}^\tau, v_{s|s+1}^\tau, \varepsilon_{s|s+1}^\tau)^\tau = (\beta_{s-1|s-1}^\tau, 0^\tau, 0^\tau)^\tau + (x_s^\tau, v_{s|s+1}^\tau, \varepsilon_{s|s+1}^\tau)^\tau,$$

wobei beide Summanden stochastisch unabhängig sind und $\beta_{s-1|s-1} \sim \mathcal{N}_p(0, \bar{S}_s)$,

$$X_s := \begin{pmatrix} x_s \\ v_{s|s+1} \\ \varepsilon_{s|s+1} \end{pmatrix} \sim \mathcal{N}_{2p+q}(0, \begin{pmatrix} \bar{A}_s & \bar{A}B_s & \bar{A}C_s \\ \bar{A}B_s^\tau & \bar{B}_s & \bar{B}C_s \\ \bar{A}C_s^\tau & \bar{B}C_s^\tau & \bar{C}_s \end{pmatrix}).$$

Dies nutzen wir aus, indem wir beim Übergang von Zeitpunkt $s-1$ auf s an Werten aus der Vergangenheit bis $s-1$ nur $\beta_{s-1|s-1}^{(j)}$ behalten. Daraufhin sampeln wir unabhängig dazu $X_s^{(i)}$ und berechnen darauf aufbauend $\Lambda_{F,s|\bar{s}}^{(i,j)}$, $\Lambda_{Q,s|\bar{s}}^{(i,j)}$, und $\Lambda_{V,s|\bar{s}}^{(i,j)}$, also $R_s^{(i,j)}$. Durch anschließendes Mitteln von $R_s^{(i,j)}$ $\min\left\{1, \frac{b}{|R_s^{(i,j)}|}\right\}$ erhalten wir Schätzwerte für $c_s(\beta_{s-1|s-1}^{(j)})$; dieses Gitter können wir nun für andere Werte von $\beta_{s-1|s-1}$ zur Interpolation benutzen. In höheren Dimensionen wird dieses Gitter aber viel zu dünn sein; andererseits interessiert für die spätere Schätzung nur $c_s(\beta_{s-1|s-1}^{\natural})$ mit $\beta_{s-1|s-1}^{\natural}$ dem auf den tatsächlichen Beobachtungen ausgewerteten Filter. Daher setzen wir stets $\beta_{s-1|s-1}^{(j=1)} = \beta_{s-1|s-1}^{\natural}$.

Die weiteren simulierten Werte $\beta_{s-1|s-1}^{(j)}$ benötigen wir aber, um bei der Ermittlung von A keine Schätzung eines nach $\beta_{s-1|s-1}^{\natural}$ bedingten Erwartungswertes zu erhalten.

Insgesamt erhalten wir aber hiermit Schätzwerte $\beta_{s|s}^{(i,j)}$, von denen wir im nächsten Schritt s auf $s+1$ nur mehr "ein i " behalten, indem wir setzen $\beta_{s|s}^{(j)} = \beta_{s|s}^{(1,j)}$.

Wahl von b Anders als im üblichen Parameterschätzproblem kommt es hier nicht vordringlich auf Effizienz an; daher geben wir hier auch kein Effizienzkriterium zur Wahl von b an. Dieses wäre analog zur Vorgehensweise beim rIC und mIC aber möglich — und hier auch durch die asymptotische Entwicklung abgesichert. Als Faustregel schlagen wir vor, b als $2\sqrt{\text{tr } \mathcal{I}^{-1}}/p$ zu wählen.

Blockweises Stutzen Wie beim mIC (vgl. Abschnitt 4.4) empfiehlt es sich, die einzelnen Komponenten Λ_F , Λ_V und Λ_Q separat zu stutzen. Anders als

¹⁴der Grenzfall $s = T$ ergibt sich ganz analog.

beim mIC steht für den Fall der Parameterschätzung aber eine Asymptotik als Rechtfertigung der Vorgehensweise zur Verfügung; vgl. Rieder (1994), Remark 5.5.4.

10.5.9 Adaptivität

Aus der Darstellung der L_2 -Ableitung in Δy_t -Termen in Lemma 10.5.2 folgt, daß alle drei Komponenten von Λ jeweils gerade Funktionen in Δy_r sind, in dem Sinn daß $\Lambda((-\Delta y_r)_r) = \Lambda((\Delta y_r)_r)$.

Daher gilt für jede ungerade, $\mathbb{R}^{(m+1)p}$ -wertige Funktion g_s in Δy_r $r \leq s$ — also $g_s((-\Delta y_r)_r) = -g_s((\Delta y_r)_r)$ — für die $E|g_s||\Delta y_{r_1}||\Delta y_{r_2}| < \infty$ für $r_i \leq s$ ist, mit $b \in \mathbb{R}^{(m+1)p}$, $c \in \mathbb{R}^{p^2 + \binom{p}{2} + \binom{q}{2}}$ und r, s beliebig

$$E[b^\tau g_s c^\tau \Lambda_{r|r}] = 0.$$

Da nun aber alle bisher vorgestellten robusten und nicht robusten Verfahren zur Schätzung von β_s (egal ob Glätter, Filter oder Vorhersage) ungerade Funktionen im oben genannten Sinn sind, folgt die schon erwähnte robuste Adaptivität, die es uns erlaubt, zur Schätzung der Hyper-Parameter suboptimale, \sqrt{T} -konsistente Schätzungen vorzunehmen.

Schluß

Zusammenfassung der Ergebnisse

Sind die Anforderungen der Einleitung erfüllt?

Mit dem robusten EM-Algorithmus 10.5.9 haben wir nun alle “Bauteile” für ein Attackieren der Probleme aus der Einleitung beisammen. Mit dem rLS-Filter, dem rIC-Filter und dem mIC-Glätter und -Filter, letzterer zumindest in der block-gestutzten Variante mit fester Fensterweite, haben wir Verfahren entwickelt, die nach einer “off-line” Kalibrierung in der Lage sind, den Anforderungen der Einleitung zu genügen:

Muß man nur mit AO's rechnen, sind die beiden strikt rekursiven Varianten vollkommen ausreichend. Sie zeichnen sich durch einfache, “on-line” berechenbare Korrektur- und Prädiktionsschritte aus, die jeweils durch ihre einfache Gestalt auch für den Praktiker transparente Ergebnisse liefern. Bei Strukturbrüchen sind die (strikt) rekursiven Verfahren nicht geeignet, aber mit dem mIC-Filter steht ein nur wenig aufwendigeres Verfahren zur Verfügung, das auch diesen Ansprüchen genügt.

Für die zur Erstellung von Konfidenzbändern benötigten Fehlerkovarianzen können wir mit Monte-Carlo Integration jeweils im idealen Modell Näherungen “off-line” ermitteln¹⁵.

Sofern wir von der Zeitinvarianz der Hyper-Parameter über einen hinreichend langen Zeitraum ausgehen dürfen, sind diese Verfahren auch bei deren Unkenntnis anwendbar, indem wir diese als Nebenparameter mit Algorithmus 10.5.9 schätzen.

eingeführte Begriffe / Verfahren

In (1) und (2) definieren wir das Zustandsraummodell, das dieser Arbeit zugrundeliegt, und führen mit (V1) – (V4) die verwendeten Verteilungsannahmen ein. Der klassische Kalman-Filter (12) wird als optimal unter allen linearen Schätzern rekapituliert. In den Abschnitten 1.4.1 und 1.4.2 führen wir mit den AO's und IO's die beiden wichtigsten Typen von Ausreißern ein.

Ausgehend von der Optimalität unter allen linearen Schätzern des klassischen Kalman-Filters definieren wir in (4) den rLS-Filter, indem wir den Kor-

¹⁵Dieses ist in der vorliegenden Arbeit nicht ausgeführt, jedoch haben wir jeweils bereits bei der Kalibrierung der Stutzhöhe nach dem Anscombe-Kriterium eine solche Näherung durchführen müssen.

rekturschritt durch

$$\beta_{t|t}^{\text{rLS}} = \beta_{t|t-1}^{\text{rLS}} + M_t^{\text{rLS}} \Delta y_t^{\text{rLS}} \min\left\{1, \frac{b^{\text{rLS}}}{|M_t^{\text{rLS}} \Delta y_t^{\text{rLS}}|}\right\}$$

ersetzen, wobei sich die Matrix M^{rLS} als Lösung des Optimierungsproblems (3) ergibt. Zur Wahl der Clippingschwelle b^{rLS} geben wir ein Versicherungskriterium (5), sowie zur Wahl von K Bedingung (6) an. Wegen der komplizierten Optimierung in M in (3) schlagen wir vereinfachend vor, statt M^{rLS} den Kalman-Gain M^{KK} zu verwenden, und entsprechend die Stutzhöhe b nach (7) zu wählen.

Basierend auf dem Begriff der Influenzkurve im Regressionsmodell aus Definition 3.1.2 erklären wir in (18) den rIC-Filter als One-Step-Schätzer zur Hampel-Krasker Influenzkurve ψ^{rIC} aus (15):

$$\beta_{t|t}^{\text{rIC}} = \beta_{t|t-1}^{\text{rIC}} + \psi^{\text{rIC}} \begin{pmatrix} \beta_{t|t-1}^{\text{KK}} - \beta_{t|t-1}^{\text{rIC}} \\ y_t - Z_t \beta_{t|t-1}^{\text{rIC}} \end{pmatrix}$$

Zur Wahl der Clipping-Höhe geben wir wieder ein an der ARE angelehntes Kriterium in (19) an. In (23) definieren wir eine separat gestutzte, speziell auf AO's abgestimmte Variante des rIC-Filters.

Mit dem mIC-Glätter und -Filter definieren wir in einem entsprechend größeren Regressionsmodell in (8) eine (generalisiert) rekursive One-Step-Robustifizierung, die auch mit IO's zurecht kommt:

$$\beta_{T-m:T|T}^{\text{mIC}} = \beta_{T-m:T|T-1}^{\text{mIC}} + \psi_{\beta_{T-m:T|T-1}^{\text{mIC}}} (y_{T-m:T}, \beta_{T-m|T-m-1}^{\text{KK}})$$

Um diese auch rechentechnisch handhabbar zu halten, fixieren wir dazu die Fensterweite m und verwenden eine blockgestutzte Variante. Die Kalibrierung erfolgt nach (26). Für höhere Dimensionen m und p geben wir in Gestalt des LDM-Algorithmus E.2.1 ein Verfahren an, das die in diesem Fall auftretende spezielle Block-Tridiagonalgestalt der Matrix A^{-1} aus (20) und (21) respektiert.

In den Definitionen 7.2.1 bis 7.2.4 führen wir verschiedene *Stabilitäts*-Begriffe der Lyapunov-Theorie ein und übernehmen aus der Kontrolltheorie im zeitinvarianten Fall in Definitionen 7.2.10 bis 7.2.13 Begriffe wie *vollständige Beobachtbarkeit*, *Konstruierbarkeit* usw., die wir wiederum in Definition 7.2.20 auf den Fall zeitvariabler Hyper-Parameter erweitern. In Definition 7.3.2 definieren wir Markovketten auf beliebigen Zustandsräumen und identifizieren in Proposition 7.3.3 unsere Verfahren als Feller-Markovketten. In Definition 7.3.4 stellen wir die Verallgemeinerungen der wesentlichen Begriffe der Theorie der Markovketten auf diskreten Zustandsräumen für den Fall beliebiger Zustandsräume dar und geben mit diesen in Proposition 7.3.5 die entsprechende Verallgemeinerung des Stationaritätssatzes für Markovketten mit diskretem Zustandsraum wieder. Zusammen mit dem der stochastischen Kontrolltheorie entlehnten Begriff der *schwachen / lokalen stochastischen Kontrollierbarkeit* sind dies die Hilfsmittel zum Nachweis der asymptotischen Stationarität unserer Verfahren.

In Abschnitt 8.1.1 stellen wir den Minimax-Ansatz von Huber (1964) und den "Lemma 5"-Ansatz von Hampel (1968) dar, mit denen wir die Optimalitätsfrage unserer Verfahren klären wollen. Dazu reduzieren wir das Problem, einen Zustand im Zustandsraummodell zu schätzen, zunächst in (8.1) auf das Problem, bei beobachtetem Y , $Y = X + \varepsilon$, eine Schätzung für das zufällige X zu

bestimmen. Neben den bereits eingeführten AO–Ausreißern untersuchen wir diese Problematik auch unter SO–Ausreißern, die wir in Abschnitt 8.1.3 einführen und bei denen nicht $\mathcal{L}(\varepsilon_t)$, sondern direkt $\mathcal{L}(Y_t)$ kontaminiert wird. Die Einzschrittlösung im “Lemma 5”–Ansatz für SO’s veranlaßt uns, in Abschnitt 8.2.2 einen weiteren Filter, den L5–Filter, zu definieren, für dessen Clippingschwelle b^{L5} wir wieder ein Versicherungskriterium (21) vorschlagen. Dieser L5–Filter stellt sich aber in Situationen mit $\rho(F) > 1$ als ungeeignet heraus. Um im Minimax–Ansatz für SO’s ein praktikables Verfahren zu erhalten, erschweren wir in (56) und (57) die Optimierungsaufgabe, indem wir zusätzlich auch spezielle Abweichungen in $\mathcal{L}(\Delta\beta)$ zulassen.

Im zweiten Teil dieser Arbeit geht es zunächst um die asymptotischen, statistischen Eigenschaften des EM–Algorithmus und um dessen Robustifizierung. Um hierzu die Begriffe des lokal–asymptotischen Ansatzes von Le Cam und Hájek zur Verfügung zu haben, definieren wir zunächst in Definition 9.2.1 den Hilbertraum $\mathcal{L}_2^k(\mathcal{A})$ und darauf aufbauend den Begriff der L_2 –Differenzierbarkeit in Definition 9.2.2. Um auch mit ganz allgemeinen Arten von Missings zurechtzukommen, verallgemeinern wir den Begriff der L_2 –Differenzierbarkeit in Definition 9.3.4 auf Dreiecksschemata. Dazu zitieren wir mit Theorem 9.3.5 die Tatsache, daß aus dieser L_2 –Differenzierbarkeit die in Definition A.5.2 und Bemerkung A.5.3(c) eingeführte LAN–Eigenschaft folgt. In Abschnitt 9.4.1 führen wir das m –Stichprobenproblem ein, bei dem ein unabhängiger stochastischer Auswahlmechanismus entscheidet, aus welcher der m Stichproben die Beobachtungen stammen, um in Anlehnung daran in Abschnitt 9.4.2 Missings vom Typ *MCAR* nach Little und Rubin (1987) einführen zu können. Von diesem weisen wir die L_2 –Differenzierbarkeit in Proposition 9.4.3 nach. Mithilfe der damit zur Verfügung stehenden L_2 –Ableitung definieren wir in Definition 9.4.4 den Begriff der *Influenzkurve* und in Definition 9.4.6 den des *as. linearen Schätzers* in diesem Modell mit Missings. In Algorithmus 9.5.1 schreiben wir den EM–Algorithmus von Dempster et al. (1977) in unserer Notation auf. Als Alternative zum iterierten ML–Schätzer schlagen wir in (40) einen One–Step–Schätzer vor und setzen diesen in den *One–Step–EM–Algorithmus* 9.5.5 um. Bei der Robustifizierung des EM–Algorithmus’ können wir analog zum linearen Regressionsmodell auch im *MCAR*–Modell “Fehlerfreie–Variablen”– und “Fehler–in–den–Variablen”–Ansätze verfolgen, die wir kurz in Abschnitt 9.5.5 skizzieren. In (47) geben wir für letzteren die Lösung für das spezielle Optimierungsproblem (45) an.

Um mit der asymptotischen Theorie bei Missings auch Modelle mit Zeitabhängigkeit zu erfassen, verallgemeinern wir in Abschnitt 10.2.1 Le Cams Wurzeldichtenkalkül für Dreiecksschemata von Martingaldifferenzen und übertragen dann den Begriff der L_2 –Differenzierbarkeit in Definition 10.2.5 auf diesen Kontext und decken damit auch vergrößerte Modelle — insbesondere das Zustandsraummodell — ab. Für Modelle mit Zeitabhängigkeit führen wir mit der LAMN–Eigenschaft von Jeganathan (1982) in Definition 10.2.2 eine Verallgemeinerung der LAN–Eigenschaft ein. In Abschnitt 10.2.4 setzen wir dann im Kontext dieser Dreiecksschemata von Martingaldifferenzen in Definition 10.2.11 den Begriff der *Influenzkurve* und in Definition 10.2.12 den des *as. linearen Schätzers* um und können mit diesen optimale Verfahren in diesem Kontext charakterisieren. Die in Abschnitt 10.3.3 definierten Glättungsfiltrierungen erweisen sich als wichtig, wenn es um die Frage geht, ob die Fisher–Information des vergrößerten Modells

vollen Rang hat. In Definition 10.4.1 führen wir hierzu den auf Rothenberg (1971) zurückgehenden Begriff der *lokalen Identifizierbarkeit* ein. Mit den Algorithmen 10.5.7 und 10.5.9 werden ein allein auf Glättern bis maximal Ordnung 2 beruhender, klassisch-optimaler One-Step-EM-Algorithmus und ein robuster One-Step-EM-Algorithmus präsentiert.

Weil wir sie an mehreren Stellen der Arbeit gebraucht haben, haben wir einige Begriffe zur asymptotischen Statistik multivariater Zufallsvariablen in den Anhang verlegt. Dort definieren wir in Abschnitt B.1 die Operatoren $\text{vec}(\cdot)$ und $\text{svec}(\cdot)$, die Kommutatormatrix und das Kroneckerprodukt für Matrizen. In Definition B.2.6 führen wir die Fisher-Information des multivariaten Lokations-/Skalenmodells ohne die Voraussetzung der Lebesgue-Stetigkeit, geschweige denn der Differenzierbarkeit der Log-Likelihoods ein.

Theoretische und empirische Resultate dieser Arbeit

Daß der klassische Kalman-Filter unter (V3) und (V4) sogar optimal unter allen entsprechend meßbaren Verfahren ist, liegt daran, daß unter diesen Verteilungsannahmen der bedingte Erwartungswert linear ist. In Theorem 1.3.1 charakterisieren wir im stochastischen Regressionsmodell $Y = ZX + \varepsilon$ diese Eigenschaft. Unter dieser Linearität, die beim rLS-Filter nur im Schritt $t = 1$ exakt erfüllt ist, weisen wir mit Proposition 2.3.1 nach, daß der rLS-Filter dann robust-optimal ist. In Abschnitt 2.5 rechtfertigen wir zumindest in einigen Beispielen unter (V3) bzw. (V4) die Verwendung von M^{KK} im rLS-Filter anstelle von M^{rLS} , indem wir zeigen, daß man beim Übergang von M^{rLS} zu M^{KK} bestenfalls unwesentlich schlechtere MSE's erreicht.

Für den mIC-Glätter und -Filter weisen wir in Lemma 4.4.2 eine blockweise Cramér-Rao-Schranke nach und sichern so die vorgeschlagene Kalibrierung in (26) ab. Ebenfalls im Kontext der Behandlung von block-indizierten Matrizen sichern wir in Proposition 4.4.6 und Korollar 4.4.7 die Anwendbarkeit des LDM-Algorithmus E.2.1 ohne (Block-)Spalten-Pivotierung ab, der wiederum die block-tridiagonale Struktur des inversen Lagrangemultiplikators A^{mIC}^{-1} ausnützt und so wesentlich weniger Operationen braucht als herkömmliche Matrix-Algebra.

In der Mini-Studie in Kapitel 5 erweisen sich die rekursiven Filter rLS und rIC als gut konkurrenzfähig mit den wesentlich aufwendiger auszuwertenden Posterior-Modus Schätzern, ja sie schneiden unter Kontamination, die nicht der Eichsituation entspricht, in jedem Fall besser ab, als die an normalen Konvex-Kontaminationen geeichten Posterior-Modus Schätzer. Einzig der an der Cauchy-Verteilung geeichte Posterior-Modus kann hier noch mithalten, verschenkt aber im idealen Modell 50% mehr an Effizienz als rLS und rIC.

An statistischen Eigenschaften der bis dahin eingeführten Verfahren rLS, rIC und mIC weisen wir zunächst in den Propositionen 6.1.2, 8.2.1 und 8.4.1 deren Unverzerrtheit, sowie die des L5- und des SO-optimalen Filters (bzw. -Glätters und -Filters) nach. Für Lebesgue-stetige v_t , ε_t zeigen wir in Proposition 6.2.1, daß bei Filtern der rekursiven Bauart (7) die Filterfehler $\Delta\beta_{t|t}$ Lebesgue-stetig sind und decken damit auch den Fall des rLS- und des rIC-Filters ab. In Proposition 6.2.3 spezialisieren wir dieses Resultat für den rLS-Filter bei Dimension $p = 1$ sowie Gaußsches ε_t und v_t und können dieses Resultat dann beim Nachweis der Nichtnormalität von $\mathcal{L}(\Delta\beta^{\text{rLS}})$ in Proposition 6.3.7 verwenden. Die Lebesgue-Stetigkeit der gemeinsamen Verteilung

$\mathcal{L}(\Delta\beta^{\text{rIC}}, \Delta\beta^{\text{rKK}})$ kann in Dimension $p = 1$ in Proposition 6.2.5 für Lebesgue-stetige v_t, ε_t nachgewiesen werden. Mithilfe des asymptotischen Analogons zum Cramér–Lévy–Theorem A.2.3, Proposition A.2.4, können wir auch für beliebige Dimension p — selbst bei (V3) und (V4) — die gemeinsame (asymptotische) Normalität von $(\Delta\beta_t^{\text{rLS}}, \Delta\beta_{t+1}^{\text{rLS}})$ in Proposition 6.3.2 ausschließen. Andererseits gilt unter (V3) bzw. (V4) recht allgemein, daß bei Filtern von der Bauart (7) der Filterfehler bedingt nach den Realisationen $y_{1:t-1}^{\text{f}}$ normalverteilt ist, wie in Proposition 6.3.3 gezeigt wird.

Über die asymptotische Stationarität und Stabilität unserer Verfahren können wir folgendes zeigen: Mit der Definition C.1.1 von *positiv definit* für nicht notwendig symmetrische Matrizen folgt aus dem Sylvesterschen Trägheitssatz C.1.3 in den Lemmata 7.1.1 und 7.1.2 bereits $0 \preceq \mathbb{I}_p - \hat{M}[w]Z \preceq \mathbb{I}_p$ für Matrizen \hat{M} von der Bauart des Kalman–Gains. Für die Aussage $\rho(F(\mathbb{I}_p - \hat{M}[w]Z)) < 1$ benötigen wir dann aber die in Abschnitten 7.2 und 7.3 bereitgestellten Begriffe aus der (stochastischen) Kontrolltheorie und der Theorie der Markovketten auf beliebigem Zustandsraum. Mit diesen gelingt es, die asymptotische Stationarität und Stabilität aller unserer Verfahren rLS, rIC und mIC jeweils für hinreichend große Clippingschwellen $b^{[\cdot]}$ in den Theoremen 7.4.1, 7.5.1 und 7.6.1 nachzuweisen.

Was die Optimalitätsfrage unserer Verfahren im Zustandsraummodell betrifft, so lösen wir das Einschnitt–Optimierungsproblem im reduzierten Modell (1) im “Lemma 5”–Ansatz für SO in Abschnitt 8.2, wobei wir in (17) erhalten

$$\hat{f}^{\text{SO}} = \text{E}_{\text{id}}[X] + (\text{E}_{\text{id}}[X|Y] - \text{E}_{\text{id}}[X]) \min\left\{1, \frac{b/r}{|\text{E}_{\text{id}}[X|Y] - \text{E}_{\text{id}}[X]|}\right\}.$$

In Abschnitt 8.3 verallgemeinern wir ein Theorem von Birmiwal und Shen (1993) auf Dimension $p \geq 1$ und füllen einige “mathematische Lücken” in deren Beweis. Weiterhin können wir uns von deren Annahme lösen, die auftretenden Verteilungen seien Lebesgue–stetig, indem wir in unserem Setup nur eine Dominiertheitsvoraussetzung (25) im idealen Modell verlangen und beliebige Kontaminationen \tilde{P}^Y zulassen. Wie im Lokationsproblem fallen die “Lemma 5”–Lösung \hat{f}^{SO} und die Minimallösung f_0 zu Problem (24) zusammen. Als ungünstigste Verteilung des kontaminierten \hat{Y} erhalten wir in (47)

$$\hat{P}_0^Y(dy) = \frac{1-r}{r} P^Y(dy) \max(\tilde{\rho} |\text{E}_{\text{id}}[X|Y] - \text{E}_{\text{id}}[X]|, 1).$$

Im ursprünglichen Modell weisen wir in Proposition 8.4.1 nach, daß bei Verwendung des jeweils SO–optimalen Filters in jedem Schritt t die Verteilung des so entstehenden Filterfehlers Lebesgue–stetig ist, sofern nur ε_t und v_t es sind, was zugleich die Dominiertheitsvoraussetzung (25) sichert.

Im erweiterten SO–Modell (56) und (57) erhalten wir in Theorem 8.4.2, daß sich durch Übergang zu dieser Erweiterung der Sattelpunkt (f_0, \tilde{P}_0^Y) nicht ändert. Könnten wir nun nachweisen, daß dementsprechend $\mathcal{L}(\Delta\beta_t^{\text{rLS}})$ für jedes t in einer solchen erweiterten SO–Umgebung um eine Normalverteilung liegen würde, so hätten wir hiermit die Optimalität des rLS–Filters in diesem SO–Kontaminationsmodell. Dies analytisch nachzuweisen ist nicht gelungen, allerdings können wir numerisch in Abschnitt 8.4.4 zeigen, daß dies in den in Tabelle 8.1 dargestellten Modellen der Fall ist.

Im AO-Kontaminationsmodell können wir nur mit numerischen Verfahren Lösungen ermitteln. Aus theoretischen Überlegungen folgen Existenz und Eindeutigkeit einer Lösung sowohl im “Lemma 5”-Ansatz als auch im Minimax-Ansatz, vgl. Abschnitte 8.6.3 und 8.7.1–8.7.2. Durch Übergang zur Formulierung mit Dichten und anschließender Diskretisierung — vgl. Abschnitte 8.6.5 und 8.6.7 — können wir das Optimierungsproblem für AO’s im “Lemma 5”-Ansatz in ein endlichdimensionales, quadratisches Optimierungsproblem unter box-constraints, umwandeln. Für dieses geben wir — mit Unterstützung durch C. Zillober — die Lösung im Steady-State Modell in Abbildung 8.8 an.

Im Minimaxansatz für AO’s folgen wir Schick und Mitter (1994) und stellen in Abschnitt 8.7.2 den Zusammenhang zur Bayes-Schätzung und zu den Arbeiten von Donoho (1978), Mallows (1980), Bickel (1981), Levit (1981), Bickel und Collins (1983), sowie Marazzi (1985)¹⁶ her, wobei wir die Argumentation auf den mehrdimensionalen Fall verallgemeinern. Mithilfe des globalen Optimierungsalgorithmus C.2.1 gelingt es uns, Mallows’ Vermutung über die Gestalt der Trägerpunkte des ungünstigsten \tilde{P}^ε zu widerlegen.

Für den rIC-Filter gehen wir den Weg über verallgemeinerte Cramér-Rao-Schranken (105) und (109) und weisen in Proposition 8.8.1 und Korollar 8.8.3 nach, daß diese bestenfalls — vgl. (107) — durch $f_h^0 = E[X + \mathcal{I}_{X|Y}^{-1} \Lambda_x | Y]$ mit Gleichheit erreicht wird, und dies auch nur unter Normalität. In Abschnitt 8.8.2 erhalten wir dann den rIC-Filter als (fast) optimale Approximation an f_h^0 unter der Nebenbedingung, daß das Inkrement des One-Steps geeignet beschränkt ist.

Auf dem Weg zu einer robusten Hyper-Parameterschätzung in Teil II der Arbeit stellen wir zunächst in Theorem 9.3.2 die Übertragung der L_2 -Differenzierbarkeit auf gröbere Modelle fest. In Proposition 9.4.1 zeigen wir die L_2 -Differenzierbarkeit des m -Stichprobenproblems mit stochastischer Auswahl, um diese zusammen mit Theorem 9.3.2 in Proposition 9.4.3 für den Nachweis der L_2 -Differenzierbarkeit im MCAR-Modell zu verwenden. Mit Theorem 9.3.5 ergibt sich daraus die LAN-Eigenschaft dieses Modells. In den Propositionen 9.4.8 und 9.4.9 übertragen wir mit der as. Cramér-Rao-Schranke, dem Faltungssatz und der as. Minimaxschranke die zentralen Optimalitätsresultate des lokal-asymptotischen Ansatzes auf dieses Modell.

In Theorem 10.2.7 zeigen wir in einer Verallgemeinerung von Jeganathan (1982), daß die in Definition 10.2.5 eingeführte L_2 -Differenzierbarkeit für Modelle mit Abhängigkeit die LAMN-Eigenschaft nach sich zieht. Im Gegensatz zu Jeganathan benötigen wir hierzu weder reguläre bedingte Verteilungen noch ein dominiertes Modell. Daß sich L_2 -Differenzierbarkeit wie in Kapitel 9 auf geeignet vergrößerte Modelle überträgt, weisen wir in Theorem 10.2.10 nach. Unter diese Voraussetzungen fällt auch das lineare, zeitinvariante Zustandsraummodell, von dem wir die L_2 -Differenzierbarkeit bei der simultanen Schätzung von F , Q und V und bekanntem Z in Theorem 10.3.2 nachweisen. In Proposition 10.2.15 übertragen wir dann die as. Cramér-Rao-Schranke, den Faltungssatz und die as. Minimaxschranke auch auf dieses Modell.

Wir können das Zustandsraummodell auch als durch Variablen¹⁷ $(Y_s^{(1)}, Y_s^{(2)})_s$ erzeugt auffassen, von denen wir nur $(Y_s^{(1)})_s$ beobachten. Um in diesem Modell nicht nur Subfiltrierungen der natürlichen Filtrierung $\mathcal{F}_t = \sigma((Y_s^{(1)}, Y_s^{(2)})_{s \leq t})$ von der Gestalt $\mathcal{F}_t^{(1)} = \sigma((Y_s^{(1)})_{s \leq t})$ zulassen zu können, liefern wir in Ab-

¹⁶Nicht alle diese Arbeiten wurden hier verwendet, siehe Abschnitt 8.7.2.

¹⁷Im Zustandsraummodell identifizieren wir $Y^{(1)} \rightsquigarrow y$, $Y^{(2)} \rightsquigarrow \beta$.

schnitt 10.3.3 Argumente dafür, daß wir auch Glättungs-Filtrierungen, also $\mathcal{G}_t^m := \sigma((Y_s^{(1)})_{s \leq t+m})$, als Subfiltrierungen der natürlichen Filtrierung interpretieren dürfen.

In Theorem 10.4.3 zeigen wir, daß die lokale Identifizierbarkeit des Submodells längs der Glättungsfiltrierung \mathcal{G}_t^m in unserem Modell mit Zeitabhängigkeit hinreichend für die Regularität der Fisher-Information im durch $\sigma((Y_s^{(1)})_{s \leq t})$ erzeugten Submodell ist. Dies beweisen wir für das Zustandsraummodell in Theorem 10.4.6. Im Beweis hierzu verwenden wir die Charakterisierung der Regularität von $\text{Cov}[\text{svec}(XX^\tau)]$ aus Proposition B.5.1: Diese liegt genau dann vor, wenn X nicht fast sicher auf einer Quadrik liegt. Im Gaußschen Zustandsraummodell zeigen wir in Proposition 10.5.1, daß die Fisher-Information des Modells vollen Rang hat, sofern $\rho(F) < 1$, $\text{rk } Z \geq \text{rk } F = \text{rk } Q = p$, $\text{rk } V = q$. In Proposition 10.5.5 geben wir hierzu die Fisher-Information explizit an.

Die L_2 -Differenzierbarkeit im univariaten Lokationsmodell ist durch eine endliche (Lokations-)Fisher-Information charakterisiert, wie sich aus der Definition dieser Fisher-Information in Huber (1981) und einem Lemma von Hájek (1972) zur Absolutstetigkeit der Wurzel einer Dichte ergibt. Die Verallgemeinerung dieser Charakterisierung auf das multivariate Lokations-/Skalenmodell dürfte auch außerhalb dieser Arbeit von Interesse sein, wo wir sie u.a. beim Beweis von Theorem 10.3.2 verwenden, und findet sich in Anhang B: In Proposition B.2.2 verallgemeinern wir zunächst Hubers Definition der univariaten Lokations-Fisher-Information auf das univariate Skalenmodell. In Theorem B.2.4 geben wir dann die Verallgemeinerung dieses Begriffs auf das multivariate Lokations-/Skalenmodell. Wie im univariaten Fall charakterisiert die Endlichkeit der Fisher-Information — im multivariaten Fall in Gestalt der Größen \bar{I}_f, \bar{J}_f aus (B.6), (B.7) — auch die L_2 -Differenzierbarkeit des multivariaten Lokations-/Skalenmodells, wie wir mit Theorem B.3.4 zeigen.

Abrundung und Ausblick

Auch in dieser umfangreichen Darstellung ist es nicht möglich, alle Fragen zu diesem Thema vollständig zu beantworten. Auf die Darstellung einiger Ergebnisse im Projekt DFG Ri 332/8-1 mußte aufgrund ihres Umfangs an dieser Stelle verzichtet werden, andere Fragen bleiben auch weiterhin offen.

umfassende MSE-Studie

In Kapitel 5 haben wir unsere Verfahren zwar in einer Simulationsstudie getestet, es handelte sich dabei aber nur um den ersten Schritt, $t = 1$, in nur einem Modell, dem Steady-State Modell aus Beispiel 1.4.1, und die mIC-Varianten blieben außen vor.

Eine umfassende MSE-Studie wäre wünschenswert, würde jedoch den Rahmen dieser Arbeit sprengen. Wir haben allerdings im Zuge der Arbeit im Projekt DFG Ri 332/8-1 eine solche Studie durchgeführt, deren Ergebnisse und genauen Versuchsplan man im Internet unter

<http://www.uni-bayreuth.de/departments/math/org/mathe7/kalman> mit der Kennung `kalman` und dem Passwort `filter` einsehen kann.

rekursive Filter zur Erkennung von Strukturbrüchen

Ebenfalls nicht in diese Arbeit aufgenommen ist der von uns versuchte Ansatz, Strukturbrüche zeitverzögert auch mit den rekursiven Filtern zu identifizieren: Durch die Tatsache daß AO's immer nur eine Beobachtung betreffen, während IO's sich auf eine ganze Beobachtungsfolge auswirken, erzeugen beide auch unterschiedliche *Signaturen*¹⁸ im Δy -Prozeß. Diese werden bei den robusten Verfahren des rLS-Filters und des rIC-sep-AO sogar noch deutlicher, denn liegt ein AO vor, so wird $\Delta y^{\text{rIC/rLS}}$ geclippt, und es erfolgt keine extreme Auslenkung; in der nächsten Beobachtung sind wir dann schon auf dem richtigen Ausgangsniveau. Bei IO's hingegen folgt der gedämpfte $\Delta y^{\text{rIC/rLS}}$ einer Niveauänderung durch IO's zu langsam, und so werden mehrere Zeitpunkte folgen, an denen $\Delta y^{\text{rIC/rLS}}$ geclippt werden muß. Wird daher mehrfach hintereinander gestutzt, so deutet dies auf einen IO hin. Zieht man zum Vergleich noch jeweils Δy^{KK} heran, so kann man sehen, ob einem großen $|\Delta y^{\text{rIC/rLS}}|$, ein kleines $|\Delta y^{\text{KK}}|$ gegenübersteht, was dann ebenfalls auf einen IO deutet. Diese Entscheidung kann aber nur aufgrund eines Fensters von m aufeinander folgenden $|\Delta y|$ entschieden werden, weshalb eine entsprechende Verzögerung auftritt. Dieses Verfahren haben wir in einer Studie getestet und sind zu guten Erkennungsraten gekommen. Eine Aufbereitung in Manuskriptform steht allerdings noch aus.

offene Fragen bei der robusten Optimalität

analytische Lösung der rLS-Optimalität

Wie in Abschnitt 8.4.3 dargestellt, ist es bisher nicht gelungen, analytisch nachzuweisen, daß sich die Verteilungen der rLS-Filterfehler für jedes t als Element einer entsprechenden Normalverteilungsumgebung mit Radius r_{SO} auffassen lassen. Ein solcher Nachweis würde den rLS-Filter endgültig als optimal auszeichnen.

genauere Untersuchung der AO-Lösung im "Lemma 5" Problem

Bei der numerischen Lösung des AO-Problems in Beispiel 8.6.8 haben wir die im Nachhinein unbefriedigende Annahme gemacht, daß das optimale \hat{f}^{AO} am Rand für $y \rightarrow \pm\infty$ gegen $\pm b$ geht. Abbildung 8.8 suggeriert aber ein Abklingen des optimalen \hat{f}^{AO} auf 0, also einen Redescender, der nur durch die von uns auferlegte Randbedingung verhindert wird. Hier wäre es interessant zu sehen, ob sich dieser Anschein bestätigt, wenn wir für $y \rightarrow \pm\infty$ verlangen, daß $\hat{f}^{\text{AO}} \rightarrow 0$ geht.

optimale Lösung im AO-Fall für $p \geq 1$ und das ursprüngliche Modell

In den Abschnitten 8.6 und 8.7 haben wir alle Untersuchungen zur Optimalität im AO-Fall im reduzierten Modell durchgeführt und, als es um die konkrete Gestalt ging, diese nur numerisch für Dimension $p = 1$ gewonnen. Hier wäre eine allgemeinere, womöglich sogar analytische Lösung eine echte Erweiterung der bisherigen Resultate.

Optimalität für IO's

In der vorliegenden Arbeit haben wir uns bei den Optimalitätsuntersuchungen auf Ausreißertypen beschränkt, die sich allein auf die Beobachtung, nicht auf den Zustand auswirken, wie AO's und SO's. Eine Untersuchung des mIC-Filters hinsichtlich der Optimalität für IO's steht noch aus.

¹⁸vgl. de Jong und Penzer (1998)

offene Fragen bei der Hyper-Parameterschätzung

einfache robuste, multivariate Autokovarianzen

Ein Punkt, den wir bei der Definition von Algorithmus 10.5.9 noch offen gelassen haben, ist die Frage der robusten Startschätzung der Autokovarianz des Y_t -Prozesses. Wie im Anschluß an die Definition des Algorithmus schon dargestellt, kommt es im Rahmen der Zustandsschätzung bei dieser Startschätzung asymptotisch nur auf die Robustheit und die Konsistenz, nicht aber auf die Effizienz an. Wünschenswert wäre daher ein nicht-iterativer, einfach zu berechnender, robuster Schätzer, vergleichbar etwa dem Median oder dem Interquartilsabstand im univariaten u.i.v.-Fall.

Zur robusten Schätzung der Autokovarianz eines multivariaten Prozesses ist bisher noch sehr wenig publiziert. Neben der schon genannten Arbeit von Ma und Genton (1998), seien hier noch Chan und Wei (1992) und Ferguson et al. (2000) erwähnt.

Studie unserer EM-Algorithmen im normalen Zustandsraummodell

Weil wir uns bisher noch nicht für ein einfaches Verfahren zur robusten Startschätzung der Autokovarianz des Y_t -Prozesses haben entscheiden können, steht dementsprechend eine Überprüfung von Algorithmus 10.5.9 an realen und simulierten Daten noch aus. Da wir dies dann gleich mit dem Testen von Algorithmus 10.5.7 verbinden wollen, stehen auch hier noch keine empirischen Resultate zur Verfügung.

zeitkontinuierlicher Fall

Als eine Herausforderung steht immer noch die Verallgemeinerung der hier dargestellten Ansätze, insbesondere auch des Begriffs der Influenzkurve auf den zeitkontinuierlichen Fall aus. Dabei sollten die von Jones (1986) betrachteten Erweiterungen des AR-, ARMA- und ARIMA-Modells auf kontinuierliche Zeit CAR, CARMA und CARIMA sowie deren Einbettung in zeitkontinuierliche Zustandsraummodelle hilfreich sein. Von einem solchen Ansatz dürfte man sich vielleicht auch neue Impulse, in jedem Fall aber tiefere Erkenntnis im Bereich des *Financial Engineering* und des in der Einleitung erwähnten *Risiko-Controlling* erhoffen.

erweiterte Verfügbarkeit

Schließlich wäre es wünschenswert, neben den rekursiven Verfahren auch die mIC-Varianten, sowie die Routinen zur Ermittlung einer robusten Hyper-Parameter-Schätzung allgemein zugänglich zu machen, und wir werden versuchen, im Laufe der nächsten Zeit entsprechende Macros in geeigneter Form ins Internet zu stellen.

Conclusion

Summary and results

What about the requirements from the introduction?

With the robust EM–Algorithm 10.5.9 at hand, we dispose of all components necessary to attack the problems from the introduction. The procedures developed in this thesis, the rLS filter, the rIC filter and the mIC smoother and filter — the last one at least as long as we use its blockwise–clipped variant with fixed window size — all comply with the requirements of the introduction, once they are calibrated “off–line”:

For just protecting against AO’s, the strict recursive procedures completely suffice. They are distinguished by easy, “on–line”–computable correction and prediction steps, both of which are transparent in their effects to the practitioner due to their simple form.

For structural breaks / IO’s the (strictly) recursive procedures are inadequate, but with the mIC smoother and filter, we propose a just slightly more complicated procedure that meets these requirements.

For the error covariances needed to construct confidence bands we may compute numerical solutions by Monte–Carlo integration “off–line” in the ideal model¹⁹. As long as we may assume time invariance of the hyper parameters for a sufficiently long period, these procedures are also available when the hyper parameters are not known, as these parameters may be estimated as nuisance parameters with algorithm 10.5.9.

Notions / procedures introduced

The state space model we are dealing with in this thesis is defined in (1) and (2). In (V1) – (V4) we introduce the distributional assumptions used. The classical Kalman filter (12) is recapitulated as optimal among all linear estimators. In sections 1.4.1 and 1.4.2 the most important types of outliers, AO’s and IO’s, are introduced.

Based on the optimality among all linear estimator of the classical Kalman filter, the rLS filter is defined in (4), replacing the correction step by

$$\beta_{t|t}^{\text{rLS}} = \beta_{t|t-1}^{\text{rLS}} + M_t^{\text{rLS}} \Delta y_t^{\text{rLS}} \min\left\{1, \frac{b^{\text{rLS}}}{|M_t^{\text{rLS}} \Delta y_t^{\text{rLS}}|}\right\},$$

¹⁹This is not worked out in this thesis, but already for the calibration of the clipping height of our procedures according to the Anscombe–criterion we had to compute terms of that type this way.

where matrix M^{rLS} arises as solution to the optimization problem (3). As to the choice of the clipping height b^{rLS} and the loss-function modifier K , we propose the insurance criterium (5) for b^{rLS} , and (6) for K . As optimization in M is complicated, we suggest instead using the classical Kalman gain M^{KK} , and respectively choosing b according to (7).

In (18), based on Definition 3.1.2 of a regression influence curve, we define the rIC filter as a one-step estimator to the Hampel–Krasner influence curve ψ^{rIC} from (15):

$$\beta_{t|t}^{\text{rIC}} = \beta_{t|t-1}^{\text{rIC}} + \psi^{\text{rIC}} \begin{pmatrix} \beta_{t|t-1}^{\text{KK}} - \beta_{t|t-1}^{\text{rIC}} \\ y_t - Z_t \beta_{t|t-1}^{\text{rIC}} \end{pmatrix}$$

For the clipping threshold we again propose a criterium leaning on ARE in (19). For the pure AO-case, we define a separately-clipped variant of the rIC filter in (23) particularly adapted to this situation.

The mIC smoother and filter is defined in (8) as a one-step robustification for a somewhat larger regression model as

$$\beta_{T-m:T|T}^{\text{mIC}} = \beta_{T-m:T|T-1}^{\text{mIC}} + \psi_{\beta_{T-m:T|T-1}^{\text{mIC}}} (y_{T-m:T}, \beta_{T-m|T-m-1}^{\text{KK}}).$$

It is recursive in a generalized sense and is capable of dealing with IO's. In order to keep this procedure computationally manageable, we fix its window size m and use a blockwise clipping variant. Calibration, i.e. determining b^{mIC} , is done according to (26). For higher dimensions m and p , we provide the LDM-algorithm E.2.1 that respects the special block-tridiagonal form of matrix A^{-1} from (20) and (21), as arising in the variant with blockwise clipping in a natural way.

In definitions 7.2.1 to 7.2.4, we introduce different notions of *stability* from Lyapunov-theory, followed by the notions of *complete observability*, *constructability* etc. from control theory for the time-invariant case in definitions 7.2.10 to 7.2.13. These are generalized to the time varying case in definition 7.2.20. Markov chains on arbitrary state spaces are established in definition 7.3.2, identifying our procedures as Feller Markov chains in proposition 7.3.3. In definition 7.3.4 we present the generalizations of the main notions of the theory of Markov chains on a discrete state space to the case of arbitrary spaces; the corresponding generalization to arbitrary spaces of the stationarity theorem for Markov chains on a discrete space is cited in proposition 7.3.5. Together with the notion of *weak / local stochastic controllability* from stochastic control theory, these are the tools used for the proof of asymptotic stationarity of our procedures.

In section 8.1.1, we present the minimax approach of Huber (1964) and the “Lemma 5” approach of Hampel (1968), with both of which we work to settle the question of optimality of our procedures. For this we reduce the problem of estimating a state in the state space model to the problem of estimating X based on the observation Y in the model $Y = X + \varepsilon$ in (1). Additionally to the already introduced AO-outliers, we also study our problem under SO-outliers. In section 8.1.3 these are defined as outliers, where instead of $\mathcal{L}(\varepsilon_t)$, $\mathcal{L}(Y_t)$ is contaminated directly. The one-step solution in the “Lemma 5” approach leads us to defining a further filter in section 8.2.2, the L5 filter. For the corresponding clipping threshold b^{L5} we again propose an insurance criterium (21). This L5 filter, however, proves to be inadequate in situations with $\rho(F) > 1$.

In order to find a practicable procedure in the minimax approach for SO's

for $t > 1$, we consider an even harder optimizing problem in (56) and (57), additionally allowing for special deviances in $\mathcal{L}(\Delta\beta)$.

In the second part of this thesis, we first consider the statistical properties of the EM-Algorithm and its robustification. In order to dispose of the notions of the local asymptotic approach associated with Le Cam and Hájek, in Definition 9.2.1 we define the Hilbert space $\mathcal{L}_2^k(\mathcal{A})$ and based on this the notion of L_2 -differentiability in definition 9.2.2. In order to be able to cope with arbitrary types of missings, we generalize L_2 -differentiability to triangular arrays in definition 9.3.4. In theorem 9.3.5 we state the fact, that this L_2 -differentiability implies the LAN-property from definition A.5.2 and remark A.5.3(c). In section 9.4.1, we introduce the m sample problem, where at each observation instance, an independent stochastic selection mechanism decides from which of the m samples comes the observation. According to this, in section 9.4.2 we introduce the model with missings of type *MCAR* due to Little and Rubin (1987), which is shown to be L_2 -differentiable in proposition 9.4.3. Using the L_2 -derivative thereby available for this model, we define the notions of *influence curve* and *as. linear estimator* in this model in definitions 9.4.4 and 9.4.6.

In algorithm 9.5.1 we write down the EM algorithm of Dempster et al. (1977) using our notation. As an alternative to the completely iterated MLE, we propose using a one step estimator in (40) and on basis of this define the *One-Step-EM-algorithm* 9.5.5. Analogously to the linear regression model we may pursue “error-free-variables”- and “errors-in-variables” approaches to robustification in the *MCAR*-model, both of which we shortly sketch in section 9.5.5. For the latter we present the solution to the special robust optimization problem (45) in (47).

In chapter 10, the notion of L_2 -differentiability is translated to models with time dependencies including coarsened models as in chapter 9; in particular the state space model is covered. To this we generalize Le Cam’s square root calculus of densities to triangular arrays of martingale differences in section 10.2.1 followed by a corresponding notion of L_2 -differentiability in definition 10.2.5. For models with time dependencies we introduce the LAMN-property of Jegannathan (1982) in definition 10.2.2 as a generalization of the LAN-property. In section 10.2.4 the notions of *influence curve* and *as. linear estimator* are transferred to this context in definitions 10.2.11 and 10.2.12. Smoothing filtrations as introduced in section 10.3.3 prove useful as to the question whether the Fisher information of a coarsened model is of full rank. To this we introduce the notion of *local identifiability* due to Rothenberg (1971) in definition 10.4.1.

With the algorithms 10.5.7 and 10.5.9, we present a classically optimal One-Step-EM-algorithm based only on smoothers up to order 2 as well as a robust One-Step-EM-algorithm.

The notions from asymptotic statistics of multivariate random variables are used at several places of this thesis and are therefore delegated to the appendix. In section B.1 we define the operators $\text{vec}(\cdot)$ and $\text{svec}(\cdot)$, the commutator matrix and the Kronecker product for matrices. In definition B.2.6 we introduce the Fisher information of the multivariate Location-/Scale model without assuming a Lebesgue-density or even differentiability of the Log-Likelihoods.

Theoretical and empirical results of this thesis

That the classical Kalman filter is even optimal among all measurable procedures under (V3) and (V4) is due to the fact that under these distributional assumptions the conditional expectation is linear. In theorem 1.3.1 we characterize this property in the stochastic regression model $Y = ZX + \varepsilon$. Under this linearity, which under (V3) and (V4) only holds for $t = 1$ for the rLS filter, we prove that the rLS filter is robust-optimal in proposition 2.3.1. In section 2.5 we justify the replacement of M^{rLS} by M^{KK} for at least (V3) and (V4), showing in several examples that using M^{KK} instead of M^{rLS} , we incur an MSE being at most only inconsiderably worse.

For the mIC smoother and filter we show a blockwise Cramér–Rao–bound in Lemma 4.4.2 and thereby guarantee the possibility of our proposed calibration (26). By proposition 4.4.6 and corollary 4.4.7, we ensure that the LDM–algorithm E.2.1 may be applied without (block–) column–pivoting, thus preserving the (block–)tridiagonal structure of the inverse Lagrange–multiplier A^{mIC}^{-1} generated by blockwise clipping.

In the mini–study of chapter 5, the recursive filters rLS and rIC compete well with the posterior–mode estimators, the evaluation of the latter being much more expensive, however. Looked upon in situations different to the proper gauging distribution, our procedures are performing even better than the posterior mode estimators gauged at normal convex combinations. The only competitor in this study is the posterior mode estimator gauged at the Cauchy distribution, which however loses more than 50% of efficiency in the ideal model w.r.t. rLS and rIC.

Starting with the statistical properties of our procedures, in propositions 6.1.2, 8.2.1 and 8.4.1, we show unbiasedness of the rLS–, the rIC–, the mIC–, the L5– and the SO–optimal filter (resp. smoother and filter). For v_t, ε_t possessing Lebesgue densities, in proposition 6.2.1 we prove that filters of recursive type (7) generate filtering errors $\Delta\beta_{t|t}$ possessing Lebesgue densities. This also covers the case of the rLS and rIC filter. In proposition 6.2.3 we specialize this result to the rLS filter in dimension $p = 1$ under Gaussian ε_t and v_t , and get an explicit form for the density which we use to prove non–normality of $\mathcal{L}(\Delta\beta^{\text{rLS}})$ in proposition 6.3.7. Lebesgue continuity of the joint distribution $\mathcal{L}(\Delta\beta^{\text{rIC}}, \Delta\beta^{\text{KK}})$ is shown for dimension $p = 1$ in proposition 6.2.5 for Lebesgue–continuous v_t, ε_t . By means of the asymptotic analogue of the Cramér–Lévy–theorem A.2.3 — proposition A.2.4 — we are able to exclude joint (asymptotic) normality of $(\Delta\beta_t^{\text{rLS}}, \Delta\beta_{t+1}^{\text{rLS}})$ for arbitrary dimension p and even under (V3) and (V4) in proposition 6.3.2. On the other hand, under (V3) resp. (V4), we have quite generally that for filters of type (7) the filtering error conditioned on the realizations $y_{1:t-1}^{\text{h}}$ is normally distributed, as shown in proposition 6.3.3.

Concerning asymptotic stationarity and stability, we use definition C.1.1 of *positive definit* for not necessarily symmetric matrices, by Sylvester’s law of inertia C.1.3 in order to deduce that $0 \preceq \mathbb{I}_p - \hat{M}[w]Z \preceq \mathbb{I}_p$ for matrices \hat{M} of type of the Kalman gain in Lemmas 7.1.1 and 7.1.2. For the stronger assertion $\rho(F(\mathbb{I}_p - \hat{M}[w]Z)) < 1$, which entails asymptotic stability, we need the notions of (stochastic) control theory and the theory of Markov chains on an arbitrary state space, however. These notions are provided in sections 7.2 and 7.3, and with these we are able to prove asymptotic stationarity and stability of all our procedures rLS, rIC and mIC in theorems 7.4.1, 7.5.1 and 7.6.1 for sufficiently

large clipping heights $b^{[.]}$, respectively.

As to the question of optimality of our procedures in the state space model, in the “Lemma 5” approach for SO’s, we solve the one step optimization problem in the reduced model (1) in section 8.2, yielding in (17)

$$\hat{f}^{\text{SO}} = E_{\text{id}}[X] + (E_{\text{id}}[X|Y] - E_{\text{id}}[X]) \min\left\{1, \frac{b/r}{|E_{\text{id}}[X|Y] - E_{\text{id}}[X]|}\right\}.$$

In section 8.3 we generalize a theorem due to Birimiwal and Shen (1993) to dimension $p \geq 1$, closing some “mathematical gaps” in their proof. We may furthermore dispense with these authors’ assumption that all distributions involved — including the contaminating ones — possess Lebesgue densities. Instead we assume a certain domination condition (25) in the ideal model only, and allow for arbitrary contaminations \hat{P}^Y . As in the location problem, the solutions to the “Lemma 5” approach, \hat{f}^{SO} , and to the minimax approach, f_0 , coincide. As least favorable distribution for the contaminated \hat{Y} we get in (47)

$$\hat{P}_0^Y(dy) = \frac{1-r}{r} P^Y(dy) \max(\tilde{\rho} |E_{\text{id}}[X|Y] - E_{\text{id}}[X]|, 1)$$

In proposition 8.4.1 we show that in the original model, using the corresponding SO-optimal filter in each step t , the filter arising this way is unbiased. This proposition also shows that the distribution of the filtering error is Lebesgue continuous, as long as ε_t and v_t are, which at the same time entails the domination condition (25).

In theorem 8.4.2 we see that, passing to the enlarged SO model (56) and (57), the saddle point (f_0, \hat{P}_0^Y) remains unchanged. If we were able to show at that point that, for each t , $\mathcal{L}(\Delta\beta_t^{\text{rLS}})$ was an element of such an enlarged SO neighborhood around some central normal distribution, this would proof the optimality of the rLS filters in this SO contamination model. In this thesis we did not succeed in showing this analytically, numerical results in in section 8.4.4 however show that this is indeed true for all models displayed in table 8.1.

In the AO contamination model problems are much harder, and only by means of numerical procedures solutions may be obtained. For theoretical reasons existence and uniqueness of a solution follow both in the “Lemma 5” approach and in the minimax approach, c.f. sections 8.6.3 and 8.7.1–8.7.2. Passing to a formulation with densities and subsequently discretizing the problem — c.f. sections 8.6.5 and 8.6.7 — we may approximate the optimization problem for AO’s in the “Lemma 5” approach by a deterministic, quadratic optimization problem under box-constraints. With the assistance of C. Zillober, we get a numerical solution to this problem for the steady-state model as displayed in figure 8.8.

Following Schick and Mitter (1994), in the minimax approach for AO’s, in section 8.7.2 we establish the connection to Bayes estimation and to Donoho (1978), Mallows (1980), Bickel (1981), Levit (1981), Bickel and Collins (1983) and Marazzi (1985)²⁰, and generalize their arguments to the multivariate case. By means of the global optimization algorithm C.2.1 we are able to disprove Mallows’ conjecture on the form of the support of the least favorable AO-contaminating distribution \hat{P}^ε .

For the rIC filter we reason by generalized Cramér–Rao–bounds (105) and (109): In proposition 8.8.1 and corollary 8.8.3 we show that this bound may

²⁰Not all of these papers were used in this thesis, c.f. section 8.7.2.

only be attained by $f_h^0 = \mathbb{E}[X + \mathcal{I}_{X|Y}^{-1} \Lambda_x | Y]$ — c.f. (107) — and this only in the Gaussian case. In section 8.8.2 we obtain the rIC filter as (near) optimal approximation to f_h^0 under the constraint that the increment of the one-steps be suitably bounded.

On our way to robust estimation of the hyper parameters of the time-invariant state space model in part II of this thesis, in theorem 9.3.2, we state the result that L_2 -differentiability of a finer model is inherited to a submodel coarsened to a less informative sub- σ -Algebra. The L_2 -differentiability of the m -sample problem proven in proposition 9.4.1 is used together with theorem 9.3.2 in proposition 9.4.3 for the proof of L_2 -differentiability of the MCAR model. With theorem 9.3.5 this implies the LAN-property of this model. In propositions 9.4.8 and 9.4.9 we translate the central optimality results of the local-asymptotic approach in form of the as. Cramér-Rao bound, the convolution theorem and the as. minimax bound to this model.

Generalizing Jeganathan (1982) in some respects, in theorem 10.2.7 we show, that L_2 -differentiability for models with time dependencies as introduced in Definition 10.2.5 implies the LAMN-property. Contrary to Jeganathan we assume neither regular conditional distributions nor a dominated model. As in chapter 9, L_2 -differentiability is inherited to suitably coarsened models, which is shown in theorem 10.2.10. These results are used in theorem 10.3.2 to show L_2 -differentiability for simultaneous estimation of F , Q and V in the linear, time invariant state space model with known Z . In proposition 10.2.15 we translate the as. Cramér-Rao bound, the convolution theorem and the as. minimax bound to this model.

The state space model may be regarded as generated by partitioned variables²¹ of type $(Y_s^{(1)}, Y_s^{(2)})_s$, where we only observe $(Y_s^{(1)})_s$. In order to be able to cover more general subfiltrations of the natural filtration $\mathcal{F}_t = \sigma((Y_s^{(1)}, Y_s^{(2)})_{s \leq t})$ than those of form $\mathcal{F}_t^{(1)} = \sigma((Y_s^{(1)})_{s \leq t})$, in section 10.3.3 we provide arguments showing that also smoothing filtrations, i.e. $\mathcal{G}_t^m := \sigma((Y_s^{(1)})_{s \leq t+m})$, may be considered as subfiltrations of the natural filtration.

In theorem 10.4.3 we show that local identifiability of the submodel endowed with the smoothing filtration \mathcal{G}_t^m is sufficient for regularity of the Fisher information in the submodel generated by $\sigma((Y_s^{(1)})_{s \leq t})$. This property is proven for the state space model in theorem 10.4.6. In the proof of this assertion we use the characterization of regularity of $\text{Cov}[\text{svec}(XX^\tau)]$ from proposition B.5.1: This covariance is regular iff X is not lying on a conic with probability one. In proposition 10.5.1 we show that in the Gaussian state space model, the Fisher information of the model is not degenerated if $\rho(F) < 1$, $\text{rk } Z \geq \text{rk } F = \text{rk } Q = p$, $\text{rk } V = q$. For this we spell out the Fisher information explicitly in proposition 10.5.5.

Using a Lemma due to Hájek (1972) concerning absolute continuity of the square roots of a density, L_2 -differentiability in the univariate location model can be characterized by means of a finite Fisher information of location in the sense of Huber (1981). The generalization of this result to the multivariate location-/scale model in appendix B might be of interest beyond the scope of this thesis, where we use it e.g. in the proof of theorem 10.3.2: In proposition B.2.2 we first generalize the definition of Fisher information of univariate

²¹In the state space model, we identify $Y^{(1)} \rightsquigarrow y$, $Y^{(2)} \rightsquigarrow \beta$.

location of Huber (1981) to the univariate scale model. In Theorem B.2.4 we then translate the notion to the multivariate location–/scale model. As in the univariate case, finiteness of Fisher information — in the multivariate case in form of finiteness of the scalars $\bar{\mathcal{I}}_f$, $\bar{\mathcal{J}}_f$ from (B.6), (B.7) — also characterizes L_2 –differentiability of the multivariate locations–/scale model, as we demonstrate in theorem B.3.4.

Complements and perspectives

Even in this extensive thesis it is not possible to completely cover all questions arising in this topic. Due to their length we refrain from presenting some of the results of the project DFG Ri 332/8-1, other questions have not yet been settled up to now.

Extensive study of the MSE

In chapter 5 we have tested our procedures in a (mini–) simulation study; however, this has only been done for $t = 1$ in the steady state model from example 1.4.1. Furthermore only the recursive rLS and rIC have been included — the variants of the mIC have been left to a better suited framework, as they cannot be defined in a meaningful way until $t \geq m > 1$. An extensive MSE study of all our procedures in more than just one model and in which they are examined at a later time t than $t = 1$, would thus be desirable to convince the reader of the performance of our procedures in practice. This however would go beyond the scope of this thesis by far. On the other hand we do have realized such a study in the framework of project DFG Ri 332/8-1, the results of which together with the design specifications of this study may be looked up in the internet under

<http://www.uni-bayreuth.de/departments/math/org/mathe7/kalman>
using access name `kalman` and password `filter`.

Recursive filters for the detection of structural breaks

Neither is included to this thesis an approach we have tried out, in which structural breaks are to be identified with a certain delay by means of the recursive filters only:

Due to the fact that AO's always affect but one observation, while IO's continue influencing the observations for a longer period, these types of outliers generate quite different *signatures*²² in the Δy –process. These differences are even clearer for the robust procedures of rLS and rIC-sep-AO, because if a large AO occurs, $\Delta y^{rIC/rLS}$ is clipped, and no extreme peaks are entailed in $\Delta y^{rIC/rLS}$; in the following observation we then start from the correct level. With IO's, however, the attenuated $\Delta y^{rIC/rLS}$ follows a level change caused by an IO too slowly, and so for several time instances in a row, $\Delta y^{rIC/rLS}$ will have to be clipped. Runs of clippings thus indicate an IO. If for comparison we also consider Δy^{KK} , we can monitor whether on occasion of a large $|\Delta y^{rIC/rLS}|$, at the same time the classical Kalman filter produces a small $|\Delta y^{KK}|$, which also indicates an IO. This decision of course only can be made in a reasonable way on basis of a window of m succeeding $|\Delta y_t|$, which is why we have to incur a corresponding

²²c.f. de Jong and Penzer (1998)

delay.

This procedure has been tested and has yielded promising results. An elaboration to manuscript form remains yet to be done, however.

Open questions regarding robust optimality

Analytical solution of rLS-optimality

As presented in section 8.4.3, up to now we have not been able to prove analytically that the distributions of the rLS filtering error may be interpreted as an element of a corresponding neighborhood of radius r_{SO} around a certain normal distribution for each t . Such a proof would conclusively show the rLS filter to be optimal.

Closer analysis of the AO solution in the “Lemma 5” Problem

Solving the AO problem numerically in example 8.6.8, we have made the assumption, that the optimal \hat{f}^{AO} tends to $\pm b$ for $y \rightarrow \pm\infty$, which in the end has turned out to be unsatisfactory: Figure 8.8 suggests a decay of the optimal \hat{f}^{AO} to 0, i.e. a redescender, that is only inhibited by the constraint $\hat{f}^{\text{AO}} \rightarrow \pm b$ imposed by us. It would be interesting to see whether this redescending property is confirmed, if we require $\hat{f}^{\text{AO}} \rightarrow 0$ for $y \rightarrow \pm\infty$.

Solution for the AO-case for $p \geq 1$ and in the original model

In sections 8.6 and 8.7 we have limited ourselves to studying optimality for the AO-case in the reduced model and, when determining the concrete form, we have only been able to solve the problem for dimension $p = 1$. So the question of an analytic expression — perhaps even for $p \geq 1$ — is still to be settled.

Optimality for IO's

In the present thesis we have confined ourselves to considering only optimality for types of outliers affecting solely the observations, like AO's and SO's. An analysis of the mIC-filter w.r.t. IO-optimality is yet to be done.

Open questions as to the estimation of hyper parameters

Simple, multivariate covariances

One point still left open in the definition of algorithm 10.5.9, is the question of a robust starting estimator for the autocovariance of Y_t process. As is explained already in the sequel of the definition of the algorithm, in the framework of state estimation for this starting estimator asymptotically only robustness matters, while efficiency is of secondary interest. Therefore a non-iterative robust estimator easy to compute would be desirable, perhaps comparable to the median or the interquartile range in the univariate i.i.d. case.

As to robust estimation of the autocovariance of a multivariate process only little is published hitherto. Besides the paper of Ma and Genton (1998) already cited in chapter 10, we mention Chan and Wei (1992) and Ferguson et al. (2000).

Study of our EM-algorithms in the Gaussian state space model

Having not yet decided upon a simple procedure for a robust starting estimator of the autocovariance of the Y_t process, a test of the algorithm 10.5.9 with real and simulated data is yet to be done. As we want to do this test jointly with the test of algorithm 10.5.7, neither are numerical results available to the latter.

Continuous time

The generalization of the presented approach, in particular the notion of influence curve, to the setting of continuous time remains a challenging problem. To this the generalizations of AR-, ARMA- and ARIMA-Modells to continuous time as CAR, CARMA and CARIMA and their representation as time-continuous state space models, considered by Jones (1986), ought to be useful. From such an approach, we may perhaps hope for new impulses and a deeper insight into the domain of *financial engineering* and *risk control* in banking as mentioned in the introduction of this thesis.

Availability

In the XploRe quantlib `kalman.lib` the rLS filter is implemented in form of the quantlet `rLSfil`, which realizes the $rLS(M^{KK})$ filter to a given b^{rLS} , and `calibrLS`, which calibrates the rLS filter to a given δ ; also confer Ruckdeschel (2000b) for the corresponding input- and output-parameter. Like the rLS filter, the rIC filter is available within the programm package XploRe. The rIC filter is implemented in form of the quantlets `ricfil` and `calibrIC`, analogously defined as in the rLS case.

ISP-routines realizing the mIC smoothers und filters, as well as for calibrating them may be downloaded under

[http://www.uni-bayreuth.de/departments/math/org/mathe7/
RUCKDESCHEL/diss/ispmacros/mIC](http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/diss/ispmacros/mIC)

ISP-macros for the L5 filter and its calibration are available under

[http://www.uni-bayreuth.de/departments/math/org/mathe7/
RUCKDESCHEL/diss/ispmacros/L5](http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/diss/ispmacros/L5)

Finally, besides the recursive procedures and the mIC variants, we will try and put routines to the robust estimation of the hyper parameters by means of our EM-algorithms into the internet in a suitable form in the next time.

Anhang A

Wahrscheinlichkeitstheorie und asymptotische Statistik

A.1 schwache Konvergenz von Maßen

Wir stellen einige wichtige Begriffe und Tatsachen über die schwache Konvergenz von Maßen zusammen¹, die wir in der vorliegenden Arbeit ohne Beleg verwenden:

Definitionen

Sei (Ξ, d) ein metrischer Raum, \mathcal{B} die Borel- σ -Algebra und $\mathcal{C}_{[c]}(\Xi)$ die Menge aller stetigen, beschränkten Funktionen $f : \Xi \rightarrow \mathbb{R}$ [mit kompaktem Träger]. Eine Folge endlicher Maße $Q_n \in \mathcal{M}_b(\mathcal{B})$ heißt *straff*, falls es zu $\varepsilon > 0$ ein Kompaktum $K = K(\varepsilon) \subset \Xi$ gibt, so daß

$$\limsup_n Q_n(\Xi \setminus K) < \varepsilon. \quad (\text{A.1})$$

Diesen Begriff verwenden wir auch für Folgen Ξ -wertiger Zufallsgrößen auf Definitionsbereichen $(\Omega_n, \mathcal{A}_n, P_n)$ und meinen damit die Straffheit der Bildmaße $X_n \circ P_n$.

Eine Folge von Maßen $Q_n \in \mathcal{M}_b(\mathcal{B})$ heißt *schwach konvergent* gegen ein $Q_0 \in \mathcal{M}_b(\mathcal{B})$, falls für alle $f \in \mathcal{C}(\Xi)$ gilt

$$\lim_n \int f dQ_n = \int f dQ_0. \quad (\text{A.2})$$

$Q_n \in \mathcal{M}_b(\mathcal{B})$ heißt *vage konvergent* gegen ein $Q_0 \in \mathcal{M}_b(\mathcal{B})$, falls für alle $f \in \mathcal{C}_c(\Xi)$ gilt

$$\lim_n \int f dQ_n = \int f dQ_0. \quad (\text{A.3})$$

Wir schreiben für die beiden Konvergenzen

$$Q_n \xrightarrow{w} Q_0, \quad \text{resp.} \quad Q_n \xrightarrow{v} Q_0$$

Bezeichnen wir die Menge der Unstetigkeitsstellen einer meßbaren Funktion h mit D_h , so gilt

¹Die Zusammenstellung hält sich an Rieder (1994), App. A.

Proposition A.1.1 (Stetigkeitssatz) Seien (Ξ, \mathcal{B}) , $(\tilde{\Xi}, \tilde{\mathcal{B}})$ zwei metrische Räume und $h : (\Xi, \mathcal{B}) \rightarrow (\tilde{\Xi}, \tilde{\mathcal{B}})$ eine Borel-meßbare Funktion. Wir betrachten eine Folge $Q_n \in \mathcal{M}_b(\mathcal{B})$, so daß $Q_0(D_h) = 0$. Dann gilt

- (a) $Q_n \xrightarrow{w} Q_0 \implies h \circ Q_n \xrightarrow{w} h \circ Q_0$
 (b) Für $(\tilde{\Xi}, \tilde{\mathcal{B}}) = (\mathbb{R}, \mathbb{B})$ und h zusätzlich beschränkt gilt außerdem

$$\lim_n \int h dQ_n = \int h dQ_0.$$

BEWEIS: Billingsley (1968), Theoreme 5.1 und 5.2, pp. 30, 31. ////

Ein topologischer Raum (Ξ, \mathcal{B}) heißt *polnisch*, wenn er vollständig metrisierbar und separabel ist.

Theorem A.1.2 (Prokhorov) Sei $\Pi \subset \mathcal{M}_b(\mathcal{B})$ eine Familie von Maßen auf dem topologischen Raum (Ξ, \mathcal{B}) . Dann gilt

direkte Hälfte Wenn Π straff und gleichmäßig beschränkt ist, so ist Π schwach relativ kompakt, d.h. jede Folge aus Π enthält eine schwach konvergente Teilfolge.

indirekte Hälfte Ist Ξ polnisch, so folgt aus der relativen Kompaktheit von Π auch die Straffheit und die gleichmäßige Beschränktheit von Π .

BEWEIS: Billingsley (1968), Theoreme 6.1 und 6.2, p. 37. ////

Sei nun $(\Xi, \mathcal{B}) = (\mathbb{R}^k, \mathbb{B}^k)$; zu einem Maß $Q \in \mathcal{M}_b(\mathbb{B}^k)$ definieren wir die *charakteristische Funktion*

$$\hat{Q}(t) := \int \exp(it^\tau x) Q(dx),$$

für die wir bei einer Zufallsvariablen X auch $\varphi_X(t)$ schreiben.

Proposition A.1.3 Sei $Q_n \in \mathcal{M}_b(\mathcal{B})$.

(a) Wenn $Q_n \xrightarrow{w} Q_0$ für ein $Q_0 \in \mathcal{M}_b(\mathbb{B}^k)$, dann $\hat{Q}_n(t) \rightarrow \hat{Q}_0(t)$ für jedes $t \in \mathbb{R}^k$, gleichmäßig auf t -Kompakta.

(b) Wenn $\hat{Q}_n(t) \rightarrow \varphi(t)$ für jedes t und $\varphi(t)$ stetig in $t = 0$, so gilt $\varphi = \hat{Q}_0$ für ein $Q_0 \in \mathcal{M}_b(\mathbb{B}^k)$ und $Q_n \xrightarrow{w} Q_0$.

BEWEIS: Billingsley (1968), Theoreme 7.6, p. 46. ////

A.2 Eigenschaften der Normalverteilung

A.2.1 bedingte Verteilungen bei Gaußschen Vektoren

Lemma A.2.1 Sei $(T^\tau, X^\tau)^\tau$ ein endlich dimensionaler, normalverteilter Zufallsvektor mit Verteilung

$$\mathcal{L} \left(\begin{array}{c} T \\ X \end{array} \right) = \mathcal{N} \left(\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} A & C \\ C^\tau & B \end{pmatrix} \right) \quad (\text{A.4})$$

mit Mittelwerten $a \in \mathbb{R}^p$, $b \in \mathbb{R}^k$ Kovarianz $C \in \mathbb{R}^{p \times k}$, und Varianzen $A = A^\tau \in \mathbb{R}^{p \times p}$, $0 \prec B = B^\tau \in \mathbb{R}^{k \times k}$. Dann ist die reguläre bedingte Verteilung von T gegeben $X = x$ $\lambda^k(dx)$ -f.s.

$$\mathcal{L}(T|X = x) = \mathcal{N}(a + CB^{-1}(x - b), A - CB^{-1}C^\tau) \quad (\text{A.5})$$

BEWEIS: Rieder (1994), Lemma 2.2.7 ////

A.2.2 elliptische Symmetrie

Lemma A.2.2 Sei $X \sim \mathcal{N}_p(\mu, \Sigma)$. Dann ist

$$\mathcal{L}(X) = \mathcal{L}(\mu + \Sigma^{\frac{1}{2}}Yu) \quad (\text{A.6})$$

mit

$$Y \sim \text{ufo}(\mathcal{S}_p), \quad u \sim \mathcal{L}(\sqrt{\chi_p^2}), \quad u, Y \text{ sto.u.} \quad (\text{A.7})$$

BEWEIS: Zu zeigen ist $\mathcal{L}(Yu) \stackrel{!}{=} \mathcal{N}_p(0, \mathbb{I}_p)$, was aber unmittelbar mit Transformation auf Polarkoordinaten folgt. Für allgemeines μ, Σ folgt die Behauptung aus der Invarianz der Normalverteilung unter affin-linearen Transformationen.

////

A.2.3 ein asymptotisches Cramér–Lévy–Theorem

In Abschnitt 6.3.1 verwenden wir folgende Charakterisierung der Normalverteilung bzw. deren asymptotisches Pendant, Theorem A.2.4,

Theorem A.2.3 (Cramér–Lévy) Seien X, Y unabhängige Zufallsvariablen mit Werten im \mathbb{R}^p . Dann sind folgende Aussagen äquivalent:

$$X, Y \text{ sind normalverteilt} \quad \iff \quad X + Y \text{ ist normalverteilt} \quad (\text{A.8})$$

BEWEIS: Feller (1971), Theorem 1, p. 525.

////

Dies setzt sich in folgende asymptotische Proposition um

Proposition A.2.4 Seien X_n, Y_n Folgen von Zufallsvariablen mit Werten im \mathbb{R}^p , X_n für jedes n stochastisch unabhängig von Y_n ; dann haben wir folgende Äquivalenz:

(a) (X_n, Y_n) sind straff und jede schwach konvergente Teilfolge (X_{n_k}, Y_{n_k}) ist asymptotisch normalverteilt.

(b) (X_n) oder (Y_n) ist straff und für jede schwach konvergente Teilfolge X_{n_k} resp. Y_{n_k} gilt $X_{n_k} + Y_{n_k} \xrightarrow{w} \mathcal{N}(\cdot, \cdot)$

BEWEIS: “ \Rightarrow ”

(X_n, Y_n) ist straff, also ist jede der beiden Komponenten (X_n) und (Y_n) straff und nach dem Theorem von Prokhorov, Theorem A.1.2 besitzt jede Teilfolge eine schwach konvergente (Teil-)Teilfolge, die nach Voraussetzung asymptotisch normalverteilt ist. Der schwache Limes heie (X, Y) . Wegen des Stetigkeitssatzes gilt nun $(X_n + Y_n) \xrightarrow{w} (X + Y)$, was wiederum als lineare Transformation ebenfalls normalverteilt ist.

“ \Leftarrow ”

Ohne Einschränkung der Allgemeinheit sei (X_n) straff und konvergiere schwach gegen X . Nach Voraussetzung konvergiert dann $X_n + Y_n$ gegen eine normalverteilte Variable S . In charakteristischen Funktionen ausgedrckt gilt gleichmig auf t -Kompakta

$$\varphi_{X_n+Y_n}(t) = \varphi_{X_n}(t)\varphi_{Y_n}(t) \rightarrow \varphi_S(t), \quad \varphi_{X_n}(t) \rightarrow \varphi_X(t).$$

Wegen der Normalität von S aber ist $\varphi_S(t) \neq 0$ für alle t ; also ist auch $\varphi_X(t) \neq 0$; daher gilt für n hinreichend groß, $\varphi_{X_n}(t)\varphi_{Y_n}(t) \neq 0$, so daß wir durch $\varphi_{X_n}(t)$ dividieren dürfen, und so sehen, daß

$$\varphi_{Y_n}(t) = \varphi_{X_n+Y_n}(t)/\varphi_{X_n}(t) \rightarrow \varphi_S(t)/\varphi_X(t) =: \varphi_Y(t),$$

und φ_Y ist charakteristische Funktion einer Zufallsvariablen.

Nun ist der \mathbb{R}^p polnisch, und so folgt mit der indirekten Hälfte des Theorems von Prokhorov, daß Y_n und (X_n, Y_n) als schwach konvergente Folgen jeweils straff sind. Außerdem haben wir $\varphi_S(t) = \varphi_X(t)\varphi_Y(t)$; daher ist S die Summe unabhängiger Zufallsvariablen X und Y und nach Voraussetzungen normalverteilt, so daß das Cramér-Lévy-Theorem A.2.3 anwendbar ist und ergibt daß X, Y beide normalverteilt sind. ////

Bemerkung A.2.5 Unter Normalität im Theorem A.2.3 oder asymptotischer Normalität in Proposition A.2.4 wollen wir auch Varianz 0 zulassen, also erlauben, daß die Variablen fast sicher 0 respektive $o_P(n^0)$ sein dürfen. In Proposition A.2.4 ist der Fall des Zentralen Grenzwertsatzes von Lindeberg-Lévy durch letzteres abgedeckt.

Als unmittelbare Folgerung von Theorem A.2.4 notieren wir

Korollar A.2.6 Sei X_t ein d -dimensionaler $AR(p)$ Prozeß, das heißt $X_t = \sum_{k=1}^p F_k^{(t)} X_{t-k} + v_t$ für gewisse Matrizen $F_k^{(t)} \in \mathbb{R}^{p \times p}$ und einen Innovationsprozeß v_t , so daß v_t unabhängig von X_s und v_s für $s < t$ ist. Dann ist eine notwendige Bedingung für die asymptotische Normalität von X_t die asymptotische Normalität von v_t .

A.3 Stationarität und Ergodizität

Ausgehend von Breiman (1968) präsentieren wir eine Darstellung für beidseitig unendliche Prozesse mit diskretem Zeitbereich, die wir einer Vorlesung von H. Rieder entnommen haben.

Definition A.3.1 Sei $\{X_k\} = \dots, X_{-2}, X_{-1}, X_0, \dots$, ein \mathbb{R}^p -wertiger Prozeß auf (Ω, \mathcal{A}, P) .

(a) Sei π_k die Projektion auf die k -te Koordinate. Dann definieren wir den Shiftoperator $\mathcal{S} : \mathbb{R}^{\mathbb{Z}} \rightarrow \mathbb{R}^{\mathbb{Z}}$ als $\pi_k \mathcal{S}(\{X_k\}) = \pi_{k+1}(\{X_k\})$.

(b) $\{X_k\}$ heißt strikt stationär oder auch nur stationär, falls

$$\mathcal{L}(\{X_k\}) = \mathcal{L}(\mathcal{S}\{X_k\}).$$

$\{X_k\}$ heißt (strikt) stationär m -ter Ordnung, falls für jedes $t \in \mathbb{Z}$ gilt

$$\mathcal{L}(X_1, \dots, X_m) = \mathcal{L}(X_{1+t}, \dots, X_{t+m+1}).$$

(c) $\{X_k\}$ heißt schwach stationär, falls für alle $t \in \mathbb{Z}$ gilt

$$E X_t = \mu, \quad \text{Cov}(X_{h+1}, X_1) = \text{Cov}(X_{t+h}, X_t).$$

(d) $\{X_k\}$ heißt ergodisch, falls für jedes X -meßbare $A \in \mathbb{B}^{\mathbb{Z}}$ mit $\mathcal{S}(A) = A$ gilt $X \circ P(A) = 1$ oder $X \circ P(A) = 0$.

Lemma A.3.2 Sei $\{X_k\} = \dots, X_{-2}, X_{-1}, X_0, \dots$ ein stationärer, \mathbb{R}^p -wertiger Prozeß auf (Ω, \mathcal{A}, P) , χ eine $\mathbb{B}^{\mathbb{Z}}$ -meßbare Funktion. Dann ist der Prozeß $\{Y_k\}$ definiert durch $Y_k := \chi(\dots, X_{k-1}, X_k)$ stationär. Ist $\{X_k\}$ sogar ergodisch, so überträgt sich auch diese Eigenschaft auf $\{Y_k\}$.

BEWEIS: Es gilt:

$$\pi_k \circ \mathcal{S} \circ \chi \circ X = \pi_{k+1} \circ \chi \circ X = Y_{k+1} = \pi_k \circ \chi \circ \mathcal{S} \circ X$$

und damit

$$\mathcal{S} \circ \chi \circ X = \chi \circ \mathcal{S} \circ X. \quad (\text{A.9})$$

Sei \mathcal{I} die σ -Algebra der shiftinvarianten Ereignisse in $\mathbb{B}^{\mathbb{Z}}$, $X^{-1}(\mathcal{I}) =: \mathcal{I}_X \subset \mathcal{A}$ die σ -Algebra der X -invarianten Ereignisse und $Y^{-1}(\mathcal{I}) =: \mathcal{I}_Y \subset \mathcal{A}$ die der Y -invarianten Ereignisse. Dann ist

$$\mathcal{I}_Y \subset \mathcal{I}_X, \quad (\text{A.10})$$

denn mit $I_Y := Y^{-1}(I) \in \mathcal{I}_Y$ für ein $I \in \mathcal{I}$ gilt

$$I_Y = (\chi \circ X)^{-1}(I) = X^{-1} \circ \chi^{-1}(I).$$

Dabei ist wegen (9) $\chi^{-1}(I) \in \mathcal{I}$ und demnach $X^{-1} \circ \chi^{-1}(I) \in \mathcal{I}_X$. Also gilt für beliebiges $B \in \mathbb{B}^{\mathbb{Z}}$:

$$\begin{aligned} P(\mathcal{S} \circ Y \in B) &= P(\mathcal{S} \circ \chi \circ X \in B) \stackrel{(9)}{=} P(\chi \circ \mathcal{S} \circ X \in B) = \\ &= P(X \in \mathcal{S}^{-1} \circ \chi^{-1} B) = P(X \in \chi^{-1} B) = P(\chi \circ X \in B) = P(Y \in B) \end{aligned}$$

Da B beliebig war, folgt die Stationarität, und da nach (10) $\mathcal{I}_Y \subset \mathcal{I}_X$, überträgt sich auch die Ergodizität. ////

Lemma A.3.3 Sei $\{X_k\} = \dots, X_{-2}, X_{-1}, X_0, \dots$ ein stationärer, \mathbb{R}^p -wertiger Prozeß auf (Ω, \mathcal{A}, P) , $\mathcal{F} = X \circ \mathcal{A}$ die von X erzeugte Bild- σ -Algebra und für $j \in \mathbb{Z}$ sei \mathcal{F}_j die von \dots, X_{j-1}, X_j erzeugte Unter- σ -Algebra von \mathcal{F} . Weiter sei $\mathcal{G}_j \subset \mathcal{F}_{j+m}$ eine Glättungsfiltrierung im Sinne von Abschnitt 10.3.3. Dann ist auch $\{Y_k\} := (\mathbb{E}[X_k | \mathcal{G}_k])_k$ stationär. Ist $\{X_k\}$ ergodisch, so auch $\{Y_k\}$.

BEWEIS ZU LEMMA A.3.3: Es genügt Zylindermengen zu betrachten; sei dazu $B \in \mathcal{G}_k$, $B = \dots \times B_{k+m-1} \times B_{k+m}$, $B_i \in \mathbb{B}$. Dann gilt für $j \in \mathbb{Z}$

$$\begin{aligned} P(X \in B) &= P(X_{k+m} \in B_{k+m}, X_{k+m-1} \in B_{k+m-1}, \dots) = \\ &= P(X_{k+m-j} \in B_{k+m}, X_{k+m-j-1} \in B_{k+m-1}, \dots) = P(S^{-j} \circ X \in B) \end{aligned}$$

und damit gilt

$$\mathcal{G}_k = \mathcal{G}_{k-j}. \quad (\text{A.11})$$

Für die Radon-Nikodym-Gleichungen gilt demnach

$$\int_B y_k dP^{Y_k} = \int_B x_k dP^{X_k} \stackrel{\text{Stat.}}{=} \int_B x_{k-l} dP^{X_{k-l}} \stackrel{(11)}{=} \int_B y_{k-l} dP^{Y_{k-l}}$$

Die Ergodizität überträgt sich mit dem gleichen Argument wie in Lemma A.3.2. ////

A.4 gleichgradige Integrierbarkeit

Einige der Resultate in dieser Arbeit, vor allem in Kapitel 9, verwenden den Begriff der gleichgradigen Integrierbarkeit. Seien dazu

$$X_s : (\Omega_s, \mathcal{A}_s, P_s) \rightarrow (\mathbb{R}^k, \mathbb{B}^k), \quad s \in T \quad (\text{A.12})$$

eine Familie von Zufallsvariablen mit Indexbereich T . Diese Familie heißt *gleichgradig integrierbar*, falls

$$\lim_{c \rightarrow \infty} \sup_{s \in T} \int_{|X_s| > c} |X_s| dP_s = 0 \quad (\text{A.13})$$

Proposition A.4.1 Für eine Familie $\{X_s\}_{s \in T}$ aus (12) sind folgende Eigenschaften äquivalent

- (i) Die Familie $\{X_s\}_{s \in T}$ ist gleichgradig integrierbar
- (ii) (a) $\sup_s E |X_s| < \infty$
 (b) Zu $\varepsilon > 0$ gibt es ein $\delta > 0$, so daß für alle Familien $\{A_s\}$, $A_s \in \mathcal{A}_s$ mit $P_s(A_s) \leq \delta$ gilt $\int_{A_s} |X_s| dP_s \leq \varepsilon$.
- (iii) Es existiert eine positive Funktion q auf \mathbb{R}_+ so daß $\lim_{t \rightarrow \infty} q(t)/t = \infty$ und

$$\sup_s E_{P_s} [q(|X_s|)] < \infty. \quad (\text{A.14})$$

Dabei kann q monoton wachsend und konvex gewählt werden.

BEWEIS: (i) \iff (ii) e.g. Chung (1974) Thm. 4.5.3, (i) \iff (iii) geht auf de Vallée–Poussin zurück und findet sich im Fall, daß $(\Omega_s, \mathcal{A}_s, P_s) \equiv (\Omega, \mathcal{A}, P)$ bewiesen in Dellacherie und Meyer (1978), Thm. II-22. Die Verallgemeinerung auf variierende W-Räume ist klar. ////

Korollar A.4.2 Sei $|X_s| \leq |Y_s|$, Y_s gleichgradig integrierbar, dann ist auch X_s gleichgradig integrierbar.

BEWEIS:

$$0 \leq \lim_{c \rightarrow \infty} \sup_{s \in T} \int_{|X_s| > c} |X_s| dP_s \leq \lim_{c \rightarrow \infty} \sup_{s \in T} \int_{|Y_s| > c} |Y_s| dP_s = 0$$

////

Theorem A.4.3 (Vitali) Sei $0 < r < \infty$ und für eine Folge $\{X_n\}_{n \in \mathbb{N}}$ von Zufallsvariablen auf (Ω, \mathcal{A}, P) gelte $X_n = X_0 + o_P(n^0)$ und $X_n \in L_r(P)$. Dann sind folgende Aussagen äquivalent

- (i) Die Folge $\{|X_n|^r\}_{n \in \mathbb{N}}$ ist gleichgradig integrierbar.
- (ii) $\lim_{n \rightarrow \infty} E |X_n - X_0|^r = 0$
- (iii) $\lim_{n \rightarrow \infty} E |X_n|^r = E |X_0|^r < \infty$
- (iv) $\limsup_{n \rightarrow \infty} E |X_n|^r \leq E |X_0|^r < \infty$

Falls $r \geq 1$ und die Folge $\{|X_n|^r\}_{n \in \mathbb{N}}$ gleichgradig integrierbar ist, so folgt $\lim_{n \rightarrow \infty} E X_n^r = E X_0^r$.

Eigenschaft (ii) impliziert gleichgradige Integrierbarkeit.

BEWEIS: Rieder (1994), Thm. A.2.2. ////

A.5 Abriß der lokal–asymptotischen Optimalitätstheorie der Statistik

An dieser Stelle wollen wir einige Elemente des lokal–asymptotischen Ansatzes der Statistik zusammentragen, die mit den Namen Le Cam und Hájek verbunden sind und die wir Rieder (1994) entnommen haben:

Theorem A.5.1 Sei $(\Omega_n, \mathcal{A}_n)_n$ eine Folge meßbarer Räume, $P_n, Q_n \in \mathcal{M}_1(\mathcal{A}_n)$ zwei Folgen von Wahrscheinlichkeitsmaßen mit Log–Likelihoods $L_n \in \log dQ_n/dP_n$ und S_n eine Folge von $(\bar{\mathbb{R}}^p, \bar{\mathbb{B}}^p)$ –meßbarer Statistiken mit

$$\begin{pmatrix} S_n \\ L_n \end{pmatrix} \circ P_n \xrightarrow{w} \mathcal{N}\left(\begin{pmatrix} a \\ -\sigma^2/2 \end{pmatrix}, \begin{pmatrix} A & c \\ c^\tau & \sigma^2 \end{pmatrix}\right) \quad (\text{A.15})$$

für gewisse $a, c \in \mathbb{R}$, $0 < \sigma < \infty$, $A \in \mathbb{R}^{p \times p}$. Dann gilt:

$$\begin{pmatrix} S_n \\ L_n \end{pmatrix} \circ Q_n \xrightarrow{w} \mathcal{N}\left(\begin{pmatrix} a+c \\ \sigma^2/2 \end{pmatrix}, \begin{pmatrix} A & c \\ c^\tau & \sigma^2 \end{pmatrix}\right) \quad (\text{A.16})$$

Definition A.5.2 Eine Folge von parametrisierten Modellen (\mathcal{Q}_n) ,

$$\mathcal{Q}_n = \{Q_{n,t} \mid t \in \Theta\} \subset \mathcal{M}_1(\mathcal{A}_n),$$

mit offenem Parameterbereich $\emptyset \neq \Theta \subset \mathbb{R}^k$, heie as. normal, falls es eine Folge $Z = (Z_n)$ von $\mathcal{A}_n - \bar{\mathbb{B}}^k$ –mebaren, as. normalen Zufallsgroen gibt,

$$Z_n \circ Q_{n,0} \xrightarrow{w} \mathcal{N}(0, C) \quad (\text{A.17})$$

mit $C \in \mathbb{R}^{k \times k}$ einer positiv definiten Matrix, so da sich fur alle $t \in \mathbb{R}^k$ die Log–Likelihoods $L_{n,t} \in \log dQ_{n,t}/dQ_{n,0}$ stochastisch approximieren lassen als

$$L_{n,t} = t^\tau Z_n - \frac{1}{2} t^\tau C t + o_{Q_{n,0}}(n^0). \quad (\text{A.18})$$

Die Folge Z_n heit as. suffiziente Statistik, die Matrix C as. Kovarianz der Modelle (\mathcal{Q}_n) .

Wir notieren in diesem Zusammenhang auch Bemerkung 2.2.10 aus Rieder (1994) als

Bemerkung A.5.3 (a) Die Kovarianzmatrix C ist eindeutig durch (17) und (18) bestimmt. (18) impliziert auch, da eine andere Folge von Statistiken $W = (W_n)$ as. suffizient genau dann, wenn $W_n = Z_n + o_{Q_{n,0}}(n^0)$.

(b) Sehen wir von der Approximation ab, konnen wir den Begriff ‘as. suffizient’ durch das Neyman–Kriterium rechtfertigen.

(c) Wir sprechen von *lokal asymptotischer Normalitt* oder kurz *LAN*, wenn – wie im Kontext der L_2 –Differenzierbarkeit – die as. Normalitt von geeigneten lokalen Umparametrisierungen abhangt.

Definition A.5.4 Sei $S = (S_n)$, $S_n : (\Omega_n, \mathcal{A}_n) \rightarrow (\bar{\mathbb{R}}^p, \bar{\mathbb{B}}^p)$ ein as. Schatzer fur die Parametertransformation D , $D \in \mathbb{R}^{p \times k}$, $\text{rk } D = p \leq k$. S heie regulr mit Grenzverteilung $\mathcal{M} \in \mathcal{M}_\infty(\bar{\mathbb{B}}^p)$, falls fur alle $t \in \mathbb{R}$

$$(S_n - Dt) \circ Q_{n,t} \xrightarrow{w} M, \quad (\text{A.19})$$

also $S_n \circ Q_{n,t} \xrightarrow{w} M * I_{Dt}$ fur $n \rightarrow \infty$ punktweise in $t \in \Theta$.

Theorem A.5.5 (Faltungssatz) Seien (Q_n) eine Folge von as. normalen parametrisierten Modellen mit as. suffizienter Statistik Z und as. Kovarianz C . Weiter sei $D \in \mathbb{R}^{p \times k}$, $\text{rk } D = p \leq k$ und S ein as. Schätzer, der regulär für D mit Grenzverteilung M sei. Dann existiert eine Wahrscheinlichkeit $M_0 \in \mathcal{M}_1(\mathbb{B}^p)$, so daß

$$M = M_0 * \mathcal{N}(0, \Gamma), \quad \Gamma = DC^{-1}D^\tau \quad (\text{A.20})$$

und

$$(S_n - DC^{-1}Z_n) \circ Q_{n,0} \xrightarrow{w} M_0 \quad (\text{A.21})$$

Ein as. Schätzer S^* ist regulär für D und erreicht die konzentrierteste Grenzverteilung $M^* = \mathcal{N}(0, \Gamma)$ dann und nur dann, wenn

$$S_n^* = DC^{-1}Z_n + o_{Q_{n,0}}(n^0) \quad (\text{A.22})$$

Im Zusammenhang mit Definitionen A.5.2 und A.5.4 stellt sich die Frage, ob bei Existenz von regulären Schätzern bereits die as. Kovarianz C vollen Rang haben muß — oder im Kontext der L_2 -Differenzierbarkeit lokale Identifizierbarkeit im Sinn von Definition 10.4.1 vorliegen muß. Dies wird in der folgenden Proposition geklärt, die gemeinsam mit H. Rieder entstanden ist:

Proposition A.5.6 Besitzt eine Folge parametrischer Modelle (Q_n) wie in Definition A.5.2 die LAN-Eigenschaft, so kann es reguläre Schätzer nur dann geben, wenn die as. Kovarianz C vollen Rang hat.

BEWEIS: Sei $\det C = 0$, $0 \neq t_0 \in \ker C$. Dann ist für L_{n,t_0} wie in Definition A.5.2 notgedrungen

$$L_{n,t_0} = o_{Q_{n,0}}(n^0) \quad (\text{A.23})$$

Wegen der Benachbarkeit von (Q_{n,t_0}) und $(Q_{n,0})$, die sich — auch bei $Ct_0 = 0!$ — aus der LAN-Eigenschaft ergibt, ist

$$Q_{n,t_0}(dQ_{n,0} = 0) = o(n^0) \quad (\text{A.24})$$

Weiter gilt

$$\limsup_n \int \exp(L_{n,t_0}) dQ_{n,0} \leq 1 = \int 1 dQ_{n,0}, \quad (\text{A.25})$$

somit folgt aus (23) mit dem Satz von Vitali A.4.3(d)

$$\int |\exp(L_{n,t_0}) - 1| dQ_{n,0} = o(n^0) \quad (\text{A.26})$$

Dies zusammen mit (24) gibt

$$d_V(Q_{n,t_0}, Q_{n,0}) = \frac{1}{2} \int |dQ_{n,t_0} - dQ_{n,0}| = o(n^0), \quad (\text{A.27})$$

so daß insbesondere für jeden Schätzer S_n gilt

$$d_V(\mathcal{L}_{Q_{n,0}}(S_n), \mathcal{L}_{Q_{n,t_0}}(S_n)) = o(n^0) \quad (\text{A.28})$$

Angenommen nun $S_n \circ Q_{n,0} \xrightarrow{w} M$, so folgt hieraus auch $S_n \circ Q_{n,t_0} \xrightarrow{w} M$, also $(S_n - t_0) \circ Q_{n,t_0} \xrightarrow{w} M * I_{-t_0}$, so daß S_n nicht regulär sein kann. ///

Definition A.5.7 Sei \mathcal{S} die Klasse aller as. Schätzer $S = (S_n)$, $S_n : (\Omega_n, \mathcal{A}_n) \rightarrow (\mathbb{R}^p, \mathbb{B}^p)$ und \mathbf{L} die Klasse aller Funktionen $\ell : \mathbb{R}^p \rightarrow [0, \infty]$ mit

- (a) ℓ ist unterhalbstetig, also $\liminf \ell(z_n) \geq \ell(z)$ für $\mathbb{R}^p \ni z_n \rightarrow z \in \mathbb{R}^p$.
(b) ℓ ist stetig im Unendlichen, d.h. für $\mathbb{R}^p \ni z_n \rightarrow z \in \mathbb{R}^p \setminus \mathbb{R}^p$ gilt $\lim \ell(z_n) = \ell(z)$.
(c) ℓ ist symmetrisch subkonvex, also für alle $z \in \mathbb{R}^p$ und $c \geq 0$ gelte

$$\ell(z) = \ell(-z), \quad \{z \in \mathbb{R}^p \mid \ell(z) \leq c\} \text{ ist konvex} \quad (\text{A.29})$$

Bemerkung A.5.8 Weil wir es nur auf asymptotische Aussagen abgesehen haben, bei denen man stets die Unterhalbstetigkeit der Verlustfunktion annehmen muß, schließen wir diese im Unterschied zu Rieder (1994) mit in die Definition von \mathbf{L} ein und verlangen daher nicht nur Oberhalbstetigkeit, sondern Stetigkeit im Unendlichen, wie sie dort auch im as. Minimax-Theorem benötigt wird.

Theorem A.5.9 (as. Minimax-Theorem) Sei (Q_n) eine Folge von as. normalen parametrisierten Modellen mit as. suffizienter Statistik Z , und as. Kovarianz C . Weiter sei $D \in \mathbb{R}^{p \times k}$, $\text{rk } D = p \leq k$ und sei $\ell \in \mathbf{L}$ beliebig.

$$\rho_0 = \int \ell d\mathcal{N}(0, \Gamma), \quad \Gamma = DC^{-1}D^\tau. \quad (\text{A.30})$$

Dann gilt

(a)

$$\lim_{b \rightarrow \infty} \lim_{c \rightarrow \infty} \liminf_{n \rightarrow \infty} \inf_{S \in \mathcal{S}} \sup_{t \in \Theta, |t| \leq c} \int b \wedge \ell(S_n - Dt) dQ_{n,t} \geq \rho_0 \quad (\text{A.31})$$

(b) Ist $\ell : \mathbb{R}^p \rightarrow [0, \infty]$ λ^p -f.s. stetig und der Schätzer S^* lokal gleichmäßig as. normal, in dem Sinn, daß für den Prokhorovabstand d_π

$$\sup_{t \in \Theta, |t| \leq c} d_\pi(\mathcal{L}_{Q_{n,t}}(S_n^* - Dt), \mathcal{N}(0, \Gamma)) = o(n^0), \quad (\text{A.32})$$

so gilt für alle $c \in (0, \infty)$

$$\lim_{b \rightarrow \infty} \liminf_{n \rightarrow \infty} \sup_{t \in \Theta, |t| \leq c} \int b \wedge \ell(S_n^* - Dt) dQ_{n,t} = \rho_0 \quad (\text{A.33})$$

und notwendigerweise besitzt S^* die as. Darstellung

$$S_n^* = DC^{-1}Z_n + o_{Q_{n,0}}(n^0).$$

Anhang B

multivariate Begriffe für die asymptotische Statistik

B.1 Begriffe/Notation aus der multivariaten Statistik / der multilinearen Algebra

Da wir an verschiedenen Punkten dieser Arbeit matrixwertige Parameter behandeln, stellen wir an dieser Stelle Notation zusammen, die es erlaubt, je nach Zweck die Parameter als Matrizen oder als Vektoren aufzufassen. Dazu verwenden wir die Operatoren

$$\text{vec} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{mn}, \quad A = (a_{i,j})_{\substack{i=1,\dots,m \\ j=1,\dots,n}} \mapsto \text{vec}(A) = (a_{(i-1)n+j})_{\substack{i=1,\dots,m \\ j=1,\dots,n}} \quad (\text{B.1})$$

$$\text{svec} : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{\binom{n}{2}}, \quad A = (a_{i,j})_{i,j=1,\dots,n} \mapsto \text{svec}(A) = (a_{n(i-1) - \binom{i}{2} + j})_{\substack{i \leq j \\ i,j=1,\dots,n}} \quad (\text{B.2})$$

Außerdem verwenden wir die Notation $A \otimes B$ für das *Kroneckerprodukt* von A und B definiert als Matrix mit Einträgen

$$(A \otimes B)_{(i,j);(k,l)} := A_{i,k} B_{j,l}$$

Schließlich verwenden wir die *Kommutator-Matrix* (vgl. Magnus und Neudecker (1999)) $K_{p,q} \in \mathbb{R}^{pq \times pq}$ definiert über ihre Bilder:

Für $A \in \mathbb{R}^{p \times q}$ sei $K_{p,q} \text{vec}(A) = \text{vec}(A^T)$.

Bemerkung B.1.1 In vielen Büchern zur multivariaten Statistik finden sich ähnliche Operatoren; der hier verwendete $\text{vec}(\cdot)$ -Operator ist andernorts allerdings meist so definiert, daß zuerst die Spalten und dann die Zeilen durchlaufen werden, so daß dann geringfügige Unterschiede auftreten:

So gilt mit unserer Definition

$$\text{vec}(ABC) = A \otimes C^T \text{vec}(B),$$

während es in der Konvention von z.B. Searle (1997), Anderson (1993), Lütkepohl (1993), Bellman (1995), Reinsel (1997) heißen muß

$$\text{vec}(ABC) = C^T \otimes A \text{vec}(B).$$

In Graybill (1983) wiederum ist das Kroneckerprodukt als linkes direktes Produkt definiert, der $\text{vec}(\cdot)$ -Operator aber wie in den oben genannten Quellen. Unserem Operator $\text{svec}(\cdot)$ entspricht bei Searle (1997) und Lütkepohl (1993) der Operator $\text{vech}(\cdot)$ mit denselben Einschränkungen wie bei $\text{vec}(\cdot)$.

Aus der multivariaten Statistik verwenden wir den Begriff der vierten Kumulante einer multivariaten Zufallsvariablen X . Diese — in Zeichen $\kappa^{(X)}$ — schreibt sich für eine zentrierte Variable X als

$$\kappa_{i_1, i_2, j_1, j_2}^{(X)} := E[X_{i_1} X_{i_2} X_{j_1} X_{j_2}] - E[X_{i_1} X_{i_2}] E[X_{j_1} X_{j_2}] - E[X_{i_1} X_{j_1}] E[X_{i_2} X_{j_2}] - E[X_{i_1} X_{j_2}] E[X_{i_2} X_{j_1}]$$

B.2 Fisher-Information des multivariaten Lokations-/Skalen-Modells

In diesem Abschnitt wollen wir analog zu einem Resultat von Huber (c.f. Huber (1964) oder Huber (1981)) die Eigenschaft charakterisieren, daß das multivariate Lokations-/Skalenmodell zu einem $F \in \mathcal{M}_1(\mathbb{B}^k)$,

$$\mathcal{P} := \{F(\Sigma^{-1}(\cdot - \theta)) \mid \theta \in \mathbb{R}^k, \Sigma = \Sigma^\tau \succ 0\}, \tag{B.3}$$

endliche Lokations-/Skalen-Fisher-Information annimmt. Wir erinnern hierzu an

Theorem B.2.1 *Es gilt folgende Äquivalenz für die Fisher-Information im Lokationsmodell $\mathcal{P} = \{F_\theta(dx) := F(dx - \theta) \mid \theta \in \mathbb{R}\}$,*

$$\mathcal{I}(F) := \sup \frac{(\int \varphi' dF)^2}{\int \varphi^2 dF}, \quad 0 \neq \varphi \in \mathcal{D}_1, \tag{B.4}$$

mit $\mathcal{D}_1 := C_c^\infty(\mathbb{R}, \mathbb{R})$, der Menge aller beliebig oft differenzierbaren Funktionen¹ von \mathbb{R} nach \mathbb{R} mit kompaktem Träger.

- (i) $\mathcal{I}(F) < \infty$
- (ii) F besitzt eine absolutstetige Dichte f mit $\int (f'/f)^2 f dx < \infty$.

In beiden Fällen ist $\mathcal{I}(F) = \int (f'/f)^2 f dx$.

BEWEIS ZU THEOREM B.2.1: Theorem 4.2 in Huber (1981). ////

Dies verallgemeinern wir zunächst auf den eindimensionalen Skalenfall:

Proposition B.2.2 *Wir betrachten das eindimensionale Skalenmodell*

$$\mathcal{P} = \{F_\sigma(dx) := F(dx/\sigma) \mid \sigma > 0\}.$$

Dann gilt: Für die Fisher-Information im Skalenmodell,

$$\mathcal{J}(F) := \sup \frac{(\int x \varphi' dF)^2}{\int \varphi^2 dF}, \quad 0 \neq \varphi \in \mathcal{D}_1 \tag{B.5}$$

sind folgende Aussagen äquivalent

¹Im Originalbeweis verwendet Huber C_c^1 ; da aber auch \mathcal{D}_1 dicht in L_2 liegt, ist es keine Einschränkung, auf der kleineren Menge \mathcal{D}_1 zu argumentieren.

(i) $\mathcal{J}(F) < \infty$ und $F(\{0\}) = 0$.

(ii) F besitzt eine Dichte f , so daß xf absolutstetig ist mit Ableitung² $(1 + x(f'/f))f$ und $\int(1 + x(f'/f))^2 f dx < \infty$

In beiden Fällen ist $\mathcal{J}(F) = \int(1 + x(f'/f))^2 f dx$

Vor dem Beweis noch ein Lemma

Lemma B.2.3 Sei F ein Wahrscheinlichkeitsmaß auf (\mathbb{R}, \mathbb{B}) . Dann ist $\mathcal{D}_1^0 := \{\psi \mid \psi = x\varphi \text{ für ein } \varphi \in \mathcal{D}_1\}$ dicht in $L_2(F)$ genau dann, wenn $F(\{0\}) = 0$.

BEWEIS ZU LEMMA B.2.3: Angenommen $F(\{0\}) \neq 0$ und $g = \mathbb{I}_{\{0\}}$. Dann gilt für $h \in \mathcal{D}_1^0$, $\|g - h\|_{L_2(F)} \geq F(\{0\})$. Umgekehrt sei $F(\{0\}) = 0$. Es genügt zu zeigen, daß jedes $g = \mathbb{I}_A$ für $A \in \mathbb{B}$ durch $h \in \mathcal{D}_1^0$ approximierbar ist. Zu $\varepsilon > 0$ existieren wegen der σ -Stetigkeit von F Zahlen $a_1 < 0$ und $a_2 > 0$, so daß $F((a_1, a_2)) < \varepsilon/2$. Seien $A_1 := A \cap (-\infty, a_1]$, $A_2 := A \cap [a_2, \infty)$. Da \mathcal{D}_1 dicht in $L_2(F)$ ist, gibt es für $i = 1, 2$ h_i in \mathcal{D}_1 , so daß $\|\mathbb{I}_{A_i} - h_i\|_{L_2(F)} < \varepsilon/4$. Auf A_i lassen sich h_i darstellen als $x\tilde{h}_i$ mit \tilde{h}_i dort beliebig oft differenzierbar. Diese \tilde{h}_i lassen sich wiederum auf (a_1, a_2) beliebig oft differenzierbar fortsetzen durch ein $h_3 \in \mathcal{D}_1$ mit $|h_3| \leq 1$ auf (a_1, a_2) . Für das durch die Verknüpfung von h_i , $i = 1, 3, 2$ entstehende h gilt nun $\|g - h\|_{L_2(F)} \leq \varepsilon$. ////

BEWEIS ZU PROPOSITION B.2.2: (ii) \Rightarrow (i)

$$\begin{aligned} & \frac{(\int x\varphi' dF)^2}{\int \varphi^2 dF} \stackrel{(ii)}{=} \frac{(\int x\varphi' f dx)^2}{\int \varphi^2 f dx} = \\ \stackrel{\text{part. Int.}}{=} & \frac{(\int \varphi(1 + x(f'/f))f dx)^2}{\int \varphi^2 f dx} \stackrel{\text{Cauchy-Schwartz}}{\leq} \int (1 + x(f'/f))^2 f dx < \infty \end{aligned}$$

und da F eine Lebesgue-Dichte besitzt, ist insbesondere $F(\{0\}) = 0$.

(i) \Rightarrow (ii)

Wir definieren das lineare Funktional A auf der dichten Menge \mathcal{D}_1 des $L_2(F)$ als $A\varphi := -\int \varphi' x dF$. Da

$$\|A\varphi\|^2 \leq \mathcal{J}(F)\|\varphi\|_{L_2(F)}^2,$$

ist A stetig, also auf $L_2(F)$ stetig fortsetzbar. Der Satz von Riesz-Fréchet liefert die Existenz eines Elementes $g \in L_2(F)$, so daß $A\varphi = \int \varphi g dF$. Nun setzen wir

$$j(x, y) := \mathbb{I}_{y < x} g(y)/x, \quad f(x) := \int j(x, y) F(dy).$$

Offenbar ist $|j(x, y)x\varphi(x)|$ für $\varphi \in \mathcal{D}_1$ in $L_1(\lambda_\varphi(dx) \otimes F(dy))$ mit λ_φ , dem Lebesguemaß eingeschränkt auf den Träger von φ , daher folgt mit Fubini

$$-\int x\varphi'(x)f(x) dx = -\int \int_{y < x} \varphi'(x) dx g(y) F(dy) = \int \varphi(y)g(y) F(dy) = A\varphi.$$

Nach Lemma B.2.3 ist \mathcal{D}_1^0 dicht in $L_2(F)$; auf \mathcal{D}_1^0 aber definieren die beiden Maße $F(dx)$ und $f dx$ dasselbe Funktional, also stimmen die Maße überein,

²Dabei ist der Term $(1 + x(f'/f))f$ als schwache Ableitung zu lesen; diese stimmt mit der herkömmlichen überein, sofern diese existiert.

und f ist eine Version der Dichte von F .

Damit aber ist im Sinne einer Distributionsableitung³

$$A\varphi = - \int \varphi' x f dx \stackrel{\text{part.Int.}}{=} \int \varphi(1 + x(f'/f)) f dx,$$

also $g = (1 + x(f'/f))$ und $\|A\| = \int g^2 dF = \mathcal{J}(F)$. ////

Seien die Zahlen $\bar{\mathcal{I}}_f, \bar{\mathcal{J}}_f$ gegeben als

$$\bar{\mathcal{I}}_f := \sup \frac{(\int \nabla \varphi dF)^\tau (\int \nabla \varphi dF)}{\int \varphi^2 dF}, \quad 0 \stackrel{[F]}{\neq} \varphi \in \mathcal{D}_k, \quad (\text{B.6})$$

$$\bar{\mathcal{J}}_f := \sup \frac{\|\int x \nabla \varphi(x)^\tau F(dx)\|^2}{\int \varphi^2 dF}, \quad 0 \stackrel{[F]}{\neq} \varphi \in \mathcal{D}_k, \quad (\text{B.7})$$

wobei $\mathcal{D}_k := \mathcal{C}_c^\infty(\mathbb{R}^k, \mathbb{R})$ die Menge aller beliebig oft stetig, differenzierbaren Funktionen von \mathbb{R}^k nach \mathbb{R} mit kompaktem Träger ist.

$\bar{\mathcal{I}}_f$ bzw. $\bar{\mathcal{J}}_f$ werden die Rolle des Betrags der multivariaten Lokations- resp. Skalen-Fischer-Information spielen.

Theorem B.2.4 Die Endlichkeit von $\bar{\mathcal{I}}_f$, bzw. die von $\bar{\mathcal{J}}_f$ zusammen mit

$$F(\{x_i = 0\}) = 0, \quad i = 1, \dots, k \quad (\text{B.8})$$

ist äquivalent mit der Existenz einer Lebesguedichte f von F , wobei f respektive $x f$ λ^k -stetig sind und wir den letzten Ausdruck koordinatenweise verstehen. Weiter gilt λ^{k-1} -fast sicher, daß, alle Argumente bis auf eines festgehalten, f respektive $x f$ in diesem einen Argument absolutstetig ist.

Ist jeweils eine der beiden äquivalenten Bedingungen erfüllt, so ist

$$\bar{\mathcal{I}}_f = \int |\nabla f/f|^2 dF \quad \text{bzw.} \quad (\text{B.9})$$

$$\bar{\mathcal{J}}_f = \int |\mathbb{I}_k + \frac{\nabla f}{f} X^\tau|^2 dF \quad (\text{B.10})$$

Vor dem Beweis dieses Theorems noch einige Vorbereitungen:

Bei der Anwendung des Theorems auf Proposition B.3.1 benötigen wir zumindest Absolutstetigkeit von f respektive $x f$ längs (fast allen) achsenparallelen Linien. Dies wird geliefert durch

Proposition B.2.5 Jede Funktion f , die eine Distributionsableitung in $L_2(\lambda^k)$ besitzt, ist absolutstetig auf λ^k -fast allen achsenparallelen Linien. Der schwache Gradient stimmt mit dem gewöhnlichen Gradienten fast überall überein.

BEWEIS ZU PROPOSITION B.2.5: Maz'ya (1985), Theorem 1. ////

BEWEIS ZU THEOREM B.2.4: “ \Leftarrow ”

Wir zeigen den Fall $\Sigma = \mathbb{I}_k$.

$$\underline{\quad \quad \quad} \quad \left| \int \nabla \varphi dF \right|^2 = \left| \int \nabla \varphi f d\lambda \right|^2 = \left| \int -\nabla f \varphi d\lambda \right|^2 \leq$$

³genauer im Sinne einer schwachen / Sobolev-Ableitung; die Distributionsableitung ist nur über den durch sie induzierten linearen Operator erklärt; verlangt man noch nicht die Existenz eines (lokal) integralen φ' , so spricht man von einer schwachen Ableitung.

$$\begin{aligned}
&\leq \bar{\mathcal{I}}_f \int \varphi^2 dF \\
\| \int \nabla \varphi X^\tau dF \|^2 &= \| \int \nabla \varphi X^\tau f d\lambda \|^2 = \| \int -(\mathbb{I} + \frac{\nabla f}{f} X^\tau) f \varphi d\lambda \|^2 \leq \\
&\leq \bar{\mathcal{J}}_f \int \varphi^2 dF
\end{aligned}$$

und wegen der λ^k -Stetigkeit von F gilt $F(\{e_i = 0\}) = 0$ für alle i .

“ \Rightarrow ”

Weil der \mathbb{R}^k polnisch ist, existieren reguläre bedingte Verteilungen, und wir können $F(dx_1, \dots, dx_k)$ schreiben als

$$F(dx_1, \dots, dx_k) = \prod_{j=1}^{k-1} F_{j|j+1:k}(dx_j|x_{j+1}, \dots, x_k) F_k(dx_k) \quad (\text{B.11})$$

mit F_k der Randverteilung von X_k und $F_{j|j+1:k}$ einer regulären, bedingten Verteilung von X_j , gegeben $X_{j+1} = x_{j+1}, \dots, X_k = x_k$. Weiter schreiben wir im folgenden $y_{i:j}$ für den Vektor $(y_i, \dots, y_j)^\tau$.

Wir zeigen zunächst den multivariaten Lokationsfall und gehen dabei vor wie in Huber (1981), wobei wir allerdings die Argumentation etwas verkürzen können: Wegen $\bar{\mathcal{I}}_f < \infty$ gilt für ein $C \in (0, \infty)$ für $\partial_j \psi = \frac{\partial}{\partial x_j} \psi$, $\psi \in \mathcal{D}_k$ nach Voraussetzung für $j = 1, \dots, k$

$$| \int \partial_j \psi dF | \leq C \left(\int \psi^2 dF \right)^{\frac{1}{2}} = C \| \psi \|_{L_2(F)}.$$

Damit sind die linearen Operatoren $\psi \mapsto T_j(\psi) := \int \partial_j \psi dF$, definiert auf der dichten Teilmenge \mathcal{D}_k des $L_2(F)$ beschränkt, also stetig auf $L_2(F)$ fortsetzbar. Nach Riesz-Fréchet gibt es damit erzeugende Elemente $-g_j \in L_2(F)$, so daß

$$T_j(\psi) = - \int \psi g_j dF. \quad (\text{B.12})$$

Damit aber gilt für $\psi \in \mathcal{D}_k$

$$\begin{aligned}
T_1(\psi) &= \int \partial_1 \psi(y) F(dy) = - \int \psi g_1 dF = \\
&= \int_{\mathbb{R}^k} \int_{\mathbb{R}} \mathbb{I}_{\{x_1 \geq y_1\}} \partial_1 \psi(x_1, y_{2:k}) \lambda(dx_1) g_1(y) F(dy)
\end{aligned}$$

Analog zum univariaten Skalenfall ist $\mathbb{I}_{\{x_1 \geq y_1\}} \partial_1 \psi(x_1, y_{2:k}) g_1(y)$ $\psi \in \mathcal{D}_k$ in $L_1(\lambda_\psi(dx) \otimes F(dy))$, und so folgt mit Fubini für Markovkerne

$$f_1(x_1, y_{2:k}) := \int_{\mathbb{R}} \mathbb{I}_{\{x_1 \geq y_1\}} g_1(y) F_{1|2:k}(dy_1|y_{2:k}),$$

$$T_1(\psi) = \int f_1(x_1, y_{2:k}) \partial_1 \psi(x_1, y_{2:k}) \lambda(dx_1) F_{2:k}(dy_{2:k}).$$

Damit definieren $f_1(x_1, y_{2:k}) \lambda(dx_1) F_{2:k}(dy_{2:k})$ und $F_{1|2:k}(dx_1|y_{2:k}) F_{2:k}(dy_{2:k})$ auf der maßbestimmenden Menge \mathcal{D}_k das identische Funktional, also können wir $f_1(x_1, y_{2:k}) - F_{2:k}(dy_{2:k})$ -fast sicher — als Lebesgue-Dichte der regulären bedingten Verteilung $F_{1|2:k}(dx_1|y_{2:k})$ wählen.

Wir halten fest, daß wir an dieser Stelle im Gegensatz zum Beweis in Huber (1981) nicht nachzuweisen brauchen, daß f_1 punktweise beschränkt durch $F_{1|2:k}(dx_1|y_{2:k}) \int g_1^2 dF$ ist und für $x_1 \rightarrow \pm\infty$ gegen 0 geht; demzufolge können wir uns das Approximationsargument sparen, das Huber verwendet, um zu zeigen, daß $A1 = \int g dF = 0$, (4.5) in seinem Beweis.

Sei nun schon gezeigt:

$$F_{1:j|j+1:k}(dx_{1:j}|y_{j+1:k}) = f_{1:j}(x_{1:j}, y_{j+1:k}) \lambda^j(dx_{1:j}) \quad [F_{2:k}(dy_{2:k})]$$

Es gilt für $\psi \in \mathcal{D}_k$ nach Induktionsvoraussetzung

$$\begin{aligned} T_{j+1}(\psi) &= \int \partial_{j+1} \psi(y) F_{1:k}(dy_{1:k}) = \\ &\stackrel{\text{(B.12)}}{=} - \int g_{j+1}(y) \psi(y) F_{1:k}(dy_{1:k}) = \\ &\stackrel{\text{I.V.}}{=} - \int g_{j+1}(y) \psi(y) f_{1:j}(y) \lambda^j(dy_{1:j}) F_{j+1:k}(dy_{j+1:k}) = \\ &= \int_{\mathbb{R}^k} g_{j+1}(y) \int_{\mathbb{R}} \mathbf{I}_{\{x_{j+1} \geq y_{j+1}\}} \partial_{j+1} \psi(y_{1:j}, x_{j+1}, y_{j+2:k}) \lambda(dx_{j+1}) \times \\ &\quad \times f_{1:j}(y) \lambda^j(dy_{1:j}) F_{j+1|j+2:k}(dy_{j+1}|y_{j+2:k}) F_{j+2:k}(dy_{j+2:k}) \end{aligned}$$

Wieder gilt Fubini, und wir können setzen

$$f_{1:j+1}(y_{1:j}, x_{j+1}, y_{j+2:k}) = \int_{\mathbb{R}} g_{j+1}(y) \mathbf{I}_{\{x_{j+1} \geq y_{j+1}\}} f_{1:j}(y) F_{j+1|j+2:k}(dy_{j+1}|y_{j+2:k})$$

Damit gilt, wenn wir wieder die Integrationsvariable x_{j+1} durch y_{j+1} ersetzen,

$$T_{j+1}(\psi) = \int \partial_{j+1} \psi(y) f_{1:j+1}(y) \lambda^{j+1}(dy_{1:j+1}) F_{j+2:k}(dy_{j+2:k}).$$

Wie für $j = 1$ können wir also $f_{1:j+1}(y) - F_{2:k}(dy_{2:k})$ -fast sicher — als Lebesgue-Dichte der regulären bedingten Verteilung $F_{1:j+1|j+2:k}(dy_{1:j+1}|y_{j+2:k})$ wählen.

Damit erhalten wir, daß dF und $f_{1:k} d\lambda^k$ auf \mathcal{D}_k dasselbe Integral definieren, also ist $f_{1:k}$ tatsächlich als λ^k -Dichte von F wählbar. Insgesamt gilt nach Konstruktion λ^{k-1} -fast sicher:

$$\begin{aligned} f_{1:k}(x) &= \int \prod_{j=1}^k \mathbf{I}_{\{x_j \geq y_j\}} g_j(x_{1:j-1}, y_{j:k}) F_{j|j+1:k}(dy_j|y_{j+1:k}) = \\ &= \int \prod_{j=1}^k \mathbf{I}_{\{x_j \geq y_j\}} g_j(x_{1:j-1}, y_{j:k}) F(dy) = \\ &= \int \prod_{j=1}^k \mathbf{I}_{\{x_j \geq y_j\}} g_j(x_{1:j-1}, y_{j:k}) f_{1:k}(y) \lambda^k(dy), \end{aligned}$$

also ist auch $f_{1:k}$ λ^k -stetig mit verallgemeinertem⁴ Gradienten

$$\nabla f_{1:k} = (\partial_j f_{1:k})_j := (f_{1:k} g_j)_j, \quad (\text{B.13})$$

denn es gilt für $\psi \in \mathcal{D}_k$, indem wir diese Setzung für $\partial_j f_{1:k}$ verwenden:

$$\int \psi \partial_j f_{1:k} d\lambda^k = \int \psi f_{1:k} g_j d\lambda^k = \int \psi g_j dF = - \int \partial_j \psi dF = - \int \partial_j \psi f_{1:k} d\lambda^k.$$

Für die Skalen–Fisherinformation gehen wir analog vor:

Wie im Lokationsfall zeigen wir induktiv, daß aus der Stetigkeit der auf L_2 fortgesetzten linearen Operatoren $\psi \mapsto S_j(\psi) := \int x_j \partial_j \psi dF$, mit Riesz–Fréchet und erzeugenden Elementen $\tilde{g}_j \in L_2(F)$, so daß $S_j(\psi) = - \int \tilde{g}_j \psi dF$, folgt

$$F(dx) = f_{1:k}(x) \lambda^k(dx) \quad (\text{B.14})$$

mit

$$f_{1:k}(x) = \int \prod_{j=1}^k \mathbb{I}_{\{x_j \geq y_j\}} \frac{1}{x_j} \tilde{g}_j(x_{1:j-1}, y_{j:k}) f_{1:k}(y) \lambda^k(dy). \quad (\text{B.15})$$

Dabei benützen wir, daß $f_{1:i}(x_{1:i-1}, x_i, y_{2:k}) \lambda^i(dx_{1:i}) F_{i+1:k}(dy_{i+1:k})$ und $F_{1:i|i+1:k}(dx_{1:i}|y_{i+1:k}) F_{i+1:k}(dy_{i+1:k})$ dasselbe Funktional auf der Menge

$$\mathcal{D}_k(i)^0 := \{\varphi : \varphi = x_i \psi \text{ für ein } \psi \in \mathcal{D}_k\}$$

definieren. Da aber F für kein j auf Ereignisse der Form $\{x_j = 0\}$ Masse wirft, zeigt man analog zu Lemma B.2.3, daß $\mathcal{D}_k^0 = \cap_i \{\mathcal{D}_k^0(i)\}$ maßbestimmend ist und demnach (14) gilt.

Nach (15) ist auch $f_{1:k}$ λ^k –stetig mit verallgemeinertem⁴ Gradienten

$$\nabla f_{1:k} = (\partial_j f_{1:k})_j = f_{1:k} \left[\frac{1}{x_j} (\tilde{g}_j - 1) \right]_j,$$

denn es gilt für $\psi \in \mathcal{D}_k$

$$\begin{aligned} \int \psi x_j \partial_j f_{1:k} d\lambda^k &= \int \psi f_{1:k} [\tilde{g}_j - 1] d\lambda^k = \\ &= - \int [\psi + x_j \partial_j \psi] f_{1:k} d\lambda^k = - \int \partial_j [\psi x_j] f_{1:k} d\lambda^k. \end{aligned}$$

Damit aber ist auch für jedes i das Produkt $x_i f_{1:k}$ λ^k –stetig mit Distributionsableitung $f_{1:k} \mathbb{I}_{\{i=j\}} + \frac{x_i}{x_j} (\tilde{g}_j - 1)$, denn nach der Produktregel für Distributionsableitungen, bei denen ein Faktor in $\mathcal{C}^\infty(\mathbb{R}^k, \mathbb{R})$ liegt, nämlich

$$\partial_j(pq) = q \partial_j p + p \partial_j q, \quad q \in L_{2,\text{loc}}(\lambda^k), \partial_j q \in L_{1,\text{loc}}^k(\lambda^k), p \in \mathcal{C}^\infty(\mathbb{R}^k, \mathbb{R})$$

folgt mit $p = x_j$, $q = f_{1:k}$, daß

$$\partial_j(x_i f_{1:k}(x)) = f_{1:k} \mathbb{I}_{\{i=j\}} + \frac{x_i}{x_j} (\tilde{g}_j - 1). \quad (\text{B.16})$$

Damit existieren jeweils Distributionsableitungen in $L_2(F)$, also ist Proposition B.2.5 in Kraft, und die entsprechenden Dichten sind längs achsenparallelen Linien absolutstetig.

⁴im Sinne einer schwachen / Distributions–Ableitung

Die Ausdrücke (9) und (10) sieht man so: Mit (13) und der Definition der g_j folgt

$$\begin{aligned} |\int \nabla \varphi dF|^2 &= \sum_j |\int \partial_j \varphi dF|^2 = \sum_j |\int g_j \varphi dF|^2 \leq \\ &\leq \sum_j \int g_j^2 dF \int \varphi^2 dF = \int (\frac{|\nabla f|}{f})^2 dF \int \varphi^2 dF. \end{aligned}$$

Für die Skalenterme definieren wir für $0 \neq j$ neben den ‘‘Diagonaltermen’’ S_j weitere $L_2(F)$ -stetige Abbildungen $\psi \mapsto S_{i,j}(\psi) := \int x_i \partial_j \psi dF$ mit erzeugenden Elementen $\tilde{g}_{i,j} \in L_2(F)$, so daß $S_{i,j}(\psi) = - \int \tilde{g}_{i,j} \psi dF$. Insbesondere gilt dann

$$S_{i,j}(\psi) := \int x_i \partial_j \psi f_{1:k} d\lambda^k = - \int \partial_j [x_i f_{1:k}] \psi d\lambda^k = - \int \tilde{g}_{i,j} f_{1:k} \psi d\lambda^k, \tag{B.17}$$

so daß F -fast-sicher $\tilde{g}_{i,j} = \partial_j [x_i f_{1:k}] / f_{1:k}$. Damit ergibt sich dann, indem wir $\tilde{g}_{j,j} = \tilde{g}_j$ setzen

$$\begin{aligned} |\int \nabla \varphi x^\tau dF|^2 &= \sum_{i,j} |\int x_i \partial_j \varphi dF|^2 = \sum_j |\int g_{i,j} \varphi dF|^2 \leq \\ &\leq \sum_j \int g_{i,j}^2 dF \int \varphi^2 dF = \int [\mathbb{I}_k + \frac{\nabla f}{f} X^\tau]^2 dF \int \varphi^2 dF \end{aligned}$$

////

Nach diesen Vorbereitungen können wir nun die Begriffe der multivariaten Lokations-/Skalen-Fisher-Information unabhängig von der Existenz von Dichten im Zentral-Modell einführen:

Definition B.2.6 (a) *Ist im multivariaten Lokationsmodell,*

$$\mathcal{P} := \{F(\cdot - \theta) \mid \theta \in \mathbb{R}^k\},$$

die Größe $\bar{\mathcal{I}}_f$ aus (6) endlich, so nennen wir die dann existierende Matrix in $\mathbb{R}^{k \times k}$

$$\mathcal{I}_f := \int \frac{\nabla f \nabla f^\tau}{f} d\lambda^k \tag{B.18}$$

(multivariate) Lokations-Fisher-Information.

(b) *Ist im multivariaten Skalenmodell,*

$$\mathcal{P} := \{F(\Sigma^{-1}(\cdot)) \mid \Sigma = \Sigma^\tau \succ 0 \in \mathbb{R}^{k \times k}\},$$

die Größe $\bar{\mathcal{J}}_f$ aus (7) endlich und gilt (8), so nennen wir die dann existierende Matrix $\mathbb{R}^{\binom{k}{2} \times \binom{k}{2}}$

$$\mathcal{J}_f(\mathbb{I}_k) := \int (\text{svec}[\mathbb{I}_k + \frac{\nabla f}{f} X^\tau]) (\text{svec}[\mathbb{I}_k + \frac{\nabla f}{f} X^\tau])^\tau dF \tag{B.19}$$

mit dem in (B.1) eingeführten Operator $\text{svec}(\cdot)$ (multivariate) Skalen-Fisher-Information.

(c) *Sind im multivariaten Lokations-/Skalenmodell (3) die Größen $\bar{\mathcal{I}}_f$ aus (6) und $\bar{\mathcal{J}}_f$ aus (7) endlich und gilt (8), so nennen wir die dann existierende Matrix $\mathcal{I}_F(\mathbb{I}_k) \in \mathbb{R}^{(k+\binom{k}{2}) \times (k+\binom{k}{2})}$,*

$$\mathcal{I}_F(\mathbb{I}_k) := \begin{pmatrix} \mathcal{I}_f & \mathcal{H}_f^\tau(\mathbb{I}_k) \\ \mathcal{H}_f(\mathbb{I}_k) & \mathcal{J}_f(\mathbb{I}_k) \end{pmatrix} \tag{B.20}$$

mit \mathcal{I}_f aus (18), $\mathcal{J}_f(\mathbb{I}_k)$ aus (19) und

$$\mathcal{H}_f(\mathbb{I}_k) := - \int (\text{svec}[\mathbb{I}_k + \frac{\nabla f}{f} X^\tau]) \nabla f^\tau d\lambda^k \quad (\text{B.21})$$

(multivariate) Lokations-/Skalen-Fisher-Information.

(d) Ist $\bar{\mathcal{J}}_f$ aus (7) endlich und gilt (8) [und ist $\bar{\mathcal{I}}_f$ aus (6) ebenfalls endlich], so ergibt sich die (multivariate) [Lokations-/]Skalen-Fisher-Information an der Stelle $[\theta], \Sigma, \Sigma \succ 0$ als $\mathcal{J}_f(\Sigma)$ bzw. als

$$\mathcal{I}_F(\Sigma) := \begin{pmatrix} \Sigma^{-1} \mathcal{J}_f \Sigma^{-1} & \mathcal{H}_f^\tau(\Sigma) \\ \mathcal{H}_f(\Sigma) & \mathcal{J}_f(\Sigma) \end{pmatrix} \quad (\text{B.22})$$

mit

$$\mathcal{H}_f(\Sigma) = \int (\text{svec}[\Sigma^{-1}(\mathbb{I}_k + \frac{\nabla f}{f} X^\tau)]) \nabla f^\tau \Sigma^{-1} d\lambda^k \quad (\text{B.23})$$

$$= \text{diag}[\text{svec}(\Sigma^{-1})] \mathcal{H}_f(\mathbb{I}_k) \Sigma^{-1} \quad (\text{B.24})$$

$$\mathcal{J}_f(\Sigma) = \int (\text{svec}[\Sigma^{-1}(\mathbb{I}_k + \frac{\nabla f}{f} X^\tau)]) (\text{svec}[\Sigma^{-1}(\mathbb{I}_k + \frac{\nabla f}{f} X^\tau)])^\tau dF \quad (\text{B.25})$$

$$= \text{diag}[\text{svec}(\Sigma^{-1})] \mathcal{J}_f(\mathbb{I}_k) \text{diag}[\text{svec}(\Sigma^{-1})] \quad (\text{B.26})$$

Im Vorgriff auf Proposition B.3.1 und die Lemmata B.3.2 und B.3.3 notieren wir die L_2 -Ableitungs-Kandidaten

$$\begin{aligned} \Lambda_{1,\theta,\Sigma}^{(f)}(x) &= \Sigma^{-1} \Lambda_{1,0,\mathbb{I}_k}^{(f)}(\Sigma^{-1}(x - \theta)) \\ \Lambda_{1,0,\mathbb{I}_k}^{(f)}(x) &= -\nabla f(x)/f(x) \\ \Lambda_{2,\theta,\Sigma}^{(f)}(x) &= \Sigma^{-1} \Lambda_{2,0,\mathbb{I}_k}^{(f)}(\Sigma^{-1}(x - \theta)) \\ \Lambda_{2,0,\mathbb{I}_k}^{(f)}(x) &= -(\mathbb{I}_k + \Lambda_{1,0,\mathbb{I}_k}^{(f)}(x)x^\tau) \end{aligned}$$

B.3 Glattheit des multivariaten Lokations-/Skalen-Modells

Im eindimensionalen Lokationsmodell charakterisiert die Endlichkeit der Fisher-Information aus (4) die L_2 -Differenzierbarkeit des Modells, wie folgende Proposition zeigt, die auf einem Lemma von Hájek basiert.

Proposition B.3.1 1. Es gelte $\mathcal{I}(F) < \infty$. Dann gilt für $s \rightarrow 0$

$$\int \left| \sqrt{f}(u-s) - \sqrt{f}(u) \left(1 + \frac{1}{2} \Lambda_f(u)\right) \right|^2 \lambda(du) = o(s^2) \quad (\text{B.27})$$

2. Es gelte $\mathcal{J}(F) < \infty$. Dann gilt für $s \rightarrow 0$

$$\int \left| \sqrt{1+s} \sqrt{f}((1+s)u) - \sqrt{f}(u) \left(1 + \frac{1}{2} [1-u\Lambda_f(u)]\right) \right|^2 \lambda(du) = o(s^2) \quad (\text{B.28})$$

BEWEIS ZU PROPOSITION B.3.1: (i) Rieder (1994) Proposition 2.4.1, (i) und (ii) Swensen (1980) Ch. 2, Sec. 3. ////

Dies wollen wir nun auf das multivariate Lokations-/Skalenmodell (3) verallgemeinern, indem wir die in Abschnitt B.2 definierte multivariate Fisher-Information dieses Modells verwenden.

Der Schluß auf den multivariaten Fall erfolgt, indem wir zunächst aus Proposition B.3.1 auf die "partielle" L_2 -Differenzierbarkeit schließen. Diese ist wegen der Translations-/Streckungs-Äquivarianz des Lebesguemaßes sogar automatisch stetig, wie wir zeigen werden, und damit folgt wie im endlich-dimensionalen aus der stetigen "partiellen" L_2 -Differenzierbarkeit die gewünschte totale L_2 -Differenzierbarkeit:

Lemma B.3.2 *Ein multivariates Lokationsmodell $\mathcal{P} = \{F(\cdot - \theta)\} \subset \mathcal{M}_1(\mathbb{B}^k)$ ist genau dann L_2 -differenzierbar, wenn es dies auch "partiell" in jeder Koordinate ist, d.h. für jede Koordinate j gilt*

$$\int \left(\sqrt{f}(x_1, \dots, x_j + h, \dots, x_k) - \sqrt{f}(x) \left(1 - \frac{1}{2} \Lambda_{f,j}(x)\right) \right)^2 \lambda^k(dx) = o(h)$$

mit $\Lambda_f(x) = -\nabla f/f$.

BEWEIS ZU LEMMA B.3.2: Garel und Hallin (1995), Lemma 2.1; da aber Teile des Beweises noch im nächsten Lemma verwendet können, hier die Schritte:

Wie in der herkömmlichen mehrdimensionalen Differenzierbarkeit folgt aus der stetigen, partiellen L_2 -Differenzierbarkeit die (totale) L_2 -Differenzierbarkeit. Im multivariaten Lokationsmodell sind aber die partiellen L_2 -Ableitungen automatisch L_2 -stetig: Zunächst bemerken wir, daß wir uns wegen der Translationsäquivarianz des Lebesguemaßes auf die Stelle $\theta = 0$ beschränken können, denn

$$\|\Lambda_{\theta+h} \sqrt{dP_{\theta+h}} - \Lambda_{\theta} \sqrt{dP_{\theta}}\| = \|\Lambda_h \sqrt{dP_h} - \Lambda_0 \sqrt{dP_0}\|.$$

Zur Überprüfung der L_2 -Stetigkeit führen wir zu $h = (h_1, \dots, h_k)^\tau \in \mathbb{R}^k$ die Zwischenstellen $h^0 = 0$, $h^j := (h_1, \dots, h_j, 0, \dots, 0)^\tau$ ein. Dann ist

$$\begin{aligned} & \|\Lambda_h \sqrt{dP_h} - \Lambda_0 \sqrt{dP_0}\| \leq \sum_{j=1}^k \|\Lambda_{h^j} \sqrt{dP_{h^j}} - \Lambda_{h^{j-1}} \sqrt{dP_{h^{j-1}}}\| = \\ &= \sum_{\substack{1 \leq j \leq k \\ h_j \neq 0}} \frac{2}{|h_j|} \left\| \left\{ \sqrt{dP_{h^{j-1}}} - \sqrt{dP_{h^j}} \left(1 - \frac{1}{2} h_j \Lambda_{h^j}\right) \right\} + \right. \\ & \quad \left. + \left\{ \sqrt{dP_{h^j}} - \sqrt{dP_{h^{j-1}}} \left(1 + \frac{1}{2} h_j \Lambda_{h^{j-1}}\right) \right\} \right\| \leq \\ & \leq \sum_{\substack{1 \leq j \leq k \\ h_j \neq 0}} \frac{2}{|h_j|} \left(\int \left\{ \sqrt{p_0}(x + h_j e_j) - \sqrt{p_0}(x) \left(1 - \frac{1}{2} h_j \Lambda_0(x)\right) \right\}^2 d\lambda^k \right)^{1/2} \\ & \quad + \left(\int \left\{ \sqrt{p_0}(x - h_j e_j) - \sqrt{p_0}(x) \left(1 + \frac{1}{2} h_j \Lambda_0(x)\right) \right\}^2 d\lambda^k \right)^{1/2} = \\ &= \sum_{\substack{1 \leq j \leq k \\ h_j \neq 0}} \frac{4}{|h_j|} o(h_j) = o(h^0), \end{aligned}$$

wobei p_0 die entsprechende k -dimensionale Lebesgue-Dichte von P_0 ist. ////

Lemma B.3.3 *Ein multivariates Skalenmodell $\mathcal{P} = \{F(\Sigma^{-1}\cdot) \mid \Sigma = \Sigma^\tau \succ 0\} \subset \mathcal{M}_1(\mathbb{B}^k)$ ist genau dann L_2 -differenzierbar, wenn es dies auch in jeder Koordinate ist, d.h. für jede Koordinate i, j und jedes $A \in \mathbb{R}^{p \times p}$ gilt*

$$\int \left(\sqrt{\det(\mathbb{I}_k + h\delta_{i,j}A)} \sqrt{f((\mathbb{I}_k + h\delta_{i,j}A)x)} - \sqrt{f(x)} \left(1 + \frac{1}{2}\Lambda_{\mathbb{I}_k}(x)\right) \right)^2 \lambda^k(dx) = o(h)$$

mit $\delta_{i,j} \in \mathbb{R}^{k \times k}$, $\delta_{i,j}$ die Elementarmatrix mit einer 1 in den Koordinaten i, j und sonst nur 0-Einträgen.

BEWEIS ZU LEMMA B.3.3: analog zu Lemma B.3.2; wir zeigen, daß im multivariaten Skalenmodell die partiellen L_2 -Ableitungen automatisch L_2 -stetig sind. Wir bemerken, daß wir uns wegen der Streckungsäquivarianz des Lebesguemaßes auf die Stelle $\Sigma = \mathbb{I}_k$ beschränken können, denn

$$\begin{aligned} & \|\Lambda_{\Sigma(\mathbb{I}_k+H)} \sqrt{dP_{\Sigma(\mathbb{I}_k+H)}} - \Lambda_{\Sigma} \sqrt{dP_{\Sigma}}\| = \|\Sigma^{-1} [\Lambda_{\mathbb{I}_k+H} \sqrt{dP_{\mathbb{I}_k+H}} - \Lambda_{\mathbb{I}_k} \sqrt{dP_{\mathbb{I}_k}}]\| \\ & \leq \|\Lambda_{\mathbb{I}_k+H} \sqrt{dP_{\mathbb{I}_k+H}} - \Lambda_{\mathbb{I}_k} \sqrt{dP_{\mathbb{I}_k}}\| / \text{minev}(\Sigma). \end{aligned}$$

Wie im Beweis zu Lemma B.3.2 zerlegen wir das Inkrement H in Zwischenstellen $H^{(j)}$, die sich jeweils nur in einer Koordinate unterscheiden. Dazu führen wir die Indizierung gemäß des in (B.1) eingeführten $\text{svec}(\cdot)$ -Operators durch, und definieren zu $H = H^\tau \in \mathbb{R}^{k \times k}$ die Zwischenstellen $H^{(j)}$ durch

$$\mathbb{I}_k + H = \mathbb{I}_k + H^{\binom{k}{2}} = \mathbb{I}_k + H^{(j)} + \sum_{i=j+1}^{\binom{k}{2}} h_i \Delta_i;$$

dabei ist $\Delta_i = \text{svec}^{-1}(\delta_i)$ und δ_i der i -te $\binom{k}{2}$ -dimensionale Einheitsvektor. Schließlich verwenden wir die beiden speziellen Inkremente

$$\begin{aligned} H_{j,j-1} & := h_j \Delta_j (\mathbb{I}_k + H^{(j-1)})^{-1} \\ H_{j,j} & := h_j \Delta_j (\mathbb{I}_k + H^{(j)})^{-1}. \end{aligned}$$

Damit erhalten wir

$$\begin{aligned} & \|\Lambda_{\mathbb{I}_k+H} \sqrt{dP_{\mathbb{I}_k+H}} - \Lambda_{\mathbb{I}_k} \sqrt{dP_{\mathbb{I}_k}}\| \leq \\ & \leq \sum_{\substack{1 \leq j \leq \binom{k}{2} \\ h_j \neq 0}} \frac{2}{|h_j|} \left\| \left\{ \sqrt{dP_{\mathbb{I}_k+H^{(j-1)}}} - \sqrt{dP_{\mathbb{I}_k+H^{(j)}}} \left(1 - \frac{1}{2} h_j \Delta_j \Lambda_{\mathbb{I}_k+H^{(j)}}\right) \right\} + \right. \\ & \quad \left. + \left\{ \sqrt{dP_{\mathbb{I}_k+H^{(j)}}} - \sqrt{dP_{\mathbb{I}_k+H^{(j-1)}}} \left(1 + \frac{1}{2} h_j \Delta_j \Lambda_{\mathbb{I}_k+H^{(j-1)}}\right) \right\} \right\| \leq \\ & \leq \sum_{\substack{1 \leq j \leq \binom{k}{2} \\ h_j \neq 0}} \frac{2}{|h_j|} \left(\int \left\{ \frac{\sqrt{\det(\mathbb{I}_k+H^{(j)})}}{\sqrt{\det(\mathbb{I}_k+H^{(j-1)})}} \sqrt{p_{\mathbb{I}_k}}((\mathbb{I}_k + H^{(j)})(\mathbb{I}_k + H^{(j-1)})^{-1}x) - \right. \right. \\ & \quad \left. \left. - \sqrt{p_{\mathbb{I}_k}}(x) \left(1 - \frac{1}{2} h_j \Delta_j (\mathbb{I}_k + H^{(j)})^{-1} \Lambda_{\mathbb{I}_k}(x)\right) \right\}^2 d\lambda^k \right)^{1/2} \\ & \quad + \frac{2}{|h_j|} \left(\int \left\{ \frac{\sqrt{\det(\mathbb{I}_k+H^{(j-1)})}}{\sqrt{\det(\mathbb{I}_k+H^{(j)})}} \sqrt{p_{\mathbb{I}_k}}((\mathbb{I}_k + H^{(j-1)})(\mathbb{I}_k + H^{(j)})^{-1}x) - \right. \right. \\ & \quad \left. \left. - \sqrt{p_{\mathbb{I}_k}}(x) \left(1 + \frac{1}{2} h_j \Delta_j (\mathbb{I}_k + H^{(j-1)})^{-1} \Lambda_{\mathbb{I}_k}(x)\right) \right\}^2 d\lambda^k \right)^{1/2} = \\ & = \sum_{\substack{1 \leq j \leq \binom{k}{2} \\ h_j \neq 0}} \frac{2}{|h_j|} \left(\int \left\{ \sqrt{\det[\mathbb{I}_k + H_{j,j-1}]} \sqrt{p_{\mathbb{I}_k}}([\mathbb{I}_k + H_{j,j-1}]x) - \right. \right. \end{aligned}$$

$$\begin{aligned}
 & -\sqrt{p_{\mathbb{I}_k}}(x)(1 - \frac{1}{2}H_{j,j}\Lambda_{\mathbb{I}_k}(x))\}^2 d\lambda^k)^{1/2} \\
 & + \frac{2}{|h_j|} \left(\int \left\{ \sqrt{\det[\mathbb{I}_k - H_{j,j}]} \sqrt{p_{\mathbb{I}_k}}([\mathbb{I}_k - H_{j,j}]x) - \right. \right. \\
 & \left. \left. -\sqrt{p_{\mathbb{I}_k}}(x)(1 + \frac{1}{2}H_{j,j}\Lambda_{\mathbb{I}_k}(x))\}^2 d\lambda^k \right)^{1/2} = \sum_{\substack{1 \leq j \leq \binom{k}{2} \\ h_j \neq 0}} \frac{4}{|h_j|} o(h_j) = o(H^0).
 \end{aligned}$$

////

Als Konsequenz notieren wir

Theorem B.3.4 *Das multivariate Lokations-/Skalenmodell (3) ist genau dann L_2 -differenzierbar, wenn die Größen $\bar{\mathcal{I}}_f, \bar{\mathcal{J}}_f$ aus (9) und (10) endlich sind und (8) gilt.*

B.4 generalisierte Fisher-Information für zufällige Parameter

In diesem Abschnitt wollen wir die Techniken aus dem multivariaten Lokationsfall übertragen auf den Fall, daß der "Lokations-Parameter" stochastisch ist. Seien dazu $X, Y \mathbb{R}^p$ - bzw. \mathbb{R}^q -dimensionale Zufallsvariablen, deren gemeinsame Verteilung $P^{X,Y}$ durch das Produktmaß $\mu = \mu_1 \otimes \mu_2$ mit gewissen $\mu_{1[2]} \in \mathcal{M}_\sigma(\mathbb{B}^{p[q]})$ dominiert wird. Dabei sei $E|X|^2$ endlich. Dann definieren wir

$$\bar{\mathcal{I}}_{X|Y}(P^{X,Y}) := \sup \frac{(\int \nabla_x \varphi dP^{X,Y})^\tau (\int \nabla_x \varphi dP^{X,Y})}{\int \varphi^2 dP^{X,Y}}, \quad 0 \neq \varphi \in \mathcal{D}_{p+q}, \quad [P^{X,Y}]$$

(B.29)

wobei $\mathcal{D}_k := C_c^\infty(\mathbb{R}^k, \mathbb{R})$ wie in (6), (7) die Menge aller beliebig oft stetig differenzierbaren Funktionen von \mathbb{R}^k nach \mathbb{R} mit kompaktem Träger ist und $\nabla_x = (\frac{\partial}{\partial x_i})_{i=1, \dots, p}$. Mit den gleichen Techniken wie in Theorem B.2.4 kann man zeigen

Theorem B.4.1 *Die Endlichkeit von $\bar{\mathcal{I}}_{X|Y}(P^{X,Y})$ ist äquivalent zur Existenz einer Lebesguedichte $p^{X|Y=y}$ von $P^{X|Y=y}$, wobei $p^{X|Y=y} \lambda^p$ -stetig ist. Weiter ist $p^{X|Y=y}$ für fast alle achsenparallelen Linien $g(u) = (x_{1:j-1}, u, x_{j+1:p}, y_{1:q})$ absolutstetig. Ist eine der beiden äquivalenten Bedingungen erfüllt, so ist*

$$\bar{\mathcal{I}}_{X|Y}(P^{X,Y}) = \int |\nabla_x p^{X|Y} / p^{X|Y}|^2 d\lambda^p dP^Y \tag{B.30}$$

Entsprechend nennen wir

$$\mathcal{I}_{X|Y}(P^{X,Y}) = \int (\frac{\nabla_x p^{X|Y=y}}{p^{X|Y}})^\tau \frac{\nabla_x p^{X|Y=y}}{p^{X|Y}} d\lambda^p dP^Y \tag{B.31}$$

generalisierte Fisher-Information von X gegeben Y .

B.5 Kovarianzen von Kovarianzen

Abschließend präsentieren wir ein Resultat über den Rang der Kovarianz der empirischen Kovarianz, das wir unter anderem benötigen, um Bedingungen für (10.74) und (10.75) in Abschnitt 10.4.2 angeben zu können.

Proposition B.5.1 *Sei X eine \mathbb{R}^q -wertige Zufallsgröße, $E|X|^4 < \infty$; betrachten wir $\tilde{X} := (1, X^\tau)^\tau$, sowie $X_0 := X - E[X]$ und $\tilde{X}_0 := (1, X_0^\tau)^\tau$ dann sind äquivalent*

- (a) $\text{rk Cov}\{\text{vec}(X_0 X_0^\tau)[\dots]^\tau\} < \binom{q}{2}$
- (b) $\det \text{Cov}\{\text{svec}(X_0 X_0^\tau)[\dots]^\tau\} = 0$
- (c) *Es gibt eine Matrix $0 \neq A = A^\tau \in \mathbb{R}^{(q+1) \times (q+1)}$ $P(\tilde{X}^\tau A \tilde{X} = 0) = 1$.*
- (d) *X liegt fast sicher auf einer Quadrik, also*

$$X^\tau A X + b^\tau X + c = 0$$

für gewisse $A = A^\tau \in \mathbb{R}^q$, $b \in \mathbb{R}^q$, $c \in \mathbb{R}$, wobei zusätzlich $b = -A E[X]$.

Gilt

$$E[X_{0,i} X_{0,j} X_{0,k}] = 0 \quad \forall 1 \leq i, j, k \leq q, \quad (\text{B.32})$$

so folgt aus der Tatsache, daß X fast sicher auf einer Quadrik liegt, bereits, daß X auf einer Quadrik mit $b = -A E[X]$ liegt. Dies ist insbesondere der Fall, wenn $\mathcal{L}(X) = \mathcal{L}(-X)$.

BEWEIS ZU PROPOSITION B.5.1: Zunächst bemerken wir, daß wegen der Symmetrie von $X_0 X_0^\tau$ die Matrix $C := \text{Cov}\{\text{vec}(X_0 X_0^\tau)[\dots]^\tau\}$ maximal Rang $(q^2 + q)/2$ hat. Betrachten wir aber die Untermatrix

$$C_S := \text{Cov}\{\text{svec}(X_0 X_0^\tau)[\dots]^\tau\} = (C_{i_1, i_2, j_1, j_2})_{1 \leq i_1, j_1 \leq i_2, j_2 \leq q}$$

und deren Konjugation mit beliebigen Linearformen A auf diesem Unterraum, so sehen wir

$$A^\tau C_S A = \sum_{\substack{1 \leq i_1, j_1 \\ \leq i_2, j_2 \leq q}} A_{i_1, i_2} C_S A_{j_1, j_2} = \sum_{\substack{1 \leq i_1, j_1 \leq \\ \leq i_2, j_2 \leq q}} \tilde{A}_{i_1, i_2} C \tilde{A}_{j_1, j_2}$$

mit

$$\tilde{A}_{i_1, i_2} = \begin{cases} A_{i_1, i_2}/2 & i_1 < i_2 \\ A_{i_1, i_2} & \text{für } i_1 = i_2 \\ A_{i_2, i_1}/2 & i_1 > i_2 \end{cases}$$

Damit läßt sich also jede solche Linearform als symmetrische Matrix in der vollen Raumdimension interpretieren und umgekehrt. Dies zeigt (a) \iff (b). Gilt (a) oder (b), so liegt $Y := \text{svec}(X_0 X_0^\tau) - E[\text{svec}(X_0 X_0^\tau)]$ fast sicher auf einer Hyperebene $A_s \in \mathbb{R}^{\binom{q}{2}}$. Mit obiger Überlegung gilt aber dann mit $B = \text{svec}^{-1}(A_s)$

$$\begin{aligned} 0 &\stackrel{[P]}{=} A_s^\tau Y = \text{tr } A_s^\tau Y = \text{tr } B \{X_0 X_0^\tau - E[X_0 X_0^\tau]\} = X_0^\tau B X_0 - E[X_0^\tau B X_0] = \\ &= X^\tau B X - 2E[X]^\tau B X + 2E[X]^\tau B E[X] - E[X^\tau B X] = \tilde{X}^\tau A \tilde{X} \end{aligned}$$

mit

$$A = \begin{pmatrix} 2E[X]^\tau B E[X] - E[X^\tau B X] & -E[X] B \\ (-E[X] B)^\tau & B \end{pmatrix}$$

Damit ist (c) erfüllt — und somit gleich (d). Gilt umgekehrt (d), so auch (c) mit

$$A = \begin{pmatrix} c & -B E[X] \\ (-B E[X])^\tau & B \end{pmatrix}.$$

Aus $X^\tau B X - 2 E[X]^\tau B X + c \stackrel{[P]}{=} 0$ aber folgt $X_0^\tau B X_0 - E[X]^\tau B E[X] + c \stackrel{[P]}{=} 0$. Gehen wir zu Erwartungswerten über, so ist damit $c = E[X]^\tau B E[X] - E[X_0^\tau B X_0]$, und daher liegt Y auf der durch $\text{svec}(B)$ erzeugten Hyperebene und es gelten (a) und (b).

Sei $\text{Cov}[X] = U S U^\tau$ mit $U U^\tau = U^\tau U = \mathbb{I}_q$ und $S = \text{diag}[s_1, \dots, s_q]$, wobei ohne Einschränkung $s_1, \dots, s_l > 0$, $s_{l+1}, \dots, s_q = 0$ gelte. Liegt nun X auf einer Quadrik und gilt (32), so haben wir für $db = 2(b + A E[X])$,

$$\begin{aligned} X^\tau A X + 2b^\tau X + c &\stackrel{[P]}{=} 0 \iff X_0^\tau A X_0 + db^\tau X_0 - E[X_0^\tau A X_0] \stackrel{[P]}{=} 0 \\ \iff \check{X}_0^\tau \check{A} \check{X}_0 + d\check{b}^\tau \check{X}_0 &\stackrel{[P]}{=} E[\check{X}_0^\tau \check{A} \check{X}_0] \\ \implies E[(\check{X}_0^\tau \check{A} \check{X}_0 + d\check{b}^\tau \check{X}_0)^2] &= (E[\check{X}_0^\tau \check{A} \check{X}_0])^2 \\ \iff E[(\check{X}_0^\tau \check{A} \check{X}_0)^2] + 2E[(\check{X}_0^\tau \check{A} \check{X}_0)(d\check{b}^\tau \check{X}_0)] &+ E[(d\check{b}^\tau \check{X}_0)^2] = (E[\check{X}_0^\tau \check{A} \check{X}_0])^2 \\ \iff 0 \leq \text{Cov}[\check{X}_0^\tau \check{A} \check{X}_0] = -E[(d\check{b}^\tau \check{X}_0)^2] &= -\sum_{r=1}^q s_r (d\check{b}_r)^2 \leq 0, \end{aligned}$$

wobei wir $\check{A} = U^\tau A U$, $\check{X}_0 := U^\tau X_0$, $d\check{b} := U^\tau db$ setzen.

Daher ist also $d\check{b}_r = 0$ für $r = 1, \dots, l$.

Aus $s_r = 0$, $r = l + 1, \dots, q$ folgt aber $\check{X}_{0,r} \stackrel{[P]}{=} 0$, so daß die Aussagen für beliebige Werte $d\check{b}_r$, $r = l + 1, \dots, q$ gelten, also insbesondere für $d\check{b}_r = 0$. Damit liegt X auch auf der Quadrik $(A, b' = -A E[X], c)$, A , c wie gehabt, so daß das ursprüngliche b und b' dieselbe Quadrik beschreiben. ////

Korollar B.5.2 Sei $X \sim \mathcal{N}_q(0, V)$ mit $V \succ 0$. Dann gilt

$$\det \text{Cov}\{[\text{svec}(X X^\tau)][\dots]^\tau\} > 0,$$

also $\text{rk}(\mathbb{I}_{q^2} + K_{q,q})V^{\otimes 2} = \binom{q}{2}$.

BEWEIS ZU KOROLLAR B.5.2: Da $\kappa^{(X)} = 0$ ist

$$\text{Cov}\{[\text{svec}(X X^\tau)][\dots]^\tau\} = (\mathbb{I}_{q^2} + K_{q,q})V^{\otimes 2}.$$

Sei $A = A^\tau \neq 0$. Dann ist $\text{svec}(A)$ eine generische Linearform auf $\mathbb{R}^{\binom{q}{2}}$ und

$$[\text{svec}(X X^\tau)]^\tau [\text{svec}(A)] = X^\tau A X$$

Aber $P(X^\tau A X = 0) \neq 1$, da X wegen $V \succ 0$ nicht fast sicher auf einer (möglicherweise entarteten) Quadrik liegt. ////

Damit gelten (74) und (75), wann immer v und ε nicht auf einer (möglicherweise entarteten⁵) Quadrik liegen.

⁵zum Beispiel einer affinen Hyperfläche

Bemerkung B.5.3 Daß die Eigenschaft $\text{Cov } X \succ 0$ nur im Fall $\kappa^{(X)} = 0$ für den vollen Rang der Kovarianz der empirischen Kovarianz hinreicht, zeigen folgende Beispiele:

1. $X = (Y, UY)^\tau$, $Y \sim \mathcal{N}(0, 1)$ $U \sim \pm 1$ jeweils mit Wahrscheinlichkeit $1/2$, U, Y stochastisch unabhängig. Dann ist $E[X] = 0$, $\text{Cov}(X) = \mathbb{I}_2$, $\text{svec}(XX^\tau) = (Y^2, UY^2, Y^2)^\tau$ und damit $\text{rk Cov}[\text{svec}(XX^\tau)] = 2 < 3$.
2. $X = (\sin(2\pi U), \cos(2\pi U))^\tau$, $U \sim \text{ufo}[0, 1)$, also X gleichverteilt auf dem Kreis $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$. Dann ist $E[X] = 0$, $\text{Cov}(X) = 1/2\mathbb{I}_2$,

$$\text{Cov}[\text{svec}(XX^\tau)] = \frac{1}{8} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 3 & 0 \\ -1 & 0 & 1 \end{pmatrix},$$

und damit $\text{rk Cov}[\text{svec}(X)] = 2 < 3$.

Anhang C

verwendete Sätze aus anderen Gebieten der Mathematik

C.1 der Sylvestersche Trägheitssatz und Folgerungen

Bei der Herleitung der Stabilitätsaussagen in Kapitel 7 greifen wir des öfteren auf die positiv definite Halbordnung unter Matrizen zurück. Im Unterschied zu mancher Darstellung, die positive Definitheit nur für symmetrische Matrizen einführt, wollen wir folgende Definitionen verwenden:

Definition C.1.1 Eine nicht notwendig hermitesche Matrix $A \in \mathbb{C}^{n \times n}$ heiÙe positiv semidefinit falls für alle $x \in \mathbb{C}^n$ gilt

$$\Im(x^H Ax) = 0, \quad \Re(x^H Ax) \geq 0.$$

A heiÙe positiv definit, falls sie zusätzlich vollen Rang hat und negativ (semi)definit, falls $-A$ positiv (semi)definit ist.

Dies induziert eine Halbordnung auf $\mathbb{C}^{n \times n}$: Wir setzen

$$A \prec B \iff B - A \text{ positiv definit}$$

$$A \preceq B \iff B - A \text{ positiv semidefinit}$$

$$A \succ B \iff B - A \text{ negativ (semi)definit}$$

In diesem Abschnitt werden wir einige Hilfsresultate beim Umgang mit dieser positiv definiten Halbordnung herleiten. Zunächst dazu folgendes Lemma:

Lemma C.1.2 Die Jordansche Normalform von $\mathbb{C}^{n \times n} \ni A \succeq 0$, $J(A)$, hat auf der Diagonalen nur Elemente $\lambda_i \geq 0$ stehen.

Falls A diagonalisierbar ist, ist es möglich, ein $B \in \mathbb{C}^{n \times n}$ anzugeben mit $B^2 = A$. Insbesondere ist B hermitesch wählbar (und dann eindeutig) falls A hermitesch; wir schreiben dann $A^{\frac{1}{2}} := B$. In diesem Fall ist die Bedingung,

$J(A)$ hat auf der Diagonalen nur Elemente $\lambda_i \geq 0$ stehen, auch hinreichend für $A \succeq 0$.

BEWEIS: Da jedes Element λ der Diagonale von $J(A)$ auch als Eigenwert vorkommt, gilt mit $0 \neq v \in \text{Eig}_\lambda(A)$

$$0 \leq v^H A v = \lambda |v|^2.$$

Ist A diagonalisierbar, so ist die Nebendiagonale von $J(A)$ leer. Daher können wir mit $J(A)^{\frac{1}{2}} = \text{diag}[(\sqrt{\lambda_i})_{i=1 \dots n}]$ und T , der Matrix, die A auf Jordangestalt bringt, setzen $A^{\frac{1}{2}} := T J(A)^{\frac{1}{2}} T^{-1}$ und erhalten

$$A^{\frac{1}{2}} A^{\frac{1}{2}} = T J(A)^{\frac{1}{2}} T^{-1} T J(A)^{\frac{1}{2}} T^{-1} = T J(A) T^{-1} = A.$$

Für A hermitesch folgt aus dem Satz über die Hauptachsentransformation, daß A eine eindeutige Darstellung als

$$A = \sum_{\lambda \in \sigma(A)} \lambda \text{oP}_{\text{Eig}_\lambda(A)}$$

mit $\sigma(A) := \{\lambda \in \mathbb{C} : \exists u \neq 0 : Au = \lambda u\}$ besitzt. Daraus folgt notwendig

$$A^{\frac{1}{2}} = \sum_{\lambda \in \sigma(A)} \sqrt{\lambda} \text{oP}_{\text{Eig}_\lambda(A)}.$$

Für A hermitesch kann man A durch U mit $UU^H = \mathbb{I}_n$ diagonalisieren. Daher ist

$$x^H A x = x^H U^H J(A)^{\frac{1}{2}} J(A)^{\frac{1}{2}} U x = \|J(A)^{\frac{1}{2}} U x\|^2 \geq 0$$

////

Ein wichtiges Hilfsmittel bei den folgenden Untersuchungen ist der Sylvestersche Trägheitssatz; dieser lautet

Theorem C.1.3 [Sylvester]

Sei V ein reeller (komplexer) Vektorraum der Dimension $n < \infty$. Sei s eine symmetrische (hermitesche) Bilinearform (Sesquilinearform) auf V , \mathcal{A}_i , $i = 1, 2$ Basen von V und A_i die entsprechenden Matrix-Darstellungen von s bezüglich \mathcal{A}_i . Weiter sei $p_{A_i} := \dim \bigoplus_{\lambda > 0} \text{Eig}_\lambda(A_i)$, $n_{A_i} := \dim \bigoplus_{\lambda < 0} \text{Eig}_\lambda(A_i)$, wobei $\text{Eig}_\lambda(A) := \ker(A - \lambda \mathbb{I}_n)$. Dann gilt

$$p_{A_1} = p_{A_2} \quad n_{A_1} = n_{A_2}, \quad \text{rk } A_1 = \text{rk } A_2 \quad (\text{C.1})$$

BEWEIS: z.B. Fischer (1991), Satz 6.7.4., S. 214

////

Basiswechsel bezüglich der darstellenden Matrix setzen sich so um:

Lemma C.1.4 Sei V ein endlich-dimensionaler reeller (komplexer) Vektorraum. Seien A_i , $i = 1, 2$, die Matrix-Darstellungen einer symmetrischen (hermiteschen) Bilinearform (Sesquilinearform) s auf V bezüglich der Basen \mathcal{A}_i von V . Weiter sei T die darstellende Matrix für den Basiswechsel von \mathcal{A}_1 auf \mathcal{A}_2 . Dann gilt

$$A_1 = T^H A_2 T \quad (\text{C.2})$$

BEWEIS: Sei $v^{(1)} := (v_k^{(1)})_{k=1\dots n}$ der Koordinatenvektor von $v \in V$ bezüglich \mathcal{A}_1 . Dann erhält man den Koordinatenvektor $v^{(2)} := (v_k^{(2)})_{k=1\dots n}$ von v bezüglich \mathcal{A}_2 als $v^{(2)} = Tv^{(1)}$ und analog für die Koordinatenvektoren $w^{(i)}$, $i = 1, 2$, zu $w \in V$.

Damit ist für $v, w \in V$ beliebig

$$\begin{aligned} s(v, w) &= (v^{(1)})^H A_1 w^{(1)} = (v^{(2)})^H A_2 w^{(2)} = (Tv^{(1)})^H A_2 (Tw^{(1)}) = \\ &= (v^{(1)})^H T^H A_2 T w^{(1)} \end{aligned}$$

////

Daraus ergibt sich mit Theorem C.1.3

Korollar C.1.5 Sei $A \in \mathbb{R}^{n \times n}$ symmetrisch. A ist genau dann positiv definit, wenn für jedes $T \in \mathbb{R}^{n \times n}$, $\det T \neq 0$ $T^T A T$ positiv definit ist.

Außerdem gilt folgendes Lemma

Lemma C.1.6 Sei $A \in \mathbb{R}^{n \times n}$ symmetrisch. A ist genau dann positiv definit, wenn für jedes $T \in \mathbb{R}^{n \times n}$, $\det T \neq 0$ $T^{-1} A T$ positiv definit ist.

BEWEIS: klar, da Konjugation mit T , $\det T \neq 0$ die Eigenwerte erhält, und nach Lemma C.1.2 Positiv-Definitheit für symmetrische Matrizen damit äquivalent ist, daß alle Eigenwerte positiv sind. //

Lemma C.1.7 Seien $A, B \in \mathbb{R}^{n \times n}$ symmetrisch, B zusätzlich positiv semidefinit. Dann gilt

$$AB \text{ ist positiv definit} \iff A, B \text{ sind positiv definit}$$

BEWEIS: “ \Rightarrow ”:

Wenn AB positiv definit ist, so müssen auch A, B vollen Rang besitzen, also B positiv definit sein. Damit gilt mit Korollar C.1.5

$$B^{-\frac{1}{2}} A B^{\frac{1}{2}} = B^{-\frac{1}{2}} (AB) (B^{-\frac{1}{2}})^T \succ 0,$$

und mit Lemma C.1.6 folgt

$$A = B^{\frac{1}{2}} (B^{-\frac{1}{2}} A B^{\frac{1}{2}}) B^{-\frac{1}{2}} \succ 0.$$

“ \Leftarrow ”:

Nach Lemma C.1.6 gilt $B^{-\frac{1}{2}} A B^{\frac{1}{2}} \succ 0$, und mit Korollar C.1.5 gilt

$$AB = B^{\frac{1}{2}} (B^{-\frac{1}{2}} A B^{\frac{1}{2}}) (B^{\frac{1}{2}})^T \succ 0.$$

////

Als unmittelbare Folgerung aus diesem Lemma notieren wir

Korollar C.1.8 Seien $A, B, C \in \mathbb{R}^{n \times n}$ symmetrisch, C zusätzlich positiv semidefinit. Dann gilt

$$A \succ B \implies CA \succ CB \tag{C.3}$$

$$AC \succ BC \tag{C.4}$$

Korollar C.1.9 Seien $A, B \in \mathbb{R}^{n \times n}$ symmetrisch, dann gilt

$$A \succ B \quad \implies \quad A^{-1} \prec B^{-1} \quad (\text{C.5})$$

$$A^k \succ B^k \quad \forall k \in \mathbb{N} \quad (\text{C.6})$$

BEWEIS: (5):

$$B^{-1} - A^{-1} = B^{-1}AA^{-1} - B^{-1}BA^{-1} = B^{-1}(A - B)A^{-1} \stackrel{\text{Lemma C.1.7}}{\succ} 0$$

(6): Induktion nach k ; $k = 1$ gilt nach Voraussetzung.

$$A^k - B^k = (A - B)A^{k-1} + B(A^{k-1} - B^{k-1}) \stackrel{\text{Lemma C.1.7}}{\succ} 0$$

////

C.2 Optimierung

C.2.1 ein globaler Optimierungsalgorithmus

In der globalen Optimierung steht man vor dem Problem, eine nicht notwendig konvexe Funktion $F : A \rightarrow \mathbb{R}$, $A \subset \mathbb{R}^k$, zu minimieren. Um dabei zu vermeiden, in einem lokalen Minimum hängen zu bleiben, wurden in der Literatur, vgl. Metropolis et al. (1953), Müller et al. (1986), Rinnooy Kan und Timmer (1989), verschiedene stochastische Suchalgorithmen vorgeschlagen. Wir verwenden in dieser Arbeit einen “Quick and Dirty”-Algorithmus¹ im Geiste von “simulated annealing” (c.f. Kirkpatrick et al. (1983)):

Algorithmus C.2.1 (3-Phasen-Algorithmus)

1. tue von $i = 1$ bis n_1
 - (a) ziehe unabhängig zu den bisherigen X_j , $j < i$, eine Auswertungsstelle X_i aus einer geeigneten, fixierten Verteilung aus $\mathcal{M}_1(\mathbb{B}^k \cap A)$,
 - (b) bestimme $F(X_i)$
2. fertige eine “Top- n_2 -Liste” mit den “besten” n_2 Realisationen $(x_{(i_1)}, F(x_{(i_1)})), \dots, (x_{(i_{n_2})}, F(x_{(i_{n_2})}))$ an;
3. tue von $j = 1$ bis n_3
 - (a) ziehe ein “Zentrum” I_j aus der “Top- n_2 -Liste”
 - (b) ziehe unabhängig X_j mit $|\text{Var } X_j| \downarrow$ für $j \uparrow$ und $\mathbb{E}[X_j] = x_{I_j}$
 - (c) datiere die “Top- n_2 -Liste” auf

Dieser Algorithmus unterscheidet sich vom Simulated Annealing dadurch, daß

- immer mehrere (n_2) Prozesse simultan betrachtet werden
- anstelle der üblichen “Temperaturabsenkung” mit der Rate $1/\log(n_3+n_1)$ schneller, mit $1/\sqrt{n_3}$, “abgekühlt” wird

¹in dem Sinne “Quick and Dirty”, daß keinerlei Struktur der zu optimierenden Funktion ausgenutzt wird, so daß viel zu viele Funktionsauswertungen gemacht werden;

C.2.2 ein Hilfssatz aus der konvexen Analysis

Lemma C.2.2 Sei $A \subset \mathbb{R}^p$ konvex, nicht leer, abgeschlossen, $f : B \rightarrow \mathbb{R}$ unterhalbstetig mit $B \supset A$, B kompakt, konvex. Dann gilt

- f nimmt auf B sein Minimum in einer Stelle $y_0 \in B$ an.
- $f|_A$ nimmt auf A sein Minimum in einer Stelle $x_0 \in A$ an.
- Ist $y_0 \in A$, so ist $x_0 = y_0$.
- Ist $y_0 \notin A$, und f konvex, so gibt es ein $z_0 \in \partial A$ mit $f(z_0) = \inf_{z \in A} f(z)$.
- Ist zusätzlich zu (c) $p = 1$, so gilt für dieses $z_0 \in \partial A$

$$z_0 = \underset{z}{\operatorname{argmin}}\{|z - y_0| : z \in A\} \quad (\text{C.7})$$

BEWEIS:

(a) Ohne Einschränkung sei $m := \inf f \neq \infty$. Die Mengen $C_n := \{f(y) \leq m_n\}$, $m_n \downarrow m$, bilden eine absteigende Folge nichtleerer, abgeschlossener Mengen. Angenommen $C := \bigcap_n C_n$ sei leer. Dann ist $B = C^c = \bigcup_n \{f(y) > m_n\}$, also bilden die C_n^c eine offene Überdeckung des Kompaktums B , somit gilt $B = \bigcup_{n \leq N} \{f(y) > m_n\}$ für ein $N \in \mathbb{N}$ und somit müßte gelten $\inf f = m_N > m$. Also ist C nichtleer und jedes $x \in C$ erreicht das Infimum.

(b) folgt aus (a), da $A \cap C_n$ abgeschlossen.

(c) klar.

(d) Sei $m = f(y_0)$. Dann gilt mit $x_t = (1-t)x_0 + ty_0$ $x_t \in B$ für $t \in (0, 1)$. Daher ist $f(x_t) \leq tm + (1-t)f(x_0)$. Ist $f(x_0) = m$, so folgt $f(x_t) = m$, also auch für $f(x_{t_0})$ mit $x_{t_0} \in \partial A$. Sonst gilt $f(x_t) < f(x_0)$ für alle $t \in (0, 1)$, also $x_0 \in \partial A$.

(e) $\partial A = \{e_1, e_2\}$, $e_1 \leq e_2 \in \bar{\mathbb{R}}$, und y_0 liegt in genau einem der Intervalle $[-\infty, e_1]$, $[e_2, \infty]$, ohne Einschränkung im ersten. Dann wird, wie im Beweis von (d) ersichtlich, das Minimum in $f(e_1)$ angenommen.

////

C.3 Funktionalanalysis

C.3.1 einige Tatsachen

Wir wiederholen hier einige Tatsachen aus der Funktionalanalysis, die wir im Text nicht gesondert begründen.

- Jede nichtleere, konvexe, abgeschlossene Teilmenge eines Hilbertraums enthält ein eindeutiges Element kleinster Norm. (Rudin (1974), Theorem 4.10.)
- Jede Teilmenge eines lokal konvexen, topologischen Vektorraums ist genau dann beschränkt, wenn sie schwach beschränkt ist. (Rudin (1973), Theorem 3.18.)
- Jede konvexe Teilmenge eines lokal konvexen, topologischen Vektorraums ist genau dann abgeschlossen, wenn sie schwach abgeschlossen ist. (Dunford und Schwartz (1957), Thm. V 3.13)

- [Riesz'scher Darstellungssatz] Zu jedem stetigen, linearen Funktional A auf $L_p(\mu)$, μ ein σ -finites Maß und $1 \leq p < \infty$, gibt es genau ein erzeugendes Element $g \in L_q(\mu)$, $1/q + 1/p = 1$, so daß

$$Af = \int g f d\mu \quad \forall f \in L_p(\mu); \quad \|A\|_{\text{op.}} = \|g\|_{L_q} \quad (\text{C.8})$$

(Rudin (1974), Thm. 6.16)

C.3.2 schwache Unterhalbstetigkeit von $\|\cdot\|_\infty$ in L_2

Wir geben hier einen von H. Rieder vorgeschlagenen Beweis für die schwache Unterhalbstetigkeit der $L_\infty(\mu)$ -Norm in $L_2(\mu)$, μ ein σ -finites Maß, basierend auf dem Riesz'schen Darstellungssatz. Ein weiterer Beweis ergibt sich aus Lemma 5.3.2 und Prop. 5.3.3 in Rieder (1994).

Proposition C.3.1 *Sei μ ein σ -finites Maß. Dann sind die Mengen $A_b := \{x \in L_2(\mu) : \|x\|_{L_\infty(\mu)} \leq b\}$ für $0 \leq b \leq \infty$ schwach abgeschlossen in $L_2(\mu)$.*

BEWEIS: Seien b und $x_n \in A_b$ gegeben, $x_n \rightharpoonup x$, d.h. x_n konvergiere schwach in L_2 gegen x . Zu zeigen ist $\|x\|_{L_\infty} \leq b$. Für $b = \infty$ ist nichts zu zeigen. Sei also $b < \infty$. Wir definieren die Abbildung

$$\begin{aligned} A : L_2(\mu) &\rightarrow \mathbb{R} \\ g &\mapsto A(g) := \lim \int g x_n d\mu = \int g x d\mu \end{aligned}$$

Diese Abbildung ist linear. Es gilt die elementare Abschätzung

$$|Ag| = \left| \int g x d\mu \right| = \left| \lim \int g x_n d\mu \right| \leq \liminf_n \|x_n\|_{L_\infty(\mu)} \|g\|_{L_1(\mu)}, \quad (\text{C.9})$$

also ist A , definiert auf $L_2(\mu)$, auch stetig in $L_1(\mu)$. Da aber der $L_2(\mu)$ dicht in $L_1(\mu)$ liegt, gibt es eine eindeutige, stetige Fortsetzung \bar{A} von A auf $L_1(\mu)$, und es gilt

$$\|A\|_{\text{op}} = \|\bar{A}\|_{\text{op}} \leq \liminf_n \|x_n\|_{L_\infty(\mu)} \quad (\text{C.10})$$

Weiterhin liefert der Riesz'sche Darstellungssatz die Existenz eines Elementes $y \in L_\infty(\mu)$, mit

$$\bar{A}g = \int gy d\mu, \quad \|\bar{A}\|_{\text{op}} = \|y\|_{L_\infty(\mu)} \quad (\text{C.11})$$

Andererseits erzeugt auch x das Funktional \bar{A} auf der dichten Teilmenge $L_2(\mu)$, so daß $x = y$, also gilt die Behauptung

$$\|x\|_{L_\infty(\mu)} \leq \liminf_n \|x_n\|_{L_\infty(\mu)} \leq b.$$

////

Anhang D

Bestimmung robust–optimaler IC’s bei normalverteilten Scores

D.1 Problemstellung

D.1.1 allgemeiner Fall

Eine Schwierigkeit bei der Konstruktion optimal–robuster Verfahren von der Gestalt

“Minimiere die (geeignet standardisierte) Norm der asymptotischen Kovarianz eines Schätzers im idealen Modell unter einer Schranke an den (geeignet standardisierten) Bias längs einer schrumpfenden Umgebung”

ergibt sich in der Bestimmung der Lagrange–Multiplikatoren A — und im allgemeinen auch a . Diese sorgen für die Einhaltung der Zentriertheitsbedingung $E[\psi] = 0$ und der Korrelationsbedingung $E[\psi\Lambda^\tau] = \mathbb{I}$ bei der Definition von Influenzkurven, wie wir sie an verschiedenen Stellen dieser Arbeit in verschiedenen Varianten eingeführt haben, und werden in Abhängigkeit einer Clipping–Schwelle b ermittelt. Diese wiederum bestimmt sich über den relativen Effizienzverlust im idealen Modell gegenüber der klassisch optimalen Influenzkurve im Sinne eines Anscombe–Kriteriums (vgl. Anscombe (1960)).

D.1.2 Relevanz für das Filter–Problem

In Kapiteln 3 und 4 haben wir gesehen, daß unter (V3) und (V4), die klassische Scoresfunktion Λ jeweils wieder normalverteilt ist, genauer $\Lambda^{\text{rIC}} \sim \mathcal{N}_p(0, \mathcal{I}^{\text{rIC}})$ bzw. $\Lambda^{\text{mIC}} \sim \mathcal{N}_{(m+1)p}(0, \mathcal{I}^{\text{mIC}})$ wobei $\mathcal{I}^{\text{rIC}} = \Sigma_{t|t}^{\text{KK}}$ und \mathcal{I}^{mIC} wie in (5) spezifiziert ist; wir schreiben dafür im folgenden einheitlich Λ und \mathcal{I} , sowie n für die Dimension von Λ und gehen dabei davon aus, daß $\mathcal{I} \succ 0$, so daß wir die Eigenschaften der multivariaten Normalverteilung ausnutzen können, um die Schwierigkeiten bei der Bestimmung der Lagrange–Multiplikatoren zu meistern. So ist ein Zentrierungs– a nicht erforderlich, da aus $\mathcal{L}(-\Lambda) = \mathcal{L}(\Lambda)$ bereits die

Zentriertheit der Hampel–Krasner–Influenzkurve $A\Lambda \min\{1, \frac{b}{|A\Lambda|}\}$ folgt.

Daher geht es nun zunächst um die Lösung der Fixpunkt-Gleichungen (17), (24) sowie (9). Diese lassen sich kaum je explizit berechnen, und so muß man im allgemeinen auf numerische Algorithmen zurückgreifen.

D.2 Berechnung von A

Um die Notation abzukürzen, lassen wir im folgenden den Zeitindex t weg. Alle Schritte im folgenden sind so zu verstehen, daß sie für jedes t durchzuführen sind.

D.2.1 multiplikative Struktur im Regressionsmodell

Im linearen Regressionsmodell bei symmetrischer Fehlerverteilung hat die wichtige Klasse der Hampel–Krasner–Influenzkurven — vgl. (15) die Gestalt

$$\psi := AY u \min\{1, \frac{b}{|AY u|}\}, \quad (\text{D.1})$$

denn die L_2 -Ableitung hat¹ die Gestalt $\Lambda = Y u$ mit einem skalar transformierten Residuum u und einem dazu unabhängigen \mathbb{R}^n -wertigen Regressor Y . Die Matrix A sorgt dabei dafür, daß

$$\mathbb{E}[\psi \Lambda^\tau] = A \mathbb{E}[Y Y^\tau u^2 \min\{1, \frac{b}{|AY u|}\}] \stackrel{!}{=} \mathbb{I}_p \quad (\text{D.2})$$

Diese multiplikative Struktur der L_2 -Ableitung vereinfacht die Berechnungen beträchtlich, denn wir können nach Fubini das Clipping in einer inneren Integration über die Fehlerverteilung (gegeben den Regressor) verarbeiten und anschließend über die Regressoren mitteln.

D.2.2 polare Darstellung der normalen Scores

Eine solche multiplikative Struktur der L_2 -Ableitung steht nicht in allen multivariaten Modellen von vornherein zur Verfügung. Im Fall normalverteilter Scores Λ können wir diese aber künstlich herstellen, indem wir zur polaren Darstellung von Λ übergehen: Wir setzen $\tilde{\Lambda} := \mathcal{I}^{-\frac{1}{2}} \Lambda$ mit $\mathcal{I}^{-\frac{1}{2}}$ der positiv definiten, symmetrischen Wurzel von \mathcal{I}^{-1} , so daß $\tilde{\Lambda} \sim \mathcal{N}(0, \mathbb{I}_p)$. Weiterhin definieren wir $Y := \tilde{\Lambda}/|\tilde{\Lambda}|$ und $u := |\tilde{\Lambda}|$. Dann ist nach Lemma A.2.2

$$\Lambda = \mathcal{I}^{\frac{1}{2}} Y u, \quad \text{mit } Y \sim \text{ufo}(\omega_{n-1}), \quad u^2 \sim \chi_n^2 \quad (\text{D.3})$$

Bemerkung D.2.1 Um die Bestimmung einer Matrix A kommt man im Mehrdimensionalen selbst dann nicht herum, wenn man anstelle des One-Step-Schätzers (18) bei der Schätzung von β_t versuchen würde, die M-Gleichung zu lösen, um $\beta_{t|t}^{\text{ric}}$ zu erhalten, d.h. versuchen würde ein β zu finden, so daß

$$\psi \left(\begin{array}{c} \beta_{t|t-1}^{\text{kk}} - \beta \\ y_t - Z_t \beta \end{array} \right) \stackrel{!}{=} 0. \quad (\text{D.4})$$

¹vgl. Rieder (1994), Rem. 7.4.9

Im Kontext von M-Schätzern wird im allgemeinen auf eine Korrelationsbedingung verzichtet; man stößt aber dennoch auf eine Matrix A , mit der man die Scores transformiert, in Gestalt der Inversen der Ableitung $\partial_\theta \mathbb{E}_\theta[\Lambda_\theta]$, was aber unter Regularitätsvoraussetzungen äquivalent zu unserer Korrelationsbedingung ist, wie man mit partieller Integration nachweist.

Zur Lösung dieses Problem schlagen wir folgenden Algorithmus vor:

Algorithmus D.2.2

1. Initialisierung: Berechne \mathcal{I} , $\mathcal{I}^{\frac{1}{2}}$. Setze $A_0 := \mathcal{I}^{-1}$, $i := 0$.

2. A-Schritt: Berechne

$$A_{i+1}^{-1} := \mathcal{I}^{\frac{1}{2}} \mathbb{E}[Y Y^\tau r(Y, b, A_i, \mathcal{I})] \mathcal{I}^{\frac{1}{2}} \quad (\text{D.5})$$

mit

$$r(Y, b, A, \mathcal{I}) = \mathbb{E}\left[u \min\left(u, \frac{b}{|A \mathcal{I}^{\frac{1}{2}} Y|}\right) \middle| Y\right] \quad (\text{D.6})$$

3. $i := i+1$; gehe zu Schritt 2 bis Konvergenz eintritt. Dann setze $A := \lim_i A_i$.

Für die separat gestutzten Varianten des rIC und des mIC führen wir für $i = 1, 2$ ganz analog zu den Definitionen in Abschnitt D.1.2 die Größen $\tilde{\Lambda}_i$, u_i , Y_i wie in (3) ein, nur mit C_i anstelle \mathcal{I} .

Algorithmus D.2.2 wird dementsprechend modifiziert auf:

Algorithmus D.2.3

1. Berechne C_1 , $C_1^{\frac{1}{2}}$, C_2 , $C_2^{\frac{1}{2}}$.

Setze $A_0 := (C_1 + C_2)^{-1}$; setze $i := 0$.

2. A-Schritt: Berechne

$$A_{i+1}^{-1} := C_1^{\frac{1}{2}} \mathbb{E}[Y_1 Y_1^\tau r(Y_1, b_1, A_i, C_1)] C_1^{\frac{1}{2}} + C_2^{\frac{1}{2}} \mathbb{E}[Y_2 Y_2^\tau r(Y_2, b_2, A_i, C_2)] C_2^{\frac{1}{2}} \quad (\text{D.7})$$

mit $r(Y, b, A, \mathcal{I})$ aus (6)

Hierbei benutzen wir die Tatsache, daß im idealen Modell Λ_1 und Λ_2 unabhängig sind.

D.2.3 technische Hilfsresultate

Bevor wir uns nun daran machen können, die Integrale in Schritt 2 des Algorithmus D.2.2 und D.2.3 zu berechnen, benötigen wir einige Hilfsresultate: Zunächst ist es für die Berechnung von r , und für das später eingeführte s nützlich die gestutzten Momente von u zu kennen. Dazu definieren wir

Definition D.2.4 Sei $X \sim \mathcal{N}(0, \mathbb{I}_n)$. Dann führen wir zu $u := |X|$ die folgenden drei Funktionen ein:

$$\mathcal{F}_u : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}, \quad t \mapsto \mathcal{F}(t) = \Pr(u \leq t) \quad (\text{D.8})$$

$$\mathcal{E}_u : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}, \quad t \mapsto \mathcal{E}(t) = \mathbb{E}[u \mathbb{I}_{u \leq t}] \quad (\text{D.9})$$

$$\mathcal{V}_u : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}, \quad t \mapsto \mathcal{V}(t) = \mathbb{E}[u^2 \mathbb{I}_{u \leq t}] \quad (\text{D.10})$$

Lemma D.2.5 Seien $X_i \sim \Gamma((n+i)/2)$, $i = 0, 1, 2$. Dann gilt

$$\begin{aligned}\mathcal{F}_u(t) &= \Pr(X_0 \leq t^2/2), \\ \mathcal{E}_u(t) &= \frac{\sqrt{2}\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \Pr(X_1 \leq t^2/2), \\ \mathcal{V}_u(t) &= n \Pr(X_2 \leq t^2/2)\end{aligned}\tag{D.11}$$

BEWEIS: \mathcal{F}_u , \mathcal{E}_u , \mathcal{V}_u lassen sich auch schreiben als $E[u^i \mathbf{I}_{u \leq t}]$, $i = 0, 1, 2$. Wegen der Rotationssymmetrie folgt mit Übergang zu Polarkoordinaten mit von i unabhängigen Konstanten c_0 , c

$$E[u^i \mathbf{I}_{u \leq t}] = c_0 \int \mathbf{I}_{\{|u| \leq t\}} |u|^i e^{-|u|^2/2} \lambda^n(du) = c 2^{i/2} \int \mathbf{I}_{\{0 \leq s \leq t^2/2\}} s^{\frac{n+i}{2}-1} e^{-s} \lambda(ds)$$

Dabei erhalten wir wegen $\mathcal{F}_u(\infty) = 1$ die Konstante c als $1/\Gamma(n/2)$ und erkennen im letzten Term das $\Gamma(\frac{i+n}{2})$ -fache der Dichte einer $\Gamma((n+i)/2)$ -verteilten Variable. ////

Weil aber die Γ -Verteilung nicht in allen Programmpaketen verfügbar ist wollen wir uns auf den Fall der eindimensionalen Normalverteilung zurückspielen. Hierzu dient folgende

Definition D.2.6 Für $k \in \mathbb{N}_0$, $X \sim \mathcal{N}(0, 1)$ definieren wir

$$\gamma_k : \mathbb{R} \rightarrow \mathbb{R}, \quad t \mapsto E[X^k \mathbf{I}_{X \leq t}]\tag{D.12}$$

Mit dieser Funktion können wir folgende Lemmata herleiten, deren Beweis jeweils klar ist.

Lemma D.2.7 Mit $\Phi(t)$, der Verteilungsfunktion von $\mathcal{N}(0, 1)$ und $\varphi(t)$ der dazugehörigen Lebesguedichte, bestehen für γ_k folgende Rekursionen:

$$\gamma_0(t) = \Phi(t) \quad \gamma_1(t) = -\varphi(t)\tag{D.13}$$

$$\gamma_k(t) = -t^{k-1} \varphi(t) + (k-1) \gamma_{k-2}(t) \quad \text{for } k > 1\tag{D.14}$$

Explizit lauten diese Ausdrücke für $k \geq 0$, wenn wir leere Produkte als 1, leere Summen als 0 interpretieren,

$$\gamma_{2k}(t) = 2^k \Gamma(k + \frac{1}{2}) (\Phi(t)/\Gamma(\frac{1}{2}) - \frac{t}{2} \varphi(t) \sum_{j=0}^{k-1} (\frac{t^2}{2})^j / \Gamma(j + \frac{3}{2}))\tag{D.15}$$

$$\gamma_{2k+1}(t) = -\varphi(t) 2^k k! \sum_{j=0}^k (\frac{t^2}{2})^j / j!\tag{D.16}$$

Außerdem sind die γ_k differenzierbar mit Ableitung

$$\frac{d}{dt} \gamma_k(t) = \varphi(t) t^k\tag{D.17}$$

Lemma D.2.8 Sei $X \sim \mathcal{N}(0, \mathbb{I}_n)$. Dann besitzt $u := \|X\|$ die Lebesgue-Dichte

$$h(x) := \frac{1}{2^{\frac{n}{2}-1} \Gamma(\frac{n}{2})} x^{n-1} \exp(-\frac{x^2}{2}) \mathbf{I}_{x \geq 0}\tag{D.18}$$

Korollar D.2.9 Mit γ_k aus Definition D.2.6 erhalten wir

$$\mathcal{F}_u(t) = \frac{2^{\frac{3-n}{2}} \Gamma(\frac{1}{2})}{\Gamma(\frac{n}{2})} (\gamma_{n-1}(t) - \gamma_{n-1}(0)) \quad (\text{D.19})$$

$$\mathcal{E}_u(t) = \frac{2^{\frac{3-n}{2}} \Gamma(\frac{1}{2})}{\Gamma(\frac{n}{2})} (\gamma_n(t) - \gamma_n(0)) \quad (\text{D.20})$$

$$\mathcal{V}_u(t) = \frac{2^{\frac{3-n}{2}} \Gamma(\frac{1}{2})}{\Gamma(\frac{n}{2})} (\gamma_{n+1}(t) - \gamma_{n+1}(0)) \quad (\text{D.21})$$

Außerdem ist jede der Funktionen \mathcal{F}_u , \mathcal{E}_u , \mathcal{V}_u beliebig oft differenzierbar mit global beschränkten Ableitungen, d.h. für jedes ν gibt es eine Konstante C_ν , so daß

$$\left\| \frac{d^m}{dt^m} \mathcal{F}_u \right\|_\infty, \left\| \frac{d^m}{dt^m} \mathcal{E}_u \right\|_\infty, \left\| \frac{d^m}{dt^m} \mathcal{V}_u \right\|_\infty \leq C_\nu \quad \forall m \leq \nu \quad (\text{D.22})$$

D.2.4 Berechnung der Integrale

inneres Integral

Um (6) auszuwerten, benützen wir Korollar D.2.9 und gehen dabei davon aus, daß Verfahren zur Auswertung der Dichte resp. der Verteilungsfunktion einer $\mathcal{N}(0, 1)$ -Variablen zur Verfügung stehen: Wir definieren

$$c(b, A, \mathcal{I}, Y) := b / |A \mathcal{I}^{\frac{1}{2}} Y| \quad (\text{D.23})$$

und sehen, daß $r(\cdot)$ aus (6) auch geschrieben werden kann als

$$r(Y, b, A, \mathcal{I}) = \mathcal{V}_u(c) + c(\mathbb{E}[u] - \mathcal{E}_u(c)) = \mathcal{V}_u(c) + c(\sqrt{2} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} - \mathcal{E}_u(c)) \quad (\text{D.24})$$

wobei man den Wert für $\mathbb{E}[u]$ wegen monotoner Konvergenz als $\lim_{c \rightarrow \infty} \mathcal{E}_u(c)$ interpretieren kann.

äußeres Integral

Um (5) zu berechnen, benützen wir die folgende Standard-Darstellung eines Integrals über die Oberfläche ω_{n-1} der Einheitssphäre S_{n-1} in \mathbb{R}^n in Sinoiden: Für die Winkel-Variablen $\vartheta_1, \dots, \vartheta_{n-1}$, kürzen wir $\cos(\vartheta_i)$ und $\sin(\vartheta_i)$ durch c_i resp. s_i ab. Dann gilt — vgl. Mardia et al. (1979) (15.1.3) und (15.1.5) —

Lemma D.2.10

$$\int f(y) \omega_{n-1}(dy) = \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} f(Z_n) \prod_{i=1}^{n-1} s_i^{i-1} d\vartheta_1 \dots d\vartheta_{n-1} \quad (\text{D.25})$$

mit

$$Z_n = (z_1, \dots, z_n)^\tau, \quad z_k = \begin{cases} \prod_{j=1}^{k-1} s_{n-j} c_{n-k} & \text{für } k < n \\ \prod_{j=1}^{n-1} s_{n-j} & \text{für } k = n \end{cases} \quad (\text{D.26})$$

In unserem Zusammenhang ist $f(y) = YY^{\tau}r(Y, b, A, \mathcal{I})$, wobei die Integration koordinatenweise erfolgt. Aus der Differenzierbarkeitseigenschaft in Korollar D.2.9 schließen wir, daß mit c aus (23), r beliebig oft differenzierbar in c ist.

Außerdem bemerken wir, daß r aus (6) gleichmäßig beschränkt ist durch

$$0 \leq r \leq E[u^2] = E[\chi_n^2] = n \quad (\text{D.27})$$

Offenbar ist c auch beliebig oft differenzierbar, wenn nur $|AZY|$ von 0 weg beschränkt bleibt; dies ist aber wegen der Regularität von \mathcal{I} in der Tat der Fall, denn es gilt, daß

$$\exists 0 < m < M : \quad \forall \theta \in \Theta, \forall u \in \mathbb{R}^n, |u| = 1 : \quad m \leq |\mathcal{I}_{\theta}u| \leq M \quad (\text{D.28})$$

so daß wir für $u \in \mathbb{R}^n$, $|u| = 1$ und $i \geq 0$ erhalten

$$|A_{i+1}u| = |\mathcal{I}_{\theta}^{-\frac{1}{2}}(E[YY^{\tau}r(Y, b, A_i, \mathcal{I}_{\theta})])^{-1}\mathcal{I}_{\theta}^{-\frac{1}{2}}u| \geq \frac{n}{m} \quad (\text{D.29})$$

Mit der Kettenregel folgt $f(y) \in \mathcal{C}^{\infty}$ auf ω_{n-1} . Daher können wir, um f numerisch zu integrieren, zunächst die Transformation aus Lemma D.2.10 anwenden und dann koordinatenweise integrieren, indem wir Quadratur-Verfahren hoher Ordnung und/oder Extrapolationsideen verwenden.

Ist die Dimension von Λ aber hoch, so können wir die Integration längs ω_{n-1} mithilfe Monte-Carlo durchführen.

D.2.5 Konvergenz der Algorithmen D.2.2 und D.2.3

Eine wichtige Frage ist nun die Konvergenz der Algorithmen D.2.2 und D.2.3. Ihre Beantwortung würde den Rahmen dieser Arbeit sprengen; wir zitieren daher hier nur das Resultat. Ein Beweis findet sich in Ruckdeschel (2000a).

Theorem D.2.11 *Die Algorithmen D.2.2 und D.2.3 mit Startwerten wie angegeben konvergieren für hinreichend großes b geometrisch.*

Ebenfalls für hinreichend großes $b < \infty$ kann man zur Lösung der Fixpunktgleichung ein Newtonverfahren angeben, das lokal superquadratisch konvergiert.

Bemerkung D.2.12 Die Verwendung des Newton-Verfahrens verbietet sich aus Dimensionsgründen schon für moderates p , q , da die entsprechenden Ableitungsterme dann $(pq)^2$ -dimensional werden.

D.3 Berechnung von b

Wenn man gleichzeitig b im total gestutzten Fall respektive b_1 , b_2 im separat gestutzten Fall gemäß Kriterium (19) bestimmen will, so muß dies wegen der Interdependenz von A und b simultan zu A geschehen. Da wir im separat gestutzten Fall nur den Fall $b_2 < \infty$ betrachten bezeichne im folgenden b_i einen Iterationswert für b im total gestutzten, für b_2 im separat gestutzten Fall. Ein solcher simultaner Ansatz sähe so aus:

Algorithmus D.3.1 ((A, b)-Algorithmus)

1. Finde \mathcal{I} , $b_l > b_{\min}$, $b_l < b_r < \infty$; setze $i := 0$, $A_{-1} := \mathcal{I}^{-1}$.

2. $b_i := 0.5(b_r + b_l)$,
3. Berechne $A_i = A(b_i)$ mit Algorithmus D.2.2 oder Algorithmus D.2.3, wobei als Startwert für A A_{i-1} verwendet wird.

4. Berechne

$$\text{tr} \Sigma(A) = \text{tr}(A \mathcal{I}^{\frac{1}{2}} \text{E}[Y Y^\tau s(Y, b, A, \mathcal{I})] \mathcal{I}^{\frac{1}{2}} A) \quad (\text{D.30})$$

wobei wir s definieren als

$$s(Y, b, A, \mathcal{I}) := \text{E}[u^2 \min(1, \frac{b^2}{|A \mathcal{I} Y|^2}) | Y] \quad (\text{D.31})$$

5. $i := i + 1$; ist $\text{tr} \Sigma(A) < \delta \text{tr} \mathcal{I}^{-1}$, so setze $b_r := b_i$, sonst $b_l := b_i$; Abbruch falls $b_r - b_l$ klein; andernfalls gehe zu Schritt 2.

Bei der Definition dieses Bisektions-Algorithmus' haben wir die Tatsache benützt, daß die Funktion $\text{tr} \Sigma(A(b), b)$ monoton in b ist. Dies sieht man so:

Weil die Influenzkurve, die sich als Lösung von (2) mit $b = b_1$ ergibt, zugleich das Optimierungsproblem aus dem Eingangsabschnitt D.1.1 löst, bei dem auch alle Lösungen desselben Optimierungsproblems mit $b = b_2 < b_1$ zulässige Vergleichsfunktionen sind, müssen alle $b_2 < b_1$ ein größeres $\text{tr} \Sigma(A(b_2), b_2)$ liefern als b_1 .

Bei der Berechnung von s erhalten wir ähnlich wie in (24)

$$s(Y, b, A, \mathcal{I}) = \mathcal{V}_u(c) + c(1 - \mathcal{F}_u(c)) \quad (\text{D.32})$$

so daß wir (30) genau wie in Abschnitt D.2.4 beschrieben berechnen können. Auch hier zitieren wir nur die Konvergenzresultate, wobei für den Newtonalgorithmus dasselbe gilt wie in Bemerkung D.2.12:

Theorem D.3.2 *Der kombinierte (A, b) -Algorithmus D.3.1 konvergiert für hinreichend kleines δ geometrisch.*

Ebenfalls für hinreichend kleines δ kann man zur Lösung des gesamten Problems ein Newtonverfahren angeben, das lokal superquadratisch konvergiert.

D.4 ein Resultat über die minimale Stutzhöhe in Abhängigkeit der Dimension

Bei der Wahl des Effizienzverlustes, den man einzugehen bereit ist, ist folgende Proposition im Zusammenhang mit normalen Scoresfunktionen zu beachten

Proposition D.4.1 *Sei die klassische Scoresfunktion Λ im idealen Modell normalverteilt. Dann gilt für die minimale Stutzhöhe b_{min} , für die entsprechend beschränkte Influenzkurven ψ_b existieren — d.h. $\text{E}[\psi_b \Lambda^\tau] = \mathbb{I}_p$, $\text{E} \psi = 0$, und $|\psi|_M \leq b$ — im Fall*

- $M = \mathcal{I}^{-1} : b_{\text{min}} = \frac{p \Gamma(\frac{p}{2})}{\sqrt{2} \Gamma(\frac{p+1}{2})} = \sqrt{p} + o(p^0)$
- $M = \text{Cov}(\psi) : b_{\text{min}} = \sqrt{p}$

BEWEIS: Rieder und Ruckdeschel (2000b)

////

Korollar D.4.2 *Mit wachsender Dimension p geht der maximal zu erreichende "Effizienzverlust" δ_{\max} gegen 1; d.h. der relative Informationsverlust, der sich daraus ergibt, daß man anstelle der vollen Scoresfunktion nur mit deren Richtung arbeitet, geht gegen 0. Genauer erhalten wir*

Tabelle D.1: b_{\min} und δ_{\max} in Abhängigkeit der Dimension der IC

p	b_{\min}	δ_{\max}
1	$\sqrt{\pi/2} \doteq 1.253$	$\frac{\pi}{2} - 1 \doteq 57.1\%$
2	$2\sqrt{2/\pi} \doteq 1.600$	$\frac{4}{\pi} - 1 \doteq 27.3\%$
3	$\frac{3}{2}\sqrt{\pi/2} \doteq 1.880$	$\frac{3\pi}{8} - 1 \doteq 17.8\%$
4	$\frac{8}{3}\sqrt{2/\pi} \doteq 2.128$	$\frac{32}{9\pi} - 1 \doteq 13.2\%$
5	$\frac{15}{8}\sqrt{\pi/2} \doteq 2.350$	$\frac{45\pi}{128} - 1 \doteq 10.4\%$
6	$\frac{16}{5}\sqrt{2/\pi} \doteq 2.553$	$\frac{256}{75\pi} - 1 \doteq 8.6\%$
7	$\frac{35}{16}\sqrt{\pi/2} \doteq 2.742$	$\frac{175\pi}{512} - 1 \doteq 7.4\%$
8	$\frac{128}{35}\sqrt{2/\pi} \doteq 2.918$	$\frac{4096}{1225\pi} - 1 \doteq 6.4\%$
9	$\frac{315}{128}\sqrt{\pi/2} \doteq 3.084$	$\frac{11025\pi}{32768} - 1 \doteq 5.7\%$
10	$\frac{256}{63}\sqrt{2/\pi} \doteq 3.242$	$\frac{65536}{19845\pi} - 1 \doteq 5.1\%$
20	$\frac{20\Gamma(10)}{\sqrt{2}\Gamma(10.5)} \doteq 4.528$	$\frac{\Gamma(11)\Gamma(10)}{\Gamma(10.5)^2} - 1 \doteq 2.5\%$
50	$\frac{50\Gamma(25)}{\sqrt{2}\Gamma(25.5)} \doteq 7.107$	$\frac{\Gamma(26)\Gamma(25)}{\Gamma(25.5)^2} - 1 \doteq 1.0\%$
100	$\frac{100\Gamma(50)}{\sqrt{2}\Gamma(50.5)} \doteq 10.025$	$\frac{\Gamma(51)\Gamma(50)}{\Gamma(50.5)^2} - 1 \doteq 0.5\%$
1000	$\frac{1000\Gamma(500)}{\sqrt{2}\Gamma(500.5)} \doteq 31.631$	$\frac{\Gamma(501)\Gamma(500)}{\Gamma(500.5)^2} - 1 \doteq 0.05\%$

BEWEIS: Die Gleichung, die uns b bestimmt, nimmt im informations- und im selbststandardisierten Fall folgende Gestalt an:

$$p < E[\min\{\alpha(b)|\eta|^2, b^2\}] \stackrel{!}{=} \delta p \quad \Leftrightarrow \quad 1 < E[\min\{\alpha(b)|\eta|^2, b^2\}/p] \stackrel{!}{=} \delta$$

wobei $\eta \sim \mathcal{N}_p(0; \mathbb{I})$. Aber $p \leq E[\min\{\alpha(b)|\eta|^2, b^2\}] \leq b_{\min}^2$, weil, wie wir im Abschnitt zur Bestimmung von b gesehen haben, der Ausdruck $E[\min\{\alpha(b)|\eta|^2, b^2\}]$ monoton in b ist. Andererseits ist nach Proposition D.4.1

$$\lim_{p \rightarrow \infty} b_{\min}^2/p = 1.$$

////

D.5 Verfügbarkeit

In der XploRe-Quantlib `kalman.lib` wurde dies in Form der in Tabelle D.2 aufgelisteten Quantlets realisiert, vgl. auch Ruckdeschel (2000b).

Tabelle D.2: Umsetzung der Algorithmen in XploRe

Quantlet	Realisation der Funktion
<code>itera</code>	Algorithmus D.2.2
<code>iteras</code>	Algorithmus D.2.3
<code>numint2</code> , <code>numint2m</code>	für $p = 2$ Berechnung des äußeren Integrals mit Rombergverfahren — bei standardisierten IC's ist dabei nur die Lösung der Spur von (2) nötig, was in <code>numint2m</code> ausgenutzt wird.
<code>stointp2</code> , <code>stointpm</code>	für $p > 2$ Berechnung des äußeren Integrals mit Monte-Carlo-Integration längs ω_{p-1} ; — <code>stointpm</code> entsprechend wie <code>numint2m</code>
<code>ewinn</code>	berechnet das innere Integral $r(c)$
<code>ew2inn</code>	berechnet das innere Integral $s(c)$
<code>betnormF</code>	berechnet $\mathcal{F}_u(t)$
<code>betnormE</code>	berechnet $\mathcal{E}_u(t)$
<code>betnormV</code>	berechnet $\mathcal{V}_u(t)$
<code>nmomnorm</code>	berechnet $\gamma_n(t)$

Anhang E

Numerik für Blockmatrizen

E.1 Cholesky-Zerlegung von block-indizierten Matrizen

E.1.1 Cholesky-Zerlegung für allgemeine Block-Matrizen

Es ist bekannt — vgl. Stoer (1989), S. 158 — daß der Algorithmus zur Bestimmung der Cholesky-Zerlegung¹ einer symmetrischen, positiv-definiten Matrix S , bei im eigentlichen Sinn tridiagonalen Matrizen diese Struktur erhält, und daß man nach Berechnung der Zerlegung $S = LL^T$ leicht $Sx = y$ für jede rechte Seite y lösen kann, indem wir zunächst rekursiv $Lz = y$ und dann wieder rekursiv nur in der anderen Richtung $L^T x = z$ lösen.

Bei genauerer Betrachtung dieses Algorithmus' fällt auf, daß wir nirgends die Kommutativität der Multiplikation gebraucht haben. Wir können auch vermeiden, Wurzeln zu ziehen, indem wir die Zerlegung $S = \tilde{L}D\tilde{L}^T$ benützen, bei der die Diagonale von \tilde{L} mit 1 besetzt ist und D eine Diagonalmatrix ist, deren Einträge gerade die Quadrate der normalerweise auf der Diagonale von L stehenden Einträge sind. In diesem Sinn können wir den Algorithmus für $A \in \mathbb{R}^{n \times n}$ definieren, wobei wir für $A_{i,j}$ $A[i,j]$ und für $\tilde{L}_{i,j}$ $Lt[i,j]$ schreiben

Algorithmus E.1.1 (Cholesky)

```
For i:= 1...n do
  D[i,i]:= A[i,i]
  For k := 1...i-1 do
    D[i,i] := D[i,i]-Lt[i,k] * D[k,k] * Lt[i,k]
  For j := i+1...n do
    Lt[j,i]:=A[j,i]
    For k := 1...i-1 do
      Lt[j,i]:= Lt[j,i]-Lt[j,k] * D[k,k] * Lt[i,k]
    Lt[j,i]:= 1/D[i,i] * Lt[j,i]
```

Dieser Algorithmus bleibt korrekt, wenn wir anstelle der Skalare $A_{j,i}$, $\tilde{L}_{j,i}$, etc, $p \times p$ Matrizen einsetzen, sofern nur A in Blöcke gemäß Notation 4.4.1 partitioniert werden kann. Dazu müssen folgende Modifikationen / Interpretationen

¹vgl. Stoer (1989), Abschnitt 4.3

vorgenommen werden:

- Die Ausdrücke $Lt[j,k] * D[k,k] * Lt[i,k]$ müssen durch $Lt[j,k] * D[k,k] * \text{trn}(Lt[i,k])$ ersetzt werden, wobei $\text{trn}(Lt[i,k])$ für $\tilde{L}_{(i),(k)}^T$ steht.
- Das Produktzeichen $*$ steht auch für Matrixmultiplikation.
- Die Block-Diagonale von \tilde{L} ist mit \mathbb{I}_p zu besetzen.
- $1/D[i,i]$ ist zu ersetzen durch $\text{inv}(D[i,i])$, was wiederum für $D_{(i),(i)}^{-1}$ steht.

Der Algorithmus ist wohldefiniert, denn eine symmetrische Matrix ist positiv definit genau dann, wenn alle Minoren es sind. Weil aber nach Induktion jeder der Minoren $D_{(i),(i)}$ symmetrisch und positiv definit, und daher insbesondere invertierbar ist, ist dies der Fall, denn sonst ergäbe der Multiplikationssatz für Determinanten von Blockmatrizen nicht-positive Werte für die Determinante der Minoren von A . Wie sich das rekursive Auflösen von $\tilde{L}z = y$ und $D\tilde{L}^T x = z$ für Matrixwertige Einträge umsetzt, ist klar.

Bemerkung E.1.2 Wir können Speicher sparen, indem wir $L_{j,i} := A_{i,j}^T$ zur Initialisierung von $L_{j,i}$ verwenden und dann L bis auf seine (Block-)Diagonale auf der unteren Dreieckshälfte von A speichern. D kann in (Block-)Vektorform gespeichert werden.

E.1.2 Cholesky-Zerlegung für Block-tridiagonale Matrizen

Im Fall von (Block-)tridiagonalen Matrizen A können wir den Algorithmus in der folgenden Form spezialisieren:

Algorithmus E.1.3 (Cholesky)

```
D[1,1] := A[1,1]
For i := 2... n do
  Lt[i,i-1] := A[i,i-1] inv(D[i-1,i-1])
  D[i,i] := A[i,i] - Lt[i,i-1] D[i-1,i-1] trn(Lt[i,i-1])
```

Das Lösen von $\tilde{L}z = y$ und $D\tilde{L}^T x = z$ reduziert sich auf

Algorithmus E.1.4

```
Z[1] := Y[1]
For i:= 2... n do
  Z[i] := Y[i] - Lt[i,i-1] Z[i-1]
X[n] := inv(D[n,n]) Z[n]
For i:= n-1...1 do
  X[i] := inv(D[i,i]) Z[i] - trn(Lt[i+1,i]) X[i+1]
```

Bemerkung E.1.5 (a) Aus numerischer Sicht steht im wesentlichen eine Cholesky-Zerlegung für Block-Matrizen hinter dem Algorithmus von Fahrmeir und Kaufmann (1991), in deren Abschnitt 4.

(b) Natürlich können die Schleifen in den Algorithmen E.1.3 und E.1.4 zu zwei Schleifen der Ordnung n kombiniert werden. Der Gesamtaufwand beträgt $O(np^3)$ Operationen, davon $O(n)$ Matrixinvertierungen von $p \times p$ -Matrizen, die mit $O(p^3)$ zu Buche schlagen, anstelle von $O(n^3p^3)$ Operationen, die man für das Lösen eines allgemeinen Gleichungssystems der Ordnung np braucht.

(c) Wie in Bemerkung (E.1.2) können wir Speicher sparen, indem wir nur die Ausdrücke $D_{(i),(i)}$, $\tilde{L}_{(i),(i-1)}$ in geeigneten Arrays der Dimension $p \times p \times n$ resp. $p \times p \times (n-1)$ ablegen.

E.2 LDM-Zerlegung für Block-indizierte Matrizen

Mit den Ersetzungen des vorigen Abschnitts können wir auch die klassische LR-Zerlegung² von Matrizen ins Block-Kalkül umsetzen. Allerdings ist im allgemeinen ein Permutieren der (Block-)Spalten nötig, der dann aber eine verallgemeinerte Bandstrukturen zerstört.

Um eine solche Struktur so weit wie möglich zu bewahren, kann man versuchen, den Algorithmus ohne Spaltenvertauschungen durchzuführen, was uns dann auf den LDM-Algorithmus von Golub und Van Loan (1983), Abschnitt 5.1 führt, der eine modifizierte LR-Zerlegung einer Matrix A der Form $A = LDR$ bewirkt, wobei D wieder eine (Block-)Diagonalmatrix, L eine untere, R eine obere Dreiecksmatrix ist, was den folgenden Algorithmus liefert

Algorithmus E.2.1 (LDM-Zerlegung)

```

For i := 1...n do
  For j := 1...i-1 do
    L[i,j] := A[i,j]
    For k := 1...j-1 do
      L[i,j] := L[i,j] - L[i,k] * D[k,k] * R[k,j]
    L[i,j] := L[i,j] * inv(D[j,j])
  D[i,i] := A[i,i]
  For k := 1...i-1 do
    D[i,i] := D[i,i] - L[i,k] * D[k,k] * R[k,i]
  For j := i+1...n do
    R[i,j] := A[i,j]
    For k := 1...i-1 do
      R[i,j] := R[i,j] - L[i,k] * D[k,k] * R[k,j]
    R[i,j] := inv(D[j,j]) * R[i,j]

```

Wie bei der Cholesky-Zerlegung nutzen wir eine Block-Tridiagonalgestalt von A aus:

Algorithmus E.2.2 (LDM-Zerlegung)

```

D[1,1] := A[1,1]
For i := 2...n do
  L[i,i-1] := A[i,i-1] inv(D[i-1,i-1])
  R[i-1,i] := inv(D[i-1,i-1]) A[i-1,i]
  D[i,i] := A[i,i] - A[i,i-1] inv(D[i-1,i-1]) A[i-1,i]

```

²vgl. Stoer (1989), Abschnitt 4.1

Das Lösen von $Lz = y$ und $DRx = z$ reduziert sich auf

Algorithmus E.2.3

```
Z[1] := Y[1] For i := 2... n do
    Z[i] := Y[i] - L[i,i-1] Z[i-1]
X[n] := inv(D[n,n]) Z[n] For i := n-1...1 do
    X[i] := inv(D[i,i]) Z[i] - R[i,i+1] X[i+1]
```

Bemerkung E.2.4 Damit der Algorithmus wohldefiniert ist, muß in jedem Schritt $D_{(i),(i)}$ invertierbar sein — vgl. Golub und Van Loan (1983), Theorem 5.5.1!

Literaturverzeichnis

- Aldous D.J. und Eagleson G.K. (1978): On mixing and stability of limit theorems. *Ann. Probab.*, **6**: 325–331.
- Anderson B.D.O. und Moore J.B. (1979): *Optimal filtering*. Information and System Sciences Series. Prentice-Hall.
- (1990): *Optimal control. Linear quadratic methods*. Prentice Hall Information and System Sciences Series. Prentice Hall.
- Anderson T.W. (1993): Nonnormal multivariate distributions: Inference based on elliptically contoured distributions. In: Rao C.R. (Hg.) *Multivariate analysis: future directions. Papers presented at the 7th multivariate conference held at Pennsylvania State University, University Park, PA (USA), May 5-9, 1992.*, Bd. 5 von *North-Holland Ser. Stat. Probab.*, S. 1–24. North-Holland.
- Andrews D.F., Bickel P.J., Hampel F.R., Huber P.J., Rogers W.H. und Tukey J.W. (1972): *Robust estimates of location. Survey and advances*. Princeton University Press, Princeton, N. J.
- Anscombe F.J. (1960): Rejection of outliers. *Technometrics*, **2**: 123–147.
- Aoki M. (1990): *State space modeling of time series*. Information and System Sciences Series. Springer, 2. Aufl.
- Arce G.R., Kim Y.T. und Barner K.E. (1998): Order-statistic filtering and smoothing of time-series. I. In: (Balakrishnan, N. et al., 1998), S. 525–554.
- Athans M. und Falb P.L. (1966): *Optimal control*. McGraw-Hill Electrical and Elektronik Engineering Series. McGraw-Hill.
- Athreya K.B. und Ney P. (1982): Some aspects of ergodic theory and laws of large numbers for Harris-recurrent Markov chains. In: *Nonparametric statistical inference, Budapest 1980, Vol. I.*, Bd. 32 von *Colloq. Math. Soc. Janos Bolyai*, S. 41–56.
- Başar T. und Bernhard P. (1991): *\mathcal{H}^∞ -Optimal Control and Related Minimax Design Problems*. Sytems & Control: Foundations and Applications. Birkhäuser.
- Balakrishnan, N. et al. (Hg.) (1998): *Order statistics: applications.*, Bd. 17 von *Handb. Stat.*. North-Holland.

- Barner K.E. und Arce G.R. (1998): Order-statistic filtering and smoothing of time-series. II. In: (Balakrishnan, N. et al., 1998), S. 555–602.
- Bellman R. (1995): *Introduction to matrix analysis*, Bd. 12 von *Classics in Applied Mathematics*. SIAM, Nachdruck der 1960 Orig. Aufl.
- Beran R. (1981): Efficient robust estimates in parametric models. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, **55**: 91–108.
- (1982): Robust estimation in models for independent non-identically distributed data. *Ann. Stat.*, **10**: 415–428.
- Berndt E.K., Hall B.H., Hall R.E. und Hausman J.A. (1974): Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement*, **3/4**: 653–665.
- Bickel P.J. (1981): Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Stat.*, **9**: 1301–1309.
- Bickel P.J. und Collins J.R. (1983): Minimizing Fisher information over mixtures of distributions. *Sankhya, Ser. A*, **45**: 1–19.
- Bickel P.J., Ritov Y. und Ryden T. (1998): Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Ann. Stat.*, **26**(4): 1614–1635.
- Billingsley P. (1968): *Convergence of probability measures*. John Wiley and Sons, Inc.
- Birmiwal K. und Papantoni-Kazakos P. (1994): Outlier resistant prediction for stationary processes. *Stat. Decis.*, **12**(4): 395–427.
- Birmiwal K. und Shen J. (1993): Optimal robust filtering. *Stat. Decis.*, **11**(2): 101–119.
- Bolzern P., Colaneri P. und De Nicolao G. (1996): Optimal robust filtering with time-varying parameter uncertainty. *Int. J. Control*, **63**(3): 557–576.
- Bonchelet C.G. Jr (1985): Robust recursive estimation in linear models. Tagungsbeitrag zu[r] *International Symposium on Information Theory* in Brighton, England.
- Bonchelet C.G. Jr und Dickinson B.W. (1983): An approach to robust Kalman filtering. In: *Proceedings on the 22nd IEEE Conference on Decision & Control*, Bd. 1, S. 304–305. IEEE Control Systems Society, San Antonio, TX.
- (1987): An extension to the SRIF Kalman filter. *IEEE Trans. Autom. Control*, **AC-32**: 176–179.
- Box G.E.P. und Tiao G.C. (1975): Intervention analysis with applications to economic and environmental problems. *J. Am. Stat. Assoc.*, **70**: 70–79.

- Brand C. (1998): Empirical Evidence on the Money Supply Process in Germany. A State Space Modelling Approach to the Bundesbank's Operating Procedures. Tagungsbeitrag zu[r] Tagung des Vereins für Socialpolitik in Rostock, 23. September 1998.
- Breiman L. (1968): *Probability*. Addison-Wesley Series in Statistics. Addison-Wesley.
- Brockwell P.J. und Davis R.A. (1991): *Time series: theory and methods*. Springer Series in Statistics. Springer, 2. Aufl.
- Brown L.D. (1971): Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statistics*, **42**: 855–903.
- Burmeister E. und Wall K.D. (1982): Kalman filtering estimation of unobserved rational expectations with an application to the German hyperinflation. *J. Econ.*, **20**: 255–284.
- Caines P.E. (1988): *Linear stochastic systems*. Wiley Series in Probability and Mathematical Statistics. Wiley.
- Carlin B.P., Polson N.G. und Stoffer D.S. (1992): A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *J. Am. Stat. Assoc.*, **87**: 493–500.
- Carter C.K. und Kohn R. (1994): On Gibbs sampling for state space models. *Biometrika*, **81**(3): 541–553.
- (1996): Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika*, **83**(3): 589–601.
- Çelebi M.E. und Kurz L. (1996): Robust locally optimal filters: Kalman and Bayesian estimation theory. *Inf. Sci.*, **92**(1-4): 1–32.
- Chan W.s. und Wei W.W. (1992): A comparison of some estimators of time series autocorrelations. *Comput. Stat. Data Anal.*, **14**(2): 149–163.
- Chen H.F. (1985): *Recursive Estimation and Control for Stochastic Systems*. Wiley.
- Chen J. und Patton R.J. (1996): Optimal filtering and robust fault diagnosis of stochastic systems with unknown disturbances. *IEEE Proc., Control Theory Appl.*, **143**(1): 31–36.
- Chow H.K. (1994): Robust estimation in time series: An approximation to the Gaussian sum filter. *Commun. Stat., Theory Methods*, **23**(12): 3491–3505.
- Chung K.L. (1974): *A course in probability theory*, Bd. 21 von *Probability and Mathematical Statistics*. Academic Press, a subsidiary of Harcourt Brace Jovanovich, Publishers, 2. Aufl.
- Cipra T. und Romera R. (1991): Robust Kalman filter and its application in time series analysis. *Kybernetika*, **27**(6): 481–494.
- Cooley T.F. und Prescott E.C. (1976): Estimation in the presence of stochastic parameter variation. *Econometrica*, **44**: 167–184.

- Daumer M. und Falk M. (1998): On-line change-point detection (for state space models) using multi-process Kalman filters. *Linear Algebra Appl.*, **284**(1-3): 125–135.
- de Jong P. und Penzer J. (1998): Diagnosing shocks in time series. *J. Am. Stat. Assoc.*, **93**(442): 796–806.
- Dellacherie C. und Meyer P.A. (1978): *Probabilities and potential*, Bd. 29 von *North-Holland Mathematics Studies*. North-Holland. Transl. from the French.
- Dempster A.P., Laird N.M. und Rubin D.B. (1977): Maximum likelihood from incomplete data via the EM algorithm. Discussion. *J. R. Stat. Soc., Ser. B*, **39**: 1–38.
- Donoho D.L. (1978): *The asymptotic variance formula and large-sample criteria for the design of robust estimators*. Unpublished senior thesis, Department of Statistics, Princeton University.
- Doob J.L. (1953): *Stochastic processes*. Wiley.
- Duncan D.B. und Horn S.D. (1972): Linear dynamic recursive estimation from the viewpoint of regression analysis. *J. Am. Stat. Assoc.*, **67**: 815–821.
- Dunford N. und Schwartz J.T. (1957): *Linear Operators I - General Theory*. Wiley-Interscience., New York.
- Elliott R.J., Aggoun L. und Moore J.B. (1995): *Hidden Markov models. Estimation and control*, Bd. 29 von *Applications of Mathematics*.. Springer.
- Ershov A.A. (1978): Robust filtering algorithms. *Autom. Remote Control*, **39**: 992–996.
- Ershov A.A. und Lipster R.S. (1978): Robust Kalman filter in discrete time. *Autom. Remote Control*, **39**: 359–367.
- Fahrmeir L. und Kaufmann H. (1991): On Kalman filtering, posterior mode estimation and Fisher scoring in dynamic exponential family regression. *Metrika*, **38**(1): 37–60.
- Fahrmeir L. und Künstler R. (1999): Penalized Likelihood Smoothing in Robust State Space Models. *Metrika*, **49**: 173–191.
- Feller W. (1971): *An introduction to probability theory and its applications II*. Wiley Series in Probability and Mathematical Statistics. Wiley, 2. Aufl.
- Ferguson T.S. (1967): *Mathematical statistics. A decision theoretic approach*. Academic Press.
- Ferguson T.S., Genest C. und Hallin M. (2000): Kendall's tau for serial dependence. *Canadian Journal of Statistics*, **28**(3): 587–604.
- Fischer G. (1991): *Lineare Algebra*, Bd. 17 von *Vieweg Studium. Mathematik Grundkurs*. Vieweg, Braunschweig, 9., rev. Aufl.
- Fox A.J. (1972): Outliers in time series. *J. R. Stat. Soc., Ser. B*, **34**: 350–363.

- Franke J. (1985): Minimax-robust prediction of discrete time series. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, **68**: 337–364.
- Franke J. und Poor H.V. (1984): Minimax-robust filtering and finite-length robust predictors. In: *Robust and nonlinear time series analysis. Proc. Workshop, Heidelberg/Ger. 1983*, Nr. 26 in Lect. Notes Stat. Springer.
- Franěk P. (2000): Kalman Filtering. In: Härdle W., Klinke S. und Müller M. (Hg.) *XploRe. Learning Guide.*, Kap. 10, S. 273–284. Springer.
- Fried R., Gather U., Imhoff M. und Bauer M. (2000): Statistical methods in intensive care online monitoring. Techn. Ber. 33/00, SFB 474. Department of Statistics, University of Dortmund, 44221 Dortmund, Germany.
- Gabbouj M., Coyle E.J. und Gallagher N.C. Jr (1992): An overview of median and stack filtering. *Circuits Syst. Signal Process.*, **11**(1): 7–45.
- Garel B. und Hallin M. (1995): Local asymptotic normality of multivariate AR-MA processes with a linear trend. *Ann. Inst. Stat. Math.*, **47**(3): 551–579.
- Gather U., Fried R. und Imhoff M. (2000): Online classification of states in intensive care. Techn. Ber. 15/00, SFB 474. Department of Statistics, University of Dortmund, 44221 Dortmund, Germany.
- Golub G.H. und Van Loan C.F. (1983): *Matrix computations.* North Oxford Academic, Oxford.
- Goodrich R.L. und Caines P.E. (1979): Necessary and sufficient conditions for local second-order identifiability. *IEEE Trans. Autom. Control*, **AC-24**: 125–127.
- Graybill F.A. (1983): *Matrices with applications in statistics.* The Wadsworth Statistics/Probability Series. Wadsworth, Belmont, CA, 2. Aufl.
- Gupta N.K. und Mehra R.K. (1974): Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations. *IEEE Trans. Autom. Control*, **AC-19**: 774–783.
- Hájek J. (1972): Local asymptotic minimax and admissibility in estimation. In: *Proc. 6th Berkeley Sympos. math. Statist. Probab. 1970*, Bd. 1, S. 175–194. Univ. Calif. Press.
- Hall P. (1977): Martingale invariance principles. *Ann. Probab.*, **5**: 875–887.
- Hampel F.R. (1968): *Contributions to the theory of robust estimation.* Dissertation, University of California, Berkeley, CA.
- Hampel F.R., Ronchetti E.M., Rousseeuw P.J. und Stahel W.A. (1986): *Robust statistics. The approach based on influence functions.* Wiley Series in Probability and Mathematical Statistics. Wiley.
- Hannan E.J. (1970): *Multiple time series.* Wiley Series in Probability and Mathematical Statistics. Wiley.
- Hannan E.J. und Deistler M. (1988): *The statistical theory of linear systems.* Wiley Series in Probability and Mathematical Statistics. Wiley.

- Harrison P.J. und Stevens C.F. (1976): Bayesian forecasting. Discussion. *J. R. Stat. Soc., Ser. B*, **38**: 205–247.
- Heyde C.C. und Johnstone I.M. (1979): On asymptotic posterior normality for stochastic processes. *J. R. Stat. Soc., Ser. B*, **41**: 184–189.
- Huber P.J. (1964): Robust estimation of a location parameter. *Ann. Math. Stat.*, **35**: 73–101.
- (1967): The behavior of maximum likelihood estimates under nonstandard conditions. In: Le Cam L.M. und Neyman J. (Hg.) *Proc. 5th Berkeley Symp. Math. Stat. Probab.*, Bd. I, S. 221–233. University of California Press.
- (1981): *Robust statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley.
- Hürzeler M. (1998): *Statistical methods for general state-space models*. Dissertation, Swiss Federal Institute of Technology Zurich, Zürich.
- Jazwinski A.H. (1970): *Stochastic processes and filtering theory*, Bd. 64 von *Mathematics in Science and Engineering*. Academic Press.
- Jeganathan P. (1982): On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. *Sankhya, Ser. A*, **44**: 173–212.
- (1988): On the strong approximation of the distributions of estimators in linear stochastic models, I and II: Stationary and explosive AR models. *Ann. Stat.*, **16**(3): 1283–1314.
- Jensen J.L. und Petersen N.V. (1999): Asymptotic normality of the maximum likelihood estimator in state space models. *Ann. Stat.*, **27**(2): 514–535.
- Johnson N.L. und Kotz S. (1970): *Distributions in statistics: Continuous univariate distributions. Vol. 1,2*. Houghton Mifflin Company., Boston etc.
- Jones R.H. (1986): Time series regression with unequally spaced data. In: *Essays in time series and allied processes, Pap. Hon. E. J. Hannan.*, Bd. 23A von *J. Appl. Probab. Spec. Vol.*.
- Kailath T. (1980): *Linear systems*. Prentice-Hall Information and System Sciences Series. Prentice-Hall.
- Kalman R.E. (1960): A new approach to linear filtering and prediction problems. *Journal of Basic Engineering—Transactions of the ASME*, **82**: 35–45.
- Kalman R.E. und Bucy R. (1961): New results in filtering and prediction theory. *Journal of Basic Engineering—Transactions of the ASME*, **83**: 95–108.
- Karatzas I. und Shreve S.E. (1991): *Brownian motion and stochastic calculus*, Bd. 113 von *Graduate Texts in Mathematics*. Springer, 2. Aufl.
- Kassam S.A. und Lim T.L. (1977): Robust Wiener filters. *J. Franklin Inst.*, **304**: 171–185.

- Kassam S.A. und Poor H.V. (1985): Robust techniques for signal processing: A survey. *Proc. IEEE*, **73**(3): 433–481.
- Kirkpatrick S., Gelatt C. und Vecchi M. (1983): Optimization by Simulated Annealing. *Science*, **220**: 671–680.
- Kirlin R.L. und Moghaddamjoo A. (1986): Robust adaptive Kalman filtering for systems with unknown step inputs and non-Gaussian measurement errors. *IEEE Trans. Acoust. Speech Signal Process.*, **ASSP-34**: 252–263.
- Kitagawa G. (1987): Non-Gaussian state-space modelling of nonstationary time series. *J. Am. Stat. Assoc.*, **82**: 1032–1063.
- Kneser H. (1952): Sur une théorème fondamental de la théorie des jeux. *C. R. Acad. Sci., Paris*, **243**: 2418–2420.
- Kolmogoroff A. (1941): Interpolation und Extrapolation von stationären zufälligen Folgen. *Bull. Acad. Sci. URSS, Ser. Math.*, **5**: 3–14.
- Kreiss J.P. (1987): On adaptive estimation in stationary ARMA processes. *Ann. Stat.*, **15**: 112–133.
- Künsch H.R. (2001): State space models and Hidden Markov Models. In: Barndorff-Nielsen, O. E. and Cox, D. R. and Klüppelberg, C. (Hg.) *Complex Stochastic Systems*, S. 109–173. Chapman and Hall.
- Künstler R. (1995): *Robuste Zustandsraummodelle*. Dissertation, Ludwig Maximilians Universität, München.
- Kushner H.J. (1967): *Stochastic stability and control*, Bd. 33 von *Mathematics in Science and Engineering*. Academic Press.
- (1971): *Introduction to stochastic control*. Holt, Rinehart and Winston, Inc.
- Le Cam L. (1970): On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Ann. Math. Statistics*, **41**: 802–828.
- (1986): *Asymptotic methods in statistical decision theory*. Springer Series in Statistics. Springer.
- Le Cam L. und Yang G.I. (1988): On the preservation of local asymptotic normality under information loss. *Ann. Stat.*, **16**(2): 483–520.
- Ledolter J. (1979): A recursive approach to parameter estimation in regression and time series models. *Commun. Stat., Theory Methods A*, **8**: 1227–1247.
- Levit B.Y. (1981): On asymptotic minimax estimates of the second order. *Theory Probab. Appl.*, **25**: 552–568.
- Little R.J.A. und Rubin D.B. (1987): *Statistical analysis with missing data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. Wiley.
- Loève M. (1960): *Probability theory*. The University Series in Higher Mathematics. D. Van Nostrand Company Inc., 2. Aufl.

- Lütkepohl H. (1993): *Introduction to multiple time series analysis*. Springer, 2. Aufl.
- Ma Y. und Genton M.G. (1998): Highly robust estimation of the autocovariance function. *J. Time Ser. Anal.*, **21**(6): 663–684.
- Magnus J.R. und Neudecker H. (1999): *Matrix differential calculus with applications in statistics and econometrics*. Wiley Series in Probability and Statistics. Wiley, 2. rev. Aufl.
- Mallows C.L. (1980): Some theory of nonlinear smoothers. *Ann. Stat.*, **8**: 695–715.
- Maltz F. (1972): Maximum a-posteriori estimation for tracking non-Gaussian processes. In: *Proc. 2nd Symp. nonlin. Estim. Theory Appl., San Diego 1971*, S. 207–211.
- Marazzi A. (1985): On constrained minimization of the Bayes risk for the linear model. *Stat. Decis.*, **3**: 277–296.
- Mardia K.V., Kent J.T. und Bibby J.M. (1979): *Multivariate analysis*. Probability and Mathematical Statistics. Academic Press.
- Martin R.D. (1979): Approximate conditional-mean type smoothers and interpolators. In: *Smoothing techniques for curve estimation, Proc. Workshop, Heidelberg 1979*, Bd. 757 von *Lect. Notes Math.*, S. 117–143. Springer.
- Martin R.D. und Raftery A.E. (1987): Robustness, computation and non Euclidean models. *J. Am. Stat. Assoc.*, **82**: 1044–1050. Comment.
- Martin R.D. und Yohai V.J. (1986): Influence functionals for time series (with discussion). *Ann. Stat.*, **14**: 781–818.
- Masreliez C.J. und Martin R. (1977): Robust Bayesian estimation for the linear model and robustifying the Kalman filter. *IEEE Trans. Autom. Control*, **AC-22**: 361–371.
- Maz'ya V.G. (1985): *Sobolev spaces*. Springer. Transl. from the Russian by T. O. Shaposhnikova.
- McGarty T.P. (1975): Bayesian outlier rejection and state estimation. *IEEE Trans. Autom. Control*, S. 682–687.
- McLachlan G.J. und Krishnan T. (1997): *The EM algorithm and extensions*. Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics. Wiley.
- McLeish D.L. (1974): Dependent central limit theorems and invariance principles. *Ann. of Probab.*, **2**: 620–628.
- Meinhold R.J. und Singpurwalla N.D. (1989): Robustification of Kalman filter models. *J. Am. Stat. Assoc.*, S. 479–486.
- Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H. und Teller E. (1953): Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**: 1087–1092.

- Meyn S.P. (1989): Ergodic theorems for discrete time stochastic systems using a stochastic Lyapunov function. *SIAM J. Control Optimization*, **27**(6): 1409–1439.
- Meyn S.P. und Caines P.E. (1991): Asymptotic behavior of stochastic systems possessing Markovian realizations. *SIAM J. Control Optimization*, **29**(3): 535–561.
- Meyn S.P. und Tweedie R.L. (1993): *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer.
- Meyr H. und Spies G. (1984): The structure and performance of estimators for real-time estimation of randomly varying time delay. *IEEE Trans. Acoust. Speech Signal Process.*, **ASSP-32**: 81–94.
- Millar P. (1983): The minimax principle in asymptotic statistical theory. In: *Ecole d'ete de probabilites de Saint-Flour XI - 1981*, Bd. 976 von *Lect. Notes Math.*, S. 75–265.
- Morris J.M. (1976): The Kalman filter: A robust estimator for some classes of linear quadratic problems. *IEEE Trans. Inf. Theory*, **IT-22**: 526–534.
- Müller P.H., Nollau V. und Polovinkin A.I. (1986): *Stochastische Suchverfahren. (Random search methods)*. Mathematik für Ingenieure. Verlag Harri Deutsch, Thun-Frankfurt/Main. (Lizenzausg. des VEB Fachbuchverlag Leipzig).
- Nummelin E. (1984): *General irreducible Markov chains and non-negative operators*, Bd. 83 von *Cambridge Tracts in Mathematics*. Cambridge University Press.
- Orey S. (1971): *Lecture notes on limit theorems for Markov chain transition probabilities*, Bd. 34 von *Van Nostrand Reinhold Mathematical Studies*. Van Nostrand Reinhold.
- Pagan A. (1980): Some identification and estimation results for regression models with stochastically varying coefficients. *J. Econ.*, **13**: 341–363.
- Parthasarathy K.R. (1967): *Probability measures on metric spaces*. Probability and Mathematical Statistics. A Series of Monographs and Textbooks. Academic Press.
- Pecchi L. (1993): Time series evidence on the money supply process in the USA: A new state space modelling approach. *Economic modelling*, **10**(3): 187–200.
- Peña D. und Guttman I. (1988): Bayesian approach to robustifying the Kalman filter. In: *Bayesian analysis of time series and dynamic models*, Bd. 94 von *Stat., Textb. Monogr.*, S. 227–253.
- Pupeikis R. (1998): State estimation of dynamic systems in the presence of time-varying outliers in observations. *Informatika, Vilnius*, **9**(3): 325–342.
- Rausch M. (1999): *Robuste Varianten des Kalman-Filters*. Diplomarbeit, Lehrstuhl Mathematik VII, Universität Bayreuth.

- Reinsel G.C. (1997): *Elements of multivariate time series analysis*. Springer Series in Statistics. Springer, 2. Aufl.
- Rieder H. (1977): Least favorable pairs for special capacities. *Ann. Stat.*, **5**: 909–921.
- (1994): *Robust asymptotic statistics*. Springer Series in Statistics. Springer.
- Rieder H. und Ruckdeschel P. (2000a): Short Proofs on L_r -Differentiability. angenommen von Stat. Decis.
- (2000b): The use of different standardizations in robust statistics. Unveröffentlichtes Manuskript.
- Ringwood J.V. (2000): Shape control systems for Sendzimir steel mills. *IEEE Trans. On Control Systems Technology*, **8**(1): 70–86.
- Rinnooy Kan A.H.G. und Timmer G.T. (1989): Global optimization. In: Nemhauser G., Rinnooy Kan A. und Todd M. (Hg.) *Optimization.*, Bd. 1 von *Handbooks in Operations Research and Management Science*, Kap. IX, S. 631–662. North-Holland.
- Rosenberg B. (1973): A survey of stochastic parameter regression. *Annals of Economic and Social Measurement*, **2**: 381–398.
- Rotea M.A. und Khargonekar P.P. (1995): Generalized $\mathcal{H}_2/\mathcal{H}_\infty$ control. In: Francis B.A. und Khargonekar P.P. (Hg.) *Robust Control Theory*, Bd. 66 von *IMA Volumes in Mathematics and its Applications*, S. 81–104. Springer.
- Rothenberg T.J. (1971): Identification in parametric models. *Econometrica*, **39**: 577–591.
- Ruckdeschel P. (1996): *Zur Asymptotik und Robustheit von Zeitreihen*. Diplomarbeit, Universität Bayreuth, Bayreuth.
- (2000a): Convergence of the $a-A-b$ -algorithm. Unveröffentlichtes Manuskript.
- (2000b): Robust Kalman filtering. In: Härdle W., Hlávka Z. und Klinke S. (Hg.) *XploRe. Application Guide.*, Kap. 18, S. 483–516. Springer.
- Rudin W. (1973): *Functional analysis.*. McGraw-Hill Series in Higher Mathematics. McGraw-Hill.
- (1974): *Real and complex analysis.*. McGraw-Hill Series in Higher Mathematics. McGraw-Hill, 2. Aufl.
- Sage A.P. (1970): On invariant imbedding and nonlinear maximum a posteriori filtering and smoothing. In: *Proc. Symp. nonlin. Estim. Theory Appl., San Diego*, S. 252–255.
- Schick I.C. (1989): *Robust recursive estimation of a discrete-time stochastic linear dynamic system in the presence of heavy-tailed observation noise*. Dissertation, Massachusetts Institute of Technology, Cambridge, MA.

- Schick I.C. und Mitter S.K. (1994): Robust recursive estimation in the presence of heavy-tailed observation noise. *Ann. Stat.*, **22**(2): 1045–1080.
- Schiller F. (1962): Was heißt und zu welchem Ende studiert man Universalgeschichte? Eine akademische Antrittsrede. In: Fricke G. und Göpfert H.G. (Hg.) *Sämtliche Werke*, Abschn.: Theoretische Schriften. Hanser, München. In Verb. m. Stubenrauch, H. aufgr. d. Originaldrucke herausgegeben.
- Schneider W. (1988): Analytical uses of Kalman filtering in econometrics - a survey. *Stat. Hefte*, **29**(1): 3–33.
- Searle S.R. (1997): *Linear models*. Wiley Classics Library. Wiley, Nachdruck der 1971er Aufl.
- Shirazi M.N., Sannomiya N. und Nishikawa Y. (1988): Robust ϵ -contaminated Gaussian filtering of discrete-time linear systems. *Int. J. Control*, **48**(5): 1967–1977.
- Shumway R.H. und Stoffer D.S. (1982): An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Ser. Anal.*, **3**: 253–264.
- Smith A.F.M. und West M. (1983): Monitoring renal transplants: An application of the multiprocess Kalman filter. *Biometrics*, **39**: 867–878.
- Stockinger N. und Dutter R. (1987): Robust time series analysis: A survey. *Kybernetika*, **23**. Supplement.
- Stoer J. (1989): *Numerische Mathematik 1. Eine Einführung – unter Berücksichtigung von Vorlesungen von F. L. Bauer. (Numerical mathematics. I.)*. Springer Lehrbuch. Springer, 5., rev. Aufl.
- Swamy P.A.V.B. (1971): *Statistical inference in random coefficient regression models*, Bd. 55 von *Lecture Notes in Operations Research and Mathematical Systems*. Springer.
- Swensen A.R. (1980): *Asymptotic inference for a class of stochastic processes*. Dissertation, University of California, Berkely, CA.
- Szabłowski P.J. (1990): Elliptically contoured random variables and their application to the extension of the Kalman filter. *Comput. Math. Appl.*, **19**(2): 61–72.
- Tanaka M. und Katayama T. (1987): Robust Kalman filter for linear discrete-time system with gaussian sum noises. *Int. J. Syst. Sci.*, **18**: 1721–1731.
- Tsai C. und Kurz L. (1983): An adaptive robustizing approach to Kalman filtering. *Automatica*, **19**: 279–288.
- Tukey J.W. (1960): A survey of sampling from contaminated distribution. In: Olkin I. (Hg.) *Contributions to Probability and Statistics. Essays in Honor of H. Hotelling.*, S. 448–485. Stanford University Press.
- (1977): *Exploratory data analysis.*, Bd. 7616 von *Behavioral Science: Quantitative Methods.* Addison-Wesley.

- Verdú S. und Poor H.V. (1984): On minimax robustness: A general approach and applications. *IEEE Trans. Inf. Theory*, **IT-300**: 328–340.
- Wald A. (1949): Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.*, **20**: 595–601.
- Wang X. (1994): Local asymptotic normality for multivariate linear processes. *Stochastic Processes Appl.*, **49**(2): 331–345.
- Watson M.W. (1989): Recursive solution methods for dynamic linear rational expectations models. *J. Econ.*, **41**(1): 65–89.
- Wei W.W.S. (1990): *Time series analysis. Univariate and multivariate methods*. Addison-Wesley.
- West M. (1981): Robust sequential approximate Bayesian estimation. *J. R. Stat. Soc., Ser. B*, **43**: 157–166.
- (1984): Outlier models and prior distributions in Bayesian linear regression. *J. R. Stat. Soc., Ser. B*, **46**: 431–439.
- (1985): Generalized linear models: Scale parameters, outlier accomodation and prior distributions. In: *Bayesian statistics 2, Proc. 2nd Int. Meet., Valencia/Spain 1983*, S. 531–558.
- West M. und Harrison J. (1989): *Bayesian forecasting and dynamic models*. Springer Series in Statistics. Springer.
- West M., Harrison J. und Migon H.S. (1985): Dynamic generalized linear models and Bayesian forecasting. *J. Am. Stat. Assoc.*, **80**: 73–97.
- Wiener N. (1948): *Cybernetics or control and communication in the animal and the machine*. The M.I.T. Press and John Wiley & Sons, Inc., New York-London.
- (1949): *Extrapolation, interpolation, and smoothing of stationary time series with engineering applications*. The M.I.T. Press and John Wiley & Sons, New York-London.
- Witting H. (1985): *Mathematische Statistik I: Parametrische Verfahren bei festem Stichprobenumfang*. B. G. Teubner, Stuttgart.
- Wolfowitz J. (1949): On Wald's proof of the consistency of the maximum likelihood estimate. *Ann. Math. Statist.*, **20**: 601–602.
- Wu C.J.J. (1983): On the convergence properties of the EM algorithm. *Ann. Stat.*, **11**: 95–103.
- Yang S. (1990): Efficient robust estimation of parameter in the random censorship model. *Stat. Probab. Lett.*, **10**(5): 419–426.
- (1996): One-step robust parametric estimation with application to random censoring model. *Stat. Probab. Lett.*, **26**(3): 225–232.

Personenregister

- Aggoun, L., 5
 Aldous, D.J., 195
 Anderson, B.D.O., 5, 6, 8, 35, 45, 92,93,
 95, 104, 222
 Anderson, T.W. 254
 Andrews, D.F., ii, ix
 Anscombe, F.J., 23, 275
 Aoki, M., 208, 215
 Arce, G.R., 19
 Athans, M., 4
 Athreya, K.B., 100
 Azmjakov, V. 21
- Barner, K.E. 19
 Başar, T., 16
 Bauer, M., iii, x
 Bellman, R., 254
 Beran, R., 177
 Berndt, E.K., 173
 Bernhard, P., 16
 Bibby, J.M., 279
 Bickel, P.J., ii, ix, 148, 150, 151, 203,
 232, 240
 Billingsley, P., 246
 Birniwal, K., vi, xiii, 18, 20, 114, 116,
 120, 122, 127–129, 231, 240
 Bolzern, P., 20
 Boncelet, C.G. Jr., 19, 36, 38
 Box, G.E.P., iv, xi
 Brand, C., iv, xi, 159
 Breiman, L., 248
 Brockwell, P.J., 203, 213
 Brown, L.D., 148
 Bucy, R., 6, 7
 Burmeister, E. 159
- Caines, P.E., 5, 101, 102, 205, 207,208
 Carlin, B.P., 18
 Carter, C.K., 18
 Çelebi, M.E., 18
 Chan, W-s., 235, 243
 Chen, H.F., 5, 95
 Chen, J., 20
- Chow, H.K., 17
 Chung, K.L. 250
 Cipra, T., 19, 34, 36, 38
 Cooley, T.F., 35
 Colaneri, P., 20
 Collins, J.R., 150, 151, 232, 240
 Coyle, E.J., 19
- Daumer, M., iv, xi
 Davis, R.A., 203, 213
 Deichsel, U., viii
 Deistler, M., 5
 deJong, P., iv, xi, 234, 242
 Dellacherie, C., 250
 Dempster, A.P., 169, 229, 238
 De Nicolao, G., 20
 Dickinson, B.W., 19, 36, 38
 Donoho, D.L., 150, 232, 240
 Doob, J.L., 99
 Duncan, D.B., vi, xii, 34, 46
 Dunford, N., 273
 Dutter, R., 16
- Eagleson, G.K., 195
 Elliott, R.J., 5
 Ershov, A.A., 16–18
- Fahrmeir, L., iii, x, 18, 48, 62, 63, 285
 Falb, P.L., 4
 Falk, M., iv, xi
 Feller, W., 247
 Ferguson, T., 121
 Ferguson, T.S., 235, 243
 Fischer, G., 270
 Fox, A.J., 13
 Franěk, P., 29, 45
 Franke, J., 20, 121
 Fried, R., iii, x
- Gabbouj, M., 19
 Gallagher, N.C. Jr., 19
 Gareil, B., 203, 263
 Gather, U., iii, x

- Gelatt, C., 272
 Genest, C., 235, 243
 Genton, M.G., 224, 235, 243
 Golub, G.H., 115, 286
 Goodrich, R.L., 205, 207
 Graybill, F.A., 255
 Greiber, C., viii
 Grunau, H.C., vii
 Gupta, N.K., 173, 223
 Guttman, I., 17

 Härdle, W., iii, vii, x, 13
 Hájek, J., vi, xiii, 160, 233, 241, 251, 262
 Hall, B.H., 17
 Hall, P., 195
 Hall, R.E., 17
 Hallin, M., 203, 235, 243, 263
 Hampel, F.R., ii, vi, xiii, ix, 37, 112, 228, 237
 Hannan, E.J., 5, 206, 210, 211
 Harrison, P.J., iii, 3, 6, 14, 17
 Hartmann, K., viii
 Hausman, J.A., 173
 Heyde, C.C., 87
 Horn, S.D., vi, xii, 34, 46
 Huber, P.J., ii, ix, 25, 38, 64, 112, 115, 116, 120, 172, 228, 233, 237, 241, 242, 255, 258, 259
 Hürzeler, M., 18

 Imhoff, M., iii, vii, x

 Jazwinski, A.H., 97
 Jeganathan, P., 172, 191, 192, 195, 229, 232, 238, 241
 Jensen, J.L., 6, 203
 Johnson, N.L., 183
 Johnstone, I.M., 87
 Jones, R.H., 235, 244

 Kailath, T., 93
 Kalman, R.E., 6, 7
 Karatzas, I., ii, ix
 Kassam, S.A., 16, 20, 62
 Katayama, T., 17, 18
 Kaufmann, H., 18, 284
 Kent, J.T., 279
 Khargonekar, P.P., 16
 Kim, Y.T., 19
 Kirkpatrick, S., 272
 Kirilin, R.L., 19
 Kitagawa, G., 17

 Kneser, H., 121
 Kohl, M., viii
 Kohn, R., 18
 Kolmogoroff, A., 6
 Kotz, S., 183
 Kreiß, J.P., 204
 Krishnan, T., 169, 170, 177, 185
 Künsch, H.R., 16, 18
 Künstler, R., ii, iii, vii, x, 3, 16, 18, 48, 50, 51, 62, 63
 Kurz, L., 17, 18
 Kushner, H.J., 5

 Laird, N.M., 169, 229, 238
 Le Cam, L., vi, xiii, 160–162, 166, 170, 172, 175, 177, 229, 238, 251
 Ledolter, J., 223
 Levit, B.Y., 148, 232, 240
 Lim, T.L., 20
 Lipster, R.S., 17, 18
 Little, R.J.A., 166, 169, 177, 181, 229, 238
 Loève, M., 11
 Lütkepohl, H., 5, 254

 Ma, Y., 224, 235, 243
 Magnus, J.R., 254
 Mallows, C.L., 150, 232, 240
 Maltz, F., 18
 Marazzi, A., 148, 232, 240
 Mardia, K.V., 279
 Martin, R.D., ii, ix, 14, 16, 20, 25, 26, 115, 128, 129, 154, 155
 Masreliez, C.J., ii, ix, 16, 20, 25, 115, 128, 129, 154
 Maz'ya, V.G., vii, 257
 McGarty, T.P., 114
 McLachlan, G.J., 169, 170, 177, 185
 McLeish, D.L., 198
 Mehra, R.K., 173, 223
 Meinhold, R.J., 17
 Metropolis N., 272
 Meyer, P.A., 250
 Meyn, S.P., 98–102
 Meyr, H., 16
 Migon, H.S., 14, 17
 Millar, P., 121
 Mitter, S.K., ii, ix, 6, 16, 20, 115, 128, 148, 150, 232, 240
 Moghaddamjoo, A., 19
 Moore, J.B., 5, 6, 8, 35, 45, 92, 93, 95, 104, 222
 Morris, J.M., 6, 20

- Müller, M., vii
Müller P.H., 272
- Neudecker, H., 254
Ney, P., 100
Nishikawa, Y., 17
Nollau, V., 272
Nummelin, E., 100
- Orey, S., 100
- Pagan, A., 35, 97, 207, 223
Papantoni-Kazakos, P., 20, 116, 122
Parthasarathy, K.R., xxvi
Patton, R.J., 20
Pecchi, L., iv, xi, 159
Peña, D., 17
Petersen, N.V., 6, 203
Penzer, J., iv, xi, 234, 242
Polovinkin, A.I., 272
Polson, N.G., 18
Poor, H.V., 16, 20, 62, 121
Prescott, E.C., 35
Pupeikis, R., 18
- Raftery, A.E., 16
Rausch, M., 90
Reinsel, G.C., 254
Rieder, H., ii, iii, vi, vii, ix, x, xiii, 21,
35–37, 40, 41, 47, 113, 123, 130,
138, 161–164, 168, 169, 172, 174,
176, 188, 192–196, 198, 199, 226,
245, 246, 248, 251–253, 263, 274,
276
- Ringwood, J.V., iv, xi
Rinnooy Kan, A.H.G., 272
Ritov, Y., 203
Rogers, W.H., ii, ix
Romera, R., 19, 34, 36, 38
Ronchetti, E.M., 37
Rosenberg, B., 35
Rosenbluth, A.W., 272
Rosenbluth, M.N., 272
Rotea, M.A., 16
Rothenberg, T.J., 205, 230, 238
Rousseuw, P.J., 37
Rubin, D.B., 166, 169, 177, 181, 229,
238
Rudin, W., 273
Ryden, T., 203
- Sage, A.P., 18
Sannomiya, N., 17
- Schick, I.C., ii, ix, 6, 16, 18, 20, 97, 98,
115, 128, 129, 148, 150, 232, 240
- Schittkowski, K., 145
Schneider, W., 223, 224
Schwartz, J.T., 273
Searle, S.R., 208, 254
SFB 373, iii, x, 13
SFB 475, iii, x
Shen, J., vi, xiii, 18, 20, 114, 116, 120,
127–129, 231, 240
- Shirazi, M.N., 17
Shreve, S.E., ii, ix
Shumway, R.H., 159, 187, 202, 222, 224
Singpurwalla, N.D., 17
Smith, A.F.M., iv, xi
Spies, G., 16
Stahel, W.A., 37
Stevens, C.F., 3
Stockinger, N., 16
Stoer, 135, 284, 286
Stoffer, D.S., 18, 159, 187, 202, 222,
224
Swamy, P.A.V.B., 35
Swenson, A.R., 204, 263
Szablowski, P.J., 17
- Tanaka, M., 17, 18
Teller, A.H., 272
Teller, E., 272
Tiao, G.C., iv, xi
Timmer, G.T., 272
Tsai, C., 17
Tukey, J.W., ii, ix, 19, 64
Tweedie, R.L. 98
- van Loan, C.F., 115, 285
Verdú, S., 20
Vecchi, M., 272
von Bothmer, H.C., viii
- Wald, A., 172
Wall, K.D., 159
Wang, X., 203
Watson, M.W., 159
Wei, W.W.S. iv, xi, 235, 243
West, M. iv, xi, 3, 6, 14, 17
Wiener, N., 6
Witting, H., 162
Wolfowitz, J., 172
Wu, C.J.J., 170
- Yang, G.I., 162, 166, 175, 177
Yang, S., 175, 177

Yohai, V.J., 14

Zillober, C., vii, 145, 232, 240

Index

- ACM-(type-)Filter, 20, 26, 115
- AR-Modell, 203, 235, 244
 - AR(I)MA, 235, 244
 - CAR, 235, 244
 - CAR(I)MA, 235, 244
- as. relative Effizienz
 - rLS, 23
 - rIC, 39
 - mIC, 53
 - L5, 118
- as. Kovarianz, 251
- as. linearer Schätzer
 - bei Martingaldifferenzen, 199
 - bei unabh. Beobachtungen mit Missings, 168
- as. Minimax-Schranke
 - bei Martingaldifferenzen, 199
 - bei Modellen mit LAN-Eigenschaft, 252
 - bei unabh. Beobachtungen mit Missings, 168
- as. Minimax-Schätzer
 - bei Martingaldifferenzen, 199
 - bei unabh. Beobachtungen mit Missings, 168
- as. Minimax-Theorem, 252
- as. Normalität
 - lokale, 252
 - des MLE, 171
 - der robusten Verfahren, 86
 - von Modellen, 251
- as. suffiziente Statistik, 251
- Ausreißer
 - AO, 13, 21, 22, 64, 114, 115
 - IO, 14, 51, 62, 115
 - Erkennung von, 234, 242
 - Konvex-Kontaminationsumgebung, 15, 114, 129
 - LO, 14, 57
 - PO, 14
 - SO, 14, 68, 114, 116
 - erweitertes SO-Kontaminationsmodell, 129
 - Zusammenhang SO/AO, 115
- batch, 62
- bedingte Verteilung
 - bei Gaußschen Vektoren, 246
 - Normalität der Posterior-Verteilung, 86
- bedingter Erwartungswert
 - als Lösung des MSE-Problems, 7
 - bei Gaußschen Vektoren, 246
- Berechnung
 - durch Samplingtechniken, 18
 - num. Verfahren, 17
- Eichverteilung
 - t -Verteilung, 67
 - normale Konvexkombination, 68
 - Linearität des , 10–13
- Beobachtungsgleichung, 3
- Bias
 - als Robustheitsnebenbedingung im “Lemma 5”-Ansatz, 113
 - Unverzerrtheit der robusten Verfahren, 76
 - Unverzerrtheit des L5, 117
 - Unverzerrtheit des SO-optimalen Filters, 126
 - Schranke an den, im robusten EM-Algorithmus, 176
- box-constraints, 141, 232, 240
- Cayley-Hamilton
 - Satz von, 97
- charakteristische Funktion, 246
- Cholesky-Zerlegung
 - für Block-indizierte Matrizen, 284

- für Block-tridiagonale Matrizen, 285
- Cramér-Lévy
 - Theorem, 247
- Cramér-Rao-Schranke
 - bei Martingaldifferenzen, 199
 - bei unabh. Beobachtungen mit Missings, 168
 - verallgemeinerte, 153, 154
- Diagonalisierbarkeit, 90, 115
- Distributionsableitung, 257
- Dreiecksschema, 163
- Dynamische Lineare Modelle, 3
- elliptisch-symmetrische Verteilungen, 17, 247
- EM-Algorithmus, 51, 159
 - (effizienter) One-Step, 174
 - klassischer
 - Definition, 169
 - Eigenschaften, 170
 - robuster One-Step
 - Definition, 175
 - im normalen Zustandsraummodell, 224
- Ergodizität, 192, 248
- Erreichbarkeit, 101
- erweitertes Kontaminationsmodell, 129
- Faltungssatz, 252
 - bei Martingaldifferenzen, 199
 - bei Modellen mit LAN-Eigenschaft, 252
 - bei unabh. Beobachtungen mit Missings, 168
- fat tails, 16
- Fehlerkovarianzen
 - Bestimmung von, 227, 236
- Filterproblem, 6
- Filtrierung
 - Glättungs-,
 - Definition, 200
 - Zweischritt-, 207
 - Einschritt-, 216, 222
 - LAMN-Eigenschaft für, 190
 - L_2 -Differenzierbarkeit für zeitvariable, 196
 - Wurzeldichtenkalkül für, 188
- Financial Engineering, 235, 244
- Fisher-Information
 - endliche Lokations-, 36
 - generalisierte, 265
 - (multivariate) Lokations-, 255, 261
 - (multivariate) Lokations-/Skalen-, 3, 262
 - (multivariate) Skalen-, 255, 261
 - Regularität im Zustandsraummodell, 208, 216
- Gaussian-Sum-Filter, 17
- Gibbs Sampler, 18
- gleichgradige Integrierbarkeit, 250
 - Satz von Vitali, 162, 250
 - Vallée-Poussin-Kriterium, 162, 250
- Glättungs-Filtrierung *siehe* Filtrierungen
 - Glättungsproblem, 6
- Gross-Error-Modell, 129
- Hard Rejection, 16
- Hidden-Markov-Model, 3, 203
- Hyper-Parameter, 200
 - Definition, 4
 - Schätzung der,
 - L_2 -Differenzierbarkeit des Modells, 202
 - Schätzproblem, 159, 200
 - durch EM-Algorithmus, 222
 - durch (effizienten) One-Step EM-Algorithmus, 223
 - durch robusten EM-Algorithmus, 224
 - Trägheit der Zustandsschätzung durch Mitschätzung, 51, 62
- Identifizierbarkeit (Regularität der Fisher-Information)
 - bei Glättungs-Filtrierungen, 206
 - lokale, zweiter Ordnung, 205
 - Wahl der Parametrisierung, 200
- Influenzkurve
 - bei Martingaldifferenzen, 197
 - bei unabh. Beobachtungen mit Missings, 167
- Hampel-Krasker-Typ, 37, 276
- Korrelationsbedingung, 275
- partielle, 167, 198
- redescender, 50

- Regression
 - m -Schritt, 48
 - klassisch optimale, 36
 - robust-optimale, 37
 - standardisierte
 - Informations-, 40, 41, 57, 280
 - Selbst-, 40, 42,
 - Zentriertheitsbedingung, 275
- Innovationen, 3
- Interventionsanalyse, iv, xi
- ISP, 62, 70, 120, 244
- Kalman-Filter, iv, xi, 7
 - als bedingter Erwartungswert unter (V3), (V4), 10
 - als bestes, lineares Verfahren, 8
 - als orthogonale Projektion, 8, 21
 - als Posterior-Modus unter (V3), (V4), 10
 - als Regressionsschätzer
 - KQ, 34
 - MLE, 35
 - One-Step, 37
 - extended Kalman-Filter, 6
 - Nichtrobustheit des, 13–15
 - Stationarität, 95, 98
 - Trägheit des, 62
- Kalman-Gain, 7
- Kalman-Glätter und -Filter, 8, 222
 - als bedingter Erwartungswert unter (V3), (V4), 48
 - als KQ-Schätzer, 47
 - als One-Step unter (V3), (V4), 48
 - als orthogonale Projektion, 48
 - als Posterior-Modus unter (V3), (V4), 48
 - Stationarität, 96
- Kalman-Vorhersage, 8
- Kleinste-Quadrate-Problem, 9
- Kommutator-Matrix, 254
- Kontrolltheorie, 4, 16, 92
 - stochastische, 101
 - generalisierte Beobachtungsmatrix, 101
 - erweiterte Übergangsmatrix, 101
 - Transferfunktion, 95
 - vollständige Beobachtbarkeit
 - allgem. determ. Fall, 97
 - Zeitinvarianz, 94
 - vollständige Ermittelbarkeit
 - Zeitinvarianz, 94
 - vollständige Erreichbarkeit
 - allgem. determ. Fall, 97
 - Zeitinvarianz, 93
 - vollständige Konstruierbarkeit
 - Zeitinvarianz, 94
 - vollständige Kontrollierbarkeit
 - Zeitinvarianz, 94
 - lokale stochastische, 101
 - schwache stochastische, 101
 - vollständige Stabilisierbarkeit
 - Zeitinvarianz, 94
- Konvergenz
 - schwache, 245
 - vage, 245
- Konvexkontamination, 64
- Konvexkontaminationsumgebungen, 115
- Kroneckerprodukt, 254
- Kumulante
 - vierte, 255
- L-Schätzer, 19
- L_2 -Ableitung
 - bei größeren σ -Algebren, 162
 - des hypothet. feiner filtrierten Modells, 204
 - des normalen Zustandsraummodells, 216
 - Regression
 - m -Schritt, 47
- L_2 -Differenzierbarkeit, 160
 - bei m Typen Missings, 167
 - Definition 161
 - des linearen [multiv.] Regressionsmodells, 36
 - des normalen Zustandsraummodells, 215
 - des Zustandsraummodells, 202
 - für abhängige Beobachtungen, 191
 - für abh. Beob. längs Subfiltrierungsfolgen, 196
 - für größere σ -Algebren, 167
 - für unabh. Dreiecksschemata, 163
 - im m -Stichprobenproblem, 165
 - u.i.v.-Fall, 161

- L5-Filter, 117
- Lagrange-Multiplikator
 - Bestimmung: Lagrange- A
 - Bestimmung für suboptimale IC, 177
 - Bestimmung mit Fixpunkt-Algorithmus, 280
 - Bestimmung mit Newton-Algorithmus, 280
 - Bestimmung: Lagrange- a , 177
 - Existenz 124
- LA(M)N-Eigenschaft
 - LAMN-Eigenschaft, 190
 - LAN-Eigenschaft, 160, 192, 252
 - LA(M)N-Theorem
 - für abh. Beobachtungen, 192
 - für unabh. Dreiecksschemata, 164
- LDM-Zerlegung
 - für Blockmatrizen, 286
- Le Cam
 - drittes Lemma, 251
- “Lemma 5”-Ansatz, 112
 - für SO, 116
 - für AO, 136
- Loglikelihoodentwicklung, 160, 164, 190
- M-Schätzer, 19, 36
- Markov Chain Monte Carlo, 18
- Markov-Eigenschaft, 99
- Markovkern, 99
- Markovkette, 98, 228, 237
 - (as.) homogene, 99
 - absorbierende Menge, 100
 - Definition, 99
 - Feller-, 99, 228
 - Harris-Menge, 100
 - Harris-Rekurrenz, 100
 - invariantes W-Maß, 100
 - Irreduzibilität, 99
 - Periode, 100
 - petite Menge, 100
 - φ -Irreduzibilität, 100
 - positive Harris-Rekurrenz, 100
 - Rekurrenz, 99
 - rIC-Filter, 98
 - rLS-Filter, 98
 - Transienz, 99
- MCAR, 166, 229, 232, 238, 241
- mIC
 - Definition, 49
 - feste Fensterweite, 50
 - Kalibrierung, 62
 - Symmetrie, 77
 - Unverzerrtheit, 77
 - variable Fensterweite, 49
 - Varianten, 57
 - Wahl der Stutzhöhe, 53
- Minimax-Ansatz, 19, 112, 115
 - für SO, 120
 - für AO, 145
- Minimax-Theorem
 - Kneser, 121
- Minimum Prediction Error Methods, 208
- Missing, 166
- ML-Schätzung
 - Fisher-Scoring, 173
 - Gauss-Newton-Verfahren, 173
- MM-Filter, 25
 - optimierter, 26
- MSE-Problem, 7
- MSE-Studie, 233, 242
- Nebenparameter
 - bivariates Skalenmodell, 179, 185
 - Hyper-Parameter als, 223, 224
- On-line monitoring, iii, x
- One-Step-Schätzer, 37, 173
- Optimierungsproblem
 - globales, 272
 - quadratisches unter box-constraints, 141
 - robustes, 19, 22, 38, 112, 176
 - Strafterme
 - Datentreue-, 51
 - Glattheits-, 51
- Particle Filter, 19
- physikalisches Modell, 6
- polnisch, 246
- positiv (semi)definit
 - Lyapunov-, 92
 - Matrizen, 269
- Posterior-Modus, 10, 18, 63
 - Eichverteilung, 51, 63
 - t -Verteilung, 66

- normale Konvexkombination, 66
- im Glättungs- und Filterproblem, 50
- p -Punkt-Umgebungen, 115
- Pratt
 - Satz von, 162
- Prokhorov
 - Satz von, 246
- Quadrik, 266
- Redescender, 50, 145
- Regression
 - Äquivarianz, 36, 48
 - Filtern als, 19, 34
 - Glätten und Filtern als, 46
 - Hebelpunkte, 38
 - M-Schätzer, 36
 - mit stochastischem Parameter, 19, 35
 - Modell, 36
- Regularität von Schätzern, 252
- rIC-Filter
 - Äquivarianz, 39
 - als Markovkette, 98
 - Definition, 38
 - Dichte des Filterfehlers, 81
 - in der Ministudie, 69
 - Kalibrierung, 44
 - Optimalität, 152
 - separat gestutzt, 39
 - Stationarität, 44
 - Symmetrie, 77
 - Unverzerrtheit, 77
 - Wahl der Stutzhöhe, 38
- Risiko Controlling, 235, 244
- rLS-Filter, vi, xii
 - als Markovkette, 98
 - Definition, 22
 - Dichte des Filterfehlers, 78, 79
 - Hybridfilter, 25
 - in der Ministudie, 68
 - Kalibrierung, 32
 - Optimalität, 10
 - Einschritt, 119
 - für $t > 1$, 129
 - Symmetrie, 77
 - Tuning-Parameter, 23–24
 - Unverzerrtheit, 77
- Vergleich mit L5-Filter, 118
- Robustheit
 - bei der Schätzung der Hyperparameter, 224–226
 - bisherige Ansätze im Zustandsraummodell, 15–20
 - im Regressionsmodell, 37
 - im Modell mit Missings, 175, 176
 - “Lemma 5”-Ansatz, 112
 - AO-Lösung, 136–145
 - SO-Lösung, 117
 - rIC, 155
 - Zusammenhang Minimax-Ansatz, 151
- $L_\infty / \mathcal{H}_\infty$, 16
- Minimax-Ansatz, 19, 112
 - AO-Lösung, 145–152
 - Bayes-Lösung, 147
 - SO-Lösung, 120
- Versicherungsprämie, 23
- Rombergintegration, 70
- Sattelpunkt, 115
- Shiftoperator, 248
- Signal-to-Noise-Ratio, 58, 59, 61, 62
- Signatur, 234, 242
- simulated annealing, 272
- Skalenmodell
 - normales, 177
- Splines
 - Interpolation, 70
- Splus, 115
- Stabilität, 90, 92
 - asymptotische, 92
 - exponentiell asymptotisch, 116
 - global asymptotische, 92
 - Lagrange-, 92
 - Lyapunov-, 92
- Stationarität, 90
 - Kalman-Filter, 95, 98
 - Kalman-Glätter, 96
 - schwache, 248
 - strikte, 248
- Steady-State Modell, 15
 - Definition, 15
 - in der Ministudie, 63
 - mIC im, 57
 - M^{rLS} , M^{MM} vs M^{KK} , 29

- Optimalität des rLS für $t > 1$,
128–135
- Parametrisierung der Robustheit,
135
- Stetigkeitssatz, 246
- Steuerungstheorie, 4
- stochastischer Suchalgorithmus, 272
- Straffheit, 245
- svec(\cdot), 254
- Sylvesterscher Trägheitssatz, 90
- Symmetrie, 76

- Tentfunktionen, 140
- Transformed Innovation Distribution Model, 154
- Tuning-Parameter
 - Bisektionsverfahren, 24
 - Integration
 - Monte-Carlo-, 24
 - Romberg-Verfahren, 24
- t -Verteilungen, 65

- Unverzerrtheit, *siehe* Bas76

- vec(\cdot), 254
- Verlustfunktion, 253
- Vitali
 - Satz von, 250
- Vorhersageproblem, 6

- weißes Rauschen, 4
- Wurzeldichten
 - Hilbertraum der, 161
- Wurzeldichtenkalkül
 - für größere σ -Algebren, 162
 - Martingaldifferenzen, 188
 - unabhängiger Fall, 161

- XpLoRe, iii, x, 24, 33, 45, 244, 282

- Zustandsgleichung, 3
- Zustandsraum, 98
 - in der Kontrolltheorie, 93
 - polnischer, 99
- Zustandsraummodell, iv
 - als Regressionsmodell, 35
 - Definition, 3
 - Dynamische Lineare Modelle, 3
 - Gaußsches, 4
 - HMM, 3
 - L_2 -Differenzierbarkeit, 202

- nichtlineares, 6
- robustes, 16
- vereinfachtes Modell, 113