CARL VON OSSIETZKY universität OLDENBURG

Oldenburg Discussion Papers in Economics

The Role of Novelty-Seeking Traits in Contemporary Knowledge Creation

Erkan Gören

V – 402-17

September 2017

Department of Economics
University of Oldenburg, D-26111 Oldenburg

The Role of Novelty-Seeking Traits in Contemporary Knowledge Creation*

Erkan Gören†

Institute of Economics, Carl von Ossietzky University Oldenburg

This Version: September 2017

Abstract

This paper hypothesizes and empirically establishes the persistent effects of novelty-seeking traits on cross-country differences in scientific knowledge creation. I use data on the prevalence of specific allele variants of the human DRD4 exon III gene, which population geneticists have linked to the human phenotype of novelty-seeking behavior to examine its relationship to scientific knowledge creation in society. The results suggest a positive and statistically significant linear relationship between both outcomes that is consistent with the hypothesis that the prevalence of novelty-seeking traits in society facilitates scientific knowledge creation through beneficial human behaviors related to risk-taking and explorative behavior. The empirical findings remain qualitatively unaffected when controlling for additional historical, biogeographical, and socioeconomic factors that appear as additional important determinants in the creation of scientific knowledge in society.

Keywords: DRD4 Exon III, Novelty-Seeking Traits, Entrepreneurial Activity, Knowledge Creation, Technological Progress, Economic Development, Natural Selection, Genetic Diversity

JEL Codes: N70, O30, O50, Z10

---

*I would like to thank Jürgen Bitzer for useful comments and suggestions. All remaining errors are my own.

†Carl von Ossietzky University Oldenburg, School of Computing Science, Business Administration, Economics, and Law (Faculty II), Institute of Economics, Building A5, 26111 Oldenburg, Germany, Tel.: +49-441-798-4292, e-mail: erkan.goeren@uni-oldenburg.de.

1 Introduction

"The spirit of a people, its cultural level, its social structure, the deeds its policy may prepare – all this and more is written in its fiscal history, stripped of all phrases. He who knows how to listen to its message here discerns the thunder of world history more clearly than anywhere else."
—Joseph A. Schumpeter

Models of economic growth that endogenize the pace of technological progress suggest that vertical and horizontal product innovations contribute significantly to the level of economic development (Romer, 1990; Grossman and Helpman, 1991; Aghion and Howitt, 1992). A central result is that differences in income per capita across countries are attributable to differences in the level of technological development. In these models, the rate of technological progress is determined by the knowledge stock in society overall. Several policy implications have been drawn from endogenous growth models to facilitate knowledge creation in the aggregate economy such as the need to protect the intellectual property rights of innovators, to allocate subsidies to research and development activities, and to make investments in human capital. Although these factors appear important for the accumulation of knowledge in society, they refer to proximate causes of economic development and do not reveal in greater detail why countries differ in these observed characteristics. Historical and biogeographic legacies have been identified as important causes of divergent paths in standards of living across countries.[1]

In this paper, I investigate the evolutionary origins of observed differences in the creation of knowledge across countries. The hypothesis that knowledge creation is to some extent determined by deep-rooted historical factors was posited in a recent study in Gören (2017). The author found that the prevalence of novelty-seeking traits – measured by the frequency of the 2- and 7-repeat allele variants of the human dopamine D4 receptor gene – shows a non-monotonic inverted U-shaped relationship with contemporary economic development in a cross-section of countries. This result is suggestive of the potential 'benefits' and 'costs' of novelty-seeking traits for aggregate productivity. The study reported that the negative effects of novelty-seeking traits on economic development are channeled through a higher prevalence of psychological disorders, such as Attention Deficit Hyperactivity Disorder (ADHD), which might result in educational and occupational disadvantages in modern industrialized countries. More importantly, the study attributed positive effects to technological progress resulting from the innate abilities (e.g., risk-taking, explorative, and creative behaviors) of novelty-seeking individuals that foster the accumulation of knowledge in the aggregate economy. However, the findings presented in Gören (2017) are reduced-form estimates and do not reveal in detail the mechanisms by which novelty-seeking traits affect economic development. Therefore, the central hypothesis – that the positive effects of novelty-seeking traits on aggregate productivity manifest themselves through increased knowledge creation – remains the subject of ongoing research. This paper is intended to close this important research gap and to provide additional evidence that the inverted U-shaped relationship identified in Gören (2017) is not contradictory with the proposed mechanism set forth in this paper.

In the theoretical model provided, I show that the rate of innovations is positively linked to the prevalence of

[1] See Spolaore and Wacziarg (2013) for a review of the relevant literature.

novelty-seeking traits in society. This finding is consistent with the notion that innovation activity is stimulated through beneficial economic attitudes related to risk-taking, creativity, explorative, and entrepreneurial behavior. The proposed theory build on the Schumpeterian-inspired endogenous growth model of Aghion and Howitt (1992). I extend the basic model framework to provide a microeconomically based derivation of the innovation process that describes the occurrence of innovations in the research sector as the outcome of two individual random processes regarding the probability of success and the amount of research projects channeled to the research sector, respectively. The model predicts that the rate of innovations that occur in the economy is, among others, a positive function of the prevalence of novelty-seeking traits in society.

I combine various data sources to investigate the relationship between knowledge creation and the prevalence of novelty-seeking traits in society. To measure the extent of knowledge creation in society, I use the number of scientific and technical journal articles per 1,000 people during the period 1981 to 2013 that have been published in the fields of physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences. It is generally acknowledged in the relevant literature that a country's scientific knowledge base plays an important role in the process of innovation by private firms.[2] The key explanatory variable in the empirical analysis refers to the 2- and 7-repeat allele frequencies of the human DRD4 exon III gene that population geneticists have found to be sometimes associated with the human phenotype of novelty-seeking behavior. This measure has been constructed by matching the distribution of ethnic groups in the Alesina et al. (2003) ethnicity data to the population genome data of the human dopamine D4 receptor gene in Gören (2016) using information on the classification of ethno-linguistic groups from the Global Mapping International (2010) database. In addition to the key explanatory variable, I consider a wide range of additional historical, biogeographic, and socioeconomic controls to rule out the possibility that the relationship between scientific knowledge creation and the country-level measure of novelty-seeking traits is prone to the issue of omitted variables bias.

The baseline results indicate a positive and statistically significant linear association between the number of scientific and technical journal publications per 1,000 people and the country-level DRD4 exon III 2- and 7-repeat allele frequency measure. This result is robust to the inclusion of microgeographic, land productivity, climatic, health, legal, religious, historical, and regional factors. The control variables in this list constitute alternative determinants of scientific knowledge creation that might be additionally correlated with the prevalence of novelty-seeking traits in society. Specifically, the inclusion of biogeographic controls rules out concerns that the effect of the country-level DRD4 exon III 2- and 7-repeat allele frequency measure on scientific knowledge creation simply captures the issue of geographic proximity. This possibility is of considerable importance since previous research has suggested that the between-population distribution of DRD4 exon III allele variants is the result of natural selection that has been ongoing since the exodus of the human species out of East Africa (Gören, 2016). The baseline estimate suggests that a one standard deviation increase in the country-level DRD4 exon III 2- and 7-repeat allele frequency measure would, ceteris paribus, raise the number of scientific and technical publications per 1,000 people by about 0.0678

[2]For example, Romer (1990) illustrates that product innovations such as magnetic tapes and home videocassette recorders would not have been possible without academic research in the field of electromagnetism.

journal articles. The magnitude of the estimated regression coefficient is substantial as it corresponds approximately to a 19% increase of the standard deviation of the number of scientific and technical publications per 1,000 people. I conduct a series of additional sensitivity tests to examine the robustness of the main empirical findings to various model specifications. First of all, I show that the main results are robust to the issue of genetic, ethnic, linguistic, and religious diversity. This observation effectively rules out the possibility that the estimated relationship between scientific knowledge creation and the country-level DRD4 exon III 2- and 7-repeat allele frequency measure merely reflects unobserved factors related to various aspects of societal diversity. The results are also robust to the inclusion of various technological distance controls that capture the notion of human barriers to the diffusion of knowledge across countries (i.e., genetic, linguistic, and religious distance from the technological frontier). Moreover, I provide regression coefficients that account for the country's technological and human capability factors, which appear to be of considerable importance for the creation of scientific knowledge in society. Specifically, the baseline estimates are not affected by the inclusion of factors such as national research and development efforts, infrastructure quality, and a human capital index. The estimated positive relationship between scientific knowledge creation and the country-level DRD4 exon III 2- and 7-repeat allele frequency measure is also robust when controlling for social value factors. These factors refer to variables such as openness to trade, international migration, and the protection of civil liberties in society that might reflect latent values related to openness and pluralism in society overall. I use these social capability factors to assess the country's ability to create and adopt new knowledge through beneficial cooperative activities among other approaches. I then examine the sensitivity of the main results to the issue of governance factors. Some factors have been discussed quite extensively in the political economy literature, such as corruption, executive constraints, and intellectual property rights protection in their role in shaping the incentives of economic agents to accumulate human capital. Widespread corruption, for instance, has been shown to foster pervasive rent-seeking activities that would discourage some individuals from engaging in academic research. Nevertheless, the prevalence of novelty-seeking traits continues to have a strong and statistically significant effect on the creation of scientific knowledge in society, which is independent of potential country-specific historical, biogeographic, and socioeconomic factors.

Finally, I use an instrumental variables (IVs) estimation approach to address concerns that the main findings are subject to endogeneity problems. Specifically, it is conceivable that countries with better technological opportunities might have attracted individuals with a higher prevalence of novelty-seeking traits at earlier points in their history, thus resulting in the positive correlation between scientific knowledge creation and the country-level DRD4 exon III 2- and 7-repeat allele frequency measure. I use a set of excluded biogeographic instruments that have been linked empirically to between-population variation of DRD4 exon III allele variants (Gören, 2016). Specifically, it can be established that prehistoric migratory distance from East Africa satisfies the necessary exclusion restriction in a standard IVs estimation setting. The argumentation is that migratory distance from East Africa has no direct impact on the extent of scientific knowledge creation other than through its influence on the natural selection of specific DRD4 exon III allele variants linked to the human phenotype of novelty-seeking behavior. The econometric results remain qualitatively unaffected by the proposed IVs estimation methodology. Standard statistical tests confirm that the IVs estimation approach satisfies two important conditions linked to the relevance and validity of the set of excluded instruments.

This paper contributes to the literature on the evolutionary origins of cross-country differences in standards of living. The importance of novelty-seeking traits for the rate of technological progress has been analyzed theoretically in Galor and Michalopoulos (2012). The authors argue that the kind of human behaviors frequently ascribed to novelty-seeking individuals (e.g., entrepreneurial activity and risk-taking) are the main drivers of innovation activity in the process of economic development. Based on a novel compilation of country-level DRD4 exon III allele frequencies, the study in Gören (2017) provides the first evidence of a statistically significant, inverted U-shaped association between the prevalence of novelty-seeking traits in society and comparative economic development. This observation is suggestive of the potential 'benefits' (e.g., knowledge creation) and 'costs' (e.g., educational and occupational disadvantages) experienced by novelty-seeking individuals in the aggregate economy. Other studies have documented the role of additional specific genetic markers in pre-colonial economic development, contemporary life expectancy, cultural attitudes, and innovations (Cook, 2014, 2015; Gorodnichenko and Roland, 2017). In addition, Ashraf and Galor (2013) examine the non-monotonic influence of overall human genetic diversity on pre-colonial and contemporary differences in income per capita in a cross-section of countries. Spolaore and Wacziarg (2009) and Spolaore and Wacziarg (2012) found that a country's relative genetic distance from the technological frontier contributed significantly to differences in economic development and to the diffusion of technological innovations across borders. This finding is in line with the notion that genetic distance captures deep-rooted historical and cultural differences among populations that hinder the diffusion of technological and institutional improvements across countries.

The paper is organized as follows. Section 2 presents an extended version of the basic Schumpeterian economic growth model to highlight the importance of novelty-seeking traits in the process of scientific knowledge creation. Section 3 provides a detailed description of the main variables employed in the empirical analysis. Section 4 discusses the econometric framework employed in the analysis of the relationship between scientific knowledge creation and the county-level DRD4 exon III 2- and 7-repeat allele frequency measure. The main empirical results are discussed in Section 5. Additional sensitivity tests are reported in Section 6. Finally, Section 7 concludes by summarizing the main findings.

2 The Model

In this section, I develop an endogenous economic growth model to conceptualize the main hypothesis proposed in this paper regarding the beneficial effects of novelty-seeking individuals in the process of knowledge creation in the aggregate economy. The underlying theoretical framework is the Schumpeterian-inspired model of economic growth proposed in Aghion and Howitt (1992), where growth is generated through repeated improvements in the quality of intermediate goods. This model framework is particularly convenient for studying the implications of novelty-seeking individuals in the process of economic growth. The reason is that the type of attributes related to the Schumpeterian notion of 'creative destruction' – risk-taking, creativity, and entrepreneurship – are those frequently attributed to the kind of human behaviors usually found in novelty-seeking individuals.

The underlying source of technological progress in this model of economic growth is the research sector, where the flow of workers channeled to R&D activities influence the rate of innovations that occur in the economy. It is worth mentioning that in this model framework, the arrival rate per researcher, that is, the rate at which innovations occur randomly in the aggregate economy, is exogenously given. However, this assumption might be misleading given that research productivity differs substantially across countries. Specifically, Figure 1 demonstrates the notable differences that exist in contemporary research productivity across countries, as indicated by the number of scientific and technical publications per 1,000 people in the years 1981 to 2013. The corresponding summary statistics are reported in Table 9. The mean value of scientific and technical publications per 1,000 people is about 0.1831 with a relatively large standard deviation of 0.3259 across the available set of 171 countries. As expected, research productivity is the highest among western European and Neo-European (i.e., Australia, Canada, New Zealand, and the United States) countries, whereas the lowest number of scientific and technical publications are found in African countries, which belong to one of the poorest regions in the world.

Given the fact that research productivity differs substantially across countries and time, I extend the basic model framework in Aghion and Howitt (1992) by providing a microeconomically based derivation of the source of technological progress in society. To accomplish this task, I define innovation as the outcome of two individual random processes. In the first process, the probability of successful innovations occurring out of a sequence of randomly distributed research projects has a binomial distribution. I endogenize the corresponding probability of success as a function of the extent of novelty-seeking traits in society. The second source of uncertainty refers to the number of research projects that occur randomly within the unit time interval. This circumstance is modelled according to a Poisson distribution with an endogenous arrival rate, which is a function of the country's technological and human resource base. Then, the innovation process in the aggregate economy is fully determined by the corresponding individual random processes, which follows a Poisson distribution, with a corresponding arrival rate that is a function of the prevalence of novelty-seeking traits in society. The theoretical model suggests that a higher prevalence of novelty-seeking traits in society, ceteris paribus, increases the extent of knowledge creation through a higher propensity of research projects toward success.

2.1 Theoretical Framework

Final Goods Sector. The economy produces a final product y using a single intermediate good m subject to the following Cobb-Douglas production technology

$$y_t = A_t m_t^{\alpha}, \tag{1}$$

where y is output, A the level of technology, m the amount of intermediate goods, and $0 < \alpha < 1$ indicates the output elasticity with respect to intermediate goods. For the sake of simplicity, I leave out population growth, human, and physical capital accumulation in the process of economic growth to keep the analysis as simple as possible. As usual in Schumpeterian growth models, the subscript $t = \{0, 1, 2, ...\}$ does not refer to continuous time but rather to the start of the $t-th$ innovation in the aggregate economy. The main idea behind this model framework is that the economy still uses the highest quality of intermediate goods in the production process until it is rendered obsolete by the invention of a new technology, i.e., $A_{t+1} > A_t$.

Intermediate Goods Sector. Each individual is endowed with one unit of labor that she can allocate freely between the intermediate goods and research sector. The intermediate good is produced with a linear technology that employs labor as the only production factor

$$m_t = \ell_t^M, \tag{2}$$

where ℓ_t^M denotes the flow of labor in the intermediate goods sector of the $t-th$ innovating firm. It is assumed that the intermediate goods sector is characterized by monopolistic competition. An innovating firm gets a temporal patent to earn profits in the intermediate goods sector until it is replaced by outside firms that conduct research. The duration of the patent is indefinite due to uncertainty regarding the discovery of new inventions in the research sector.

Research Sector. Growth in this model is generated from successful innovations in the research sector, which improves the quality of intermediate goods up to a productivity factor $\gamma > 1$ according to

$$A_t = A_0\gamma^t, \tag{3}$$

where A_0 refers to the initial level of technology, which might be a function of historical and biogeographic factors. In the following, I provide a microeconomically based derivation of the innovation process that models the research output as the result of a two individual random processes. This model framework proves useful to highlight the importance of the prevalence of novelty-seeking traits for the process of innovation activity and to disentangle its impact from other historical, socioeconomic, and technological factors.

Indicate with τ the continuous time variable which is of unity interval (i.e., $\Delta\tau = 1$). Denote with T_i the $i-th$ research project in the aggregate economy which can take two possible states. These two states refer to the outcome of the innovation process, which may result in a successful new invention ($T_i = 1$) or may not ($T_i = 0$). In probability theory, the Bernoulli distribution is used quite frequently to model random processes that take two possible outcomes. According to this distribution, the probability mass function of the random variable T_i with outcomes $k = \{0, 1\}$ is

$$f_{T_i}(T_i = k \mid p) = \begin{cases} p & \text{if} \quad k = 1 \quad (\Leftrightarrow \text{successful innovation}), \\ 1-p & \text{if} \quad k = 0 \quad (\Leftrightarrow \text{unsuccessful innovation}), \end{cases} \tag{4}$$

where $p \in [0, 1]$ refers to the probability of success if the $i-th$ research project results in a new invention (i.e., $T_i = 1$) and $1-p$ if not (i.e., $T_i = 0$). A main finding from endogenous growth models is that differences in income per capita across countries are attributable to differences in the level of technology, as indicated in Equation (3). The global pattern of technological performance suggests that countries differ substantially in levels of innovation activity. A possible approach to modelling such technological differences across countries is to endogenize the probability of success in the Bernoulli distribution that describes the outcome of individual research projects. Specifically, it is desirable to assume that the success probability of research projects is a function of the prevalence of novelty-seeking traits in society. The argumentation is that novelty-seeking individuals are quite successful in the creation of new knowledge or innovations due to their innate explorative, risk-taking, creative, and entrepreneurial behavior, which pushes the technological frontier forward (Galor and Michalopoulos, 2012; Gören, 2017). In particular, the probability of successful innovations is modelled according to $p = p(z, \nu)$, where ν is the prevalence of novelty-seeking traits in society. The parameter z refers to additional historical, biogeographic, technological, and human capability factors that characterize the economic environment, where individuals engage in cooperative social behavior. The specific mechanisms by which these factors influence the innovation process are discussed in greater detail below. However, it is worth mentioning that the narrative of this approach is related to the basic idea that historical, geographic, and socioeconomic circumstances might act as barriers to the diffusion of technological improvements across borders that in turn negatively affect a country's own innovation process (Basu and Weil, 1998).

Regarding the functional form of $p(z, \nu)$, I assume that $p(\cdot)$ is monotonically increasing and of diminishing marginal returns with respect to z and ν, i.e., $\partial p(z, \nu)/\partial \varphi \equiv p_{\varphi}(z, \nu) > 0$, $\partial^2 p(z, \nu)/\partial \varphi^2 \equiv p_{\varphi\varphi}(z, \nu) < 0$, and $\partial^2 p(z, \nu)/\partial \varphi \partial \varphi' \equiv p_{\varphi\varphi'}(z, \nu) = p_{\varphi'\varphi}(z, \nu) > 0$, where $\varphi, \varphi' = \{z, \nu\}$. The positive sign of the second cross partial derivatives $p_{z\nu}(z, \nu)$ and $p_{\nu z}(z, \nu)$ indicates possible complementarities between the prevalence of novelty-seeking traits and socioeconomic factors. The idea is that the benefits of novelty-seeking traits for the creation of knowledge in society might be dependent on prevailing institutional and economic circumstances that make it easier to reap gains from explorative behavior. For example, the innate impulsive nature of novelty-seeking individuals might result in educational and occupational difficulties in modern societies that are characterized by clear social hierarchies (Mannuzza and Klein, 2000). Thus, countries that are unable to develop effective educational and labor market strategies to mitigate the potential costs of novelty-seeking traits would find it hard to benefit from their explorative nature for the aggregate economy (Gören, 2017).

Usually, economies are engaged in more than one research project per unit time interval. A sequence $T = \sum_{i=1}^{N} T_i$ of N identically independently distributed ($i.i.d.$) random research projects T_i with identical individual success probabilities $p(z, \nu)$ has a binomial distribution with parameters $N \in \mathbb{N}$ and $p(z, \nu) \in [0, 1]$ with a corresponding probability mass function

$$f_T\left[T = t \mid N, p(z, \nu)\right] = \binom{N}{t} p(z, \nu)^t \left[1 - p(z, \nu)\right]^{(N-t)}, \tag{5}$$

where $t = \{0, 1, 2, ..., N\}$ is the number of successful innovations during the unit time interval. This formulation illustrates the uncertainty of the innovation process, indicating that of N random research projects, one could expect on average $\mathbb{E}\left[\sum_{i=1}^{N} T_i\right] = N \times p(z, \nu)$ successful innovations with variance $\mathbb{V}\left[\sum_{i=1}^{N} T_i\right] = N \times p(z, \nu)\left[1 - p(z, \nu)\right]$.

Next, consider another source of technological heterogeneity: It is conceivable that N, indicating the number of research projects in the unit time interval, differs across countries too. In particular, differences in the flow of researchers (ℓ_t^R), the amount of government resources channeled to R&D activities (R), and the economic environment (ψ) might explain widespread variations in the extent of research activity N across countries. In this paper, I model the number of research projects per unit time interval using a Poisson distribution. In probability theory, the Poisson distribution expresses the probability of randomly occurring events (here: number of research projects N) in a fixed time interval (here: $\Delta\tau = 1$) according to the following probability density function

$$f_N\left[N = n \mid \mu\left(\psi, \ell_t^R, R\right)\right] = \frac{\left[\mu\left(\psi, \ell_t^R, R\right)\right]^n}{n!} e^{-\mu\left(\psi, \ell_t^R, R\right)}, \tag{6}$$

with corresponding arrival rate $\mu\left(\psi, \ell_t^R, R\right) \equiv (1-\psi)\phi(\ell_t^R, R) \in \mathbb{R}_{>0}$. The arrival rate indicates the expected number of research projects per unit time interval, i.e., $\mathbb{E}\left[N\right] = (1-\psi)\phi(\ell_t^R, R)$ with variance $\mathbb{V}\left[N\right] = (1-\psi)\phi(\ell_t^R, R)$, which is identical to its expected value. The constant parameter $\psi \in [0,1]$ captures potential market distortions in the research sector (e.g., widespread corruption, rent-seeking activities, or growth-distorting government taxes) that undermine its productivity. The economy generates research projects according to the research production technology $\phi(\ell_t^R, R)$, which is subject to constant returns to scale, where $\partial\phi(\ell_t^R, R)/\partial\varphi \equiv \phi_\varphi(\ell_t^R, R) > 0$, $\partial^2\phi(\ell_t^R, R)/\partial\varphi^2 \equiv \phi_{\varphi\varphi}(\ell_t^R, R) \leq 0$, and $\partial^2\phi(\ell_t^R, R)/\partial\varphi\partial\varphi' \equiv \phi_{\varphi\varphi'}(\ell_t^R, R) = \phi_{\varphi'\varphi}(\ell_t^R, R) > 0$ with $\varphi, \varphi' = \{\ell_t^R, R\}$. The innovation process in the research sector is then determined by the corresponding probability functions of the individual random processes, as described in Equations (5) and (6), according to the following composite probability density function

$$f_T\left[T = t \mid \lambda\left(\psi, \ell_t^R, R, z, \nu\right)\right] = \sum_{n=0}^{\infty} f_T\left[T = t \mid N = n, p(z, \nu)\right] \times f_N\left[N = n \mid \mu\left(\psi, \ell_t^R, R\right)\right]. \tag{7}$$

This expression corresponds to the mixture of two individual probability distributions to a weighted probability distribution model for the occurrence in the number of *successful innovations* $t = \{0, 1, 2, ...\}$ per unit time interval in the research sector. In particular, the binomial distribution in Equation (5) with the randomly distributed parameter N is weighted over all possible realizations $n = \{0, 1, 2, ...\}$ according to the probability density function of the Poisson distribution in Equation (6) with arrival rate $\mu\left(\psi, \ell_t^R, R\right)$ with a corresponding solution given by

$$f_T\left[T = t \mid \lambda\left(\psi, \ell_t^R, R, z, \nu\right)\right] = \frac{\left[\lambda\left(\psi, \ell_t^R, R, z, \nu\right)\right]^t}{t!} e^{-\lambda\left(\psi, \ell_t^R, R, z, \nu\right)}, \tag{8}$$

where $\lambda\left(\psi, \ell_t^R, R, z, \nu\right) \equiv (1-\psi)\phi(\ell_t^R, R)p(z, \nu) \in \mathbb{R}_{>0}$. In summary, the discretely occurring events that comprise the number of successful innovations per unit time interval follow a Poisson distribution with arrival rate $\lambda\left(\psi, \ell_t^R, R, z, \nu\right)$ which, among other things, is a positive function of the prevalence of novelty-seeking traits in society ν. However, the steady-state arrival rate of the number of innovations depends on the flow of labor ℓ_t^R channeled into the research sector. This allocation process is governed by the corresponding profit considerations of economic agents in the intermediate goods and research sector, which I discuss in more detail below.

2.2 The Monopolist's Profit Maximization Problem

The intermediate goods sector is characterized by monopolistic competition. Intermediate goods are produced according to the linear technology $m_t = \ell_t^M$. The monopolist firm stays in the intermediate goods market until it is replaced by the next innovating firm. Since innovations in the economy per unit time interval occur randomly according to a Poisson distribution with arrival rate $\lambda\left(\psi, \ell_t^R, R, z, \nu\right)$, the length of time between two consecutive innovations, i.e., $\Delta(t+1,t)=\tau_{t+1}-\tau_t$, is exponentially distributed with parameter $\lambda\left(\psi,\ell_t^R,R,z,\nu\right)$, where the expected length of time between two innovations is inversely related to the Poisson arrival rate, i.e., $\mathbb{E}\left[\Delta(t+1,t)=\tau_{t+1}-\tau_t\right]=1/\left[\lambda\left(\psi,\ell_t^R,R,z,\nu\right)\right]$ with variance $\mathbb{V}\left[\Delta(t+1,t)=\tau_{t+1}-\tau_t\right]=1/\left[\lambda\left(\psi,\ell_t^R,R,z,\nu\right)\right]^2$. Thus, the length of the interval in which the monopolist firm can earn profits or remain in the intermediate goods market is negatively related to the arrival rate of innovations in the research sector.

The monopolist firm is faced with the inverse demand function in the final goods sector $p_t=\alpha A_t m_t^{\alpha-1}$ and takes as given the wage rate w_t of skilled labor. It therefore maximizes its flow of profits according to the following profit function

$$\max_{m_t}\pi_t=\left[p_t m_t-w_t m_t\right], \tag{9}$$

where π_t refers to profits of the $t-th$ innovating firm. Substituting the equation for the inverse demand function of the intermediate good in the final goods sector into the monopolist profit function and maximizing with respect to m_t yields the monopolist firm's labor demand function in the intermediate goods sector

$$\ell_t^M\equiv\widetilde{\ell}^M(\omega_t)=\left(\frac{\alpha^2}{\omega_t}\right)^{\frac{1}{(1-\alpha)}}\quad\Leftrightarrow\quad\omega_t\equiv\widetilde{\omega}(\ell_t^M)=\frac{\alpha^2}{\left(\ell_t^M\right)^{(1-\alpha)}}, \tag{10}$$

where $\omega_t=w_t/A_t$ is the productivity-adjusted wage rate. Thus, the monopolist firm's demand for skilled labor in the intermediate goods sector is inversely related to the productivity-adjusted wage rate. Given Equation (10), the monopolist profit flow is given by

$$\pi_t=A_t\widetilde{\pi}(\omega_t), \tag{11}$$

where $\widetilde{\pi}(\omega_t)=\left(\frac{1-\alpha}{\alpha}\right)\left(\frac{\alpha^2}{\omega_t^{\alpha}}\right)^{\frac{1}{(1-\alpha)}}$. This equation expresses the monopolist's profits as a negative function of the productivity-adjusted wage rate ω_t and as a positive function of the level of technology A_t.

2.3 Perfect Competition in the Research Sector

Research is conducted by outside firms that employ skilled labor in a patent-race competitive framework to earn the expected present value of future profit flows in the intermediate goods sector of the next $(t+1)-th$ innovation, denoted by V_{t+1}, according to the following expression

$$V_{t+1}=\int_0^{\infty}\widetilde{V}_{t+1}(\tau)f_{\Delta}\left[\Delta(t+2,t+1)=\tau\middle|\lambda\left(\psi,\ell_{t+1}^R,R,z,\nu\right)\right]d\tau, \tag{12}$$

where $f_{\Delta}\left[\Delta(t+2,t+1)=\tau\middle|\lambda\left(\psi,\ell_{t+1}^R,R,z,\nu\right)\right]=\lambda\left(\psi,\ell_{t+1}^R,R,z,\nu\right)e^{-\lambda\left(\psi,\ell_{t+1}^R,R,z,\nu\right)\tau}$ is the probability density function of the length of time $\Delta(t+2,t+1)=\tau_{t+2}-\tau_{t+1}\equiv\tau$ between two consecutive innovations, which has an exponential distribution with parameter $\lambda\left(\psi,\ell_{t+1}^R,R,z,\nu\right)\in\mathbb{R}_{>0}$. It is worth mentioning that the $(t+1)-th$ innovating firm will face uncertainty over its future profit flow because of research conducted by outside firms during the $(t+1)-th$ innovation interval. The present value of the future profit flow of the $(t+1)-th$ innovating firm until it is replaced by the next (t+2)-*th* innovator, denoted by $\widetilde{V}_{t+1}(\tau)$, is given by

$$\widetilde{V}_{t+1}(\tau)=\int_{\tau_{t+1}}^{\tau_{t+2}}\pi_{t+1}e^{-r(v-\tau_{t+1})}dv, \tag{13}$$

where the accent ($\sim$) placed over V indicates that the present value of the future monopolist profit flow is a random variable, because of the uncertainty regarding the length of time between two consecutive innovations, and r is the constant real interest rate. Thus, the solution to Equation (13) is given by

$$\widetilde{V}_{t+1}(\tau) = \frac{\pi_{t+1}\left[1 - e^{-r\tau}\right]}{r}. \tag{14}$$

If $\tau = (\tau_{t+2} - \tau_{t+1})$, the length of time of the $(t+1) - th$ innovation, approaches infinity in the limit, then the monopolist expected future profit flow would correspond to its value of perpetuity, i.e., π_{t+1}/r.

Substituting this expression for $\widetilde{V}_{t+1}(\tau)$ in Equation (12) and solving the integral with respect to τ gives the following solution for the monopolist firm's expected present value of future profits in the intermediate goods sector

$$V_{t+1} = \frac{\pi_{t+1}}{r + \lambda\left(\psi, \ell_{t+1}^{R}, R, z, \nu\right)}, \tag{15}$$

which is similar to the expression in Aghion and Howitt (1992) except for the fact that the Poisson arrival rate $\lambda\left(\psi, \ell_{t+1}^{R}, R, z, \nu\right)$ of successful innovations in this model setup has been derived from two individual random processes in the research sector.

After having defined the value that the outside firm can earn in the intermediate goods sector when it produces a successful innovation, we are now in the position to describe the maximization problem of the research-conducting firm that determines its research employment in the innovation process. Without loss of generality, I assume an infinitesimally small time interval $d\tau$ during the $t - th$ and $(t+1) - th$ innovations. During this time interval, the probability that the research-conducting outside firm will produce an innovation allowing it to take over the entire intermediate goods sector is approximately $\lambda\left(\psi, \ell_{t}^{R}, R, z, \nu\right) d\tau$.[3] Then, the firm's expected profit from conducting research equals $\lambda\left(\psi, \ell_{t}^{R}, R, z, \nu\right) d\tau V_{t+1}$ with corresponding costs due to research employment in the amount of $w_t d\tau \ell_t^R$ during the small time interval $d\tau$. Specifically, the research-conducting outside firm maximizes the following profit function with respect to research employment ℓ_t^R

$$\max_{\ell_t^R} \left[\lambda\left(\psi, \ell_{t}^{R}, R, z, \nu\right) d\tau V_{t+1} - w_t d\tau \ell_t^R\right]. \tag{16}$$

The solution to this maximization problem yields the famous research arbitrage equation in the Aghion and Howitt (1992) Schumpeterian growth model

$$w_t = \lambda_{\ell_t^R}\left(\psi, \ell_{t}^{R}, R, z, \nu\right) V_{t+1}, \tag{17}$$

where $\lambda_{\ell_t^R}\left(\psi, \ell_{t}^{R}, R, z, \nu\right) \equiv \partial\lambda\left(\psi, \ell_{t}^{R}, R, z, \nu\right)/\partial\ell_t^R$ refers to the first partial derivative of the Poisson arrival rate with respect to research employment. The research arbitrage equation states that outside firms will employ skilled labor in the research sector up to the point where their marginal cost of research w_t equals their expected marginal benefit of research $\lambda_{\ell_t^R}\left(\psi, \ell_{t}^{R}, R, z, \nu\right) V_{t+1}$.

[3] Strictly speaking, the probability that an outside firm will produce exactly one innovation during the length of time $d\tau$ is, according to the Poisson distribution in Equation (8), given by $f_T\left[T = 1 \mid \lambda\left(\psi, \ell_{t}^{R}, R, z, \nu\right)\right] = \lambda\left(\psi, \ell_{t}^{R}, R, z, \nu\right) d\tau e^{-\lambda\left(\psi, \ell_{t}^{R}, R, z, \nu\right) d\tau}$. Dividing both sides by $d\tau$ and taking the limit $d\tau \to 0$ yields $\lim\limits_{d\tau \to 0}\left(\frac{1}{d\tau} f_T\left[T = 1 \mid \lambda\left(\psi, \ell_{t}^{R}, R, z, \nu\right)\right]\right) = \lambda\left(\psi, \ell_{t}^{R}, R, z, \nu\right)$, which approximately equals the probability of one innovation during an infinitesimally small time duration.

2.4 Equilibrium in the Decentralized Economy

The research arbitrage equation shown in (17) together with the labor market clearing condition $L = \ell_t^R + \widetilde{\ell}^M(\omega_t)$ entirely describes the equilibrium condition in the decentralized economy. In equilibrium, skilled workers in the intermediate goods and research sector must be paid the same wage rate, which is determined by the monopolist inverse labor demand function $\widetilde{\omega}(\ell_t^M)$ in Equation (10). Using Equations (10), (14), (15), the labor market clearing condition $\ell_t^M = L - \ell_t^R$, and the fact that $A_{t+1} = \gamma A_t$, it follows for the research arbitrage equation after some basic rearrangements

$$\frac{\widetilde{\omega}(L-\ell_t^R)}{\lambda_{\ell_t^R}\left(\psi, \ell_t^R, R, z, \nu\right)} = \frac{\gamma \widetilde{\pi}\left[\widetilde{\omega}(L-\ell_{t+1}^R)\right]}{r + \lambda\left(\psi, \ell_{t+1}^R, R, z, \nu\right)}. \tag{18}$$

Notice that research employment during the current $t-th$ innovation period is a decreasing function of research employment in the $(t+1)-th$ innovation period in the future. The reason is that firms are discouraged from investing more in research today by an anticipated decrease in the flow of future profits due to a higher productivity-adjusted wage rate and a shorter length of time between two consecutive innovations, which is triggered by an increase in the Poisson arrival rate of future innovations (Aghion and Howitt, 1992).

In order to examine the model dynamics in greater detail, I make use of the following constant returns to scale function for the research production technology $\phi(\ell_t^R, R) = \ell_t^R R$, where $R > 0$ indicates a research-specific productivity parameter.[4] Substituting this expression for $\phi(\ell_t^R, R)$ in Equation (18) and defining the condition of a stationary equilibrium in which the flow of research employment is constant across innovation intervals, i.e., $\ell_t^R = \ell_{t+1}^R \equiv \widehat{\ell}^R$, yields the following steady-state solution of research employment in the decentralized economy (DE)

$$\widehat{\ell}_{DE}^R = \frac{(1-\psi)Rp(z,\nu)\gamma\left(\frac{1-\alpha}{\alpha}\right)L - r}{(1-\psi)Rp(z,\nu)\left[1+\gamma\left(\frac{1-\alpha}{\alpha}\right)\right]}, \tag{19}$$

Given the solution for the steady-state flow of research employment, the Poisson arrival rate of successful innovations in the decentralized economy is given by

$$\widehat{\lambda}_{DE}\left(\psi, \widehat{\ell}_{DE}^R, R, z, \nu\right) = \frac{(1-\psi)Rp(z,\nu)\gamma\left(\frac{1-\alpha}{\alpha}\right)L - r}{\left[1+\gamma\left(\frac{1-\alpha}{\alpha}\right)\right]}. \tag{20}$$

This expression is strictly positive if the condition $(1-\psi)Rp(z,\nu)\gamma\left(\frac{1-\alpha}{\alpha}\right)L/r > 1$ holds. Standard comparative static analysis reveals the following central result with respect to the prevalence of novelty-seeking traits in society

$$\frac{\partial \widehat{\lambda}_{DE}\left(\psi, \widehat{\ell}_{DE}^R, R, z, \nu\right)}{\partial \nu} = \frac{(1-\psi)R\frac{\partial p(z,\nu)}{\partial \nu}\gamma\left(\frac{1-\alpha}{\alpha}\right)L}{\left[1+\gamma\left(\frac{1-\alpha}{\alpha}\right)\right]} > 0, \tag{21}$$

which states that countries with a higher prevalence of novelty-seeking traits in society would have, ceteris paribus, a higher expected number of innovations in the research sector. In order to investigate this hypothesis empirically, I use data on scientific and technical publications as the main dependent variable in a large cross-section of countries. This is a partial analysis framework, in which scientific publications function as a proxy variable for the knowledge base from which innovations are generated in society. Hence, a positive and statistically significant association

[4] Another possibility would be to model the arrival rate to follow a sigmoid function (e.g., the cumulative distribution function of the logistic function) to account for threshold effects in the research sector (Aghion and Howitt, 1998).

between the number of scientific publications and the country-level measure of novelty-seeking traits would be consistent with the hypothesis of the beneficial effects of such traits in the innovation process and finally economic development.

3 Data and Variables

In this section, I provide a detailed discussion of the set of variables employed in the empirical analysis. First of all, I introduce the main dependent variable that indicates the extent of knowledge creation or innovation activity in the research sector. The main explanatory variable refers to a recent compilation of DRD4 exon III allele frequencies that population geneticists have found to be sometimes associated with the human phenotype of novelty-seeking behavior. Furthermore, this section presents a variety of biogeographic, climatic, socioeconomic, historical, and cultural variables that have attracted considerable attention in the development economics literature that appear to be equally important for both the extent of knowledge creation in society and the distribution of novelty-seeking traits across countries.

3.1 Main Dependent Variable: Number of Scientific and Technical Journal Articles

An empirical investigation of the determinants of knowledge creation requires a proper definition of research and development (R&D) activities allowing them to be identified and measured in the aggregate economy. The OECD's Frascati Manual provides a guideline for the measurement of scientific, technological, and innovation activities. According to this guideline, R&D activities are defined as "[...] creative and systematic work undertaken in order to increase the stock of knowledge – including knowledge of humankind, culture and society – and to devise new applications of available knowledge" (OECD, 2015, pp. 44–45). This definition covers three types of R&D activities – basic research, applied research, and experimental development – depending on their applicability of achieving specific goals (e.g., creation of basic knowledge, products, or processes). The output of the research activities is usually published in scientific or technical journal articles. Therefore, the main dependent variable employed here throughout the empirical analysis refers to the number of scientific and technical journal articles per 1,000 people that have been published in the fields of physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences. I use this variable to measure the level of knowledge creation in society.

3.2 Main Explanatory Variable: DRD4 Exon III 2- and 7-Repeat Allele Frequency

The main explanatory variable indicating the prevalence of novelty-seeking behavior in society refers to the frequency of a set of specific allele variants of the human dopamine D4 receptor (hereafter referred to as DRD4) gene. This particular gene has been the subject of increasing attention in molecular genetic research given its importance in the human dopaminergic region for reward, cognition, and emotion (Oak et al., 2000). The human DRD4 gene consists mainly of four encoded regions, called exons by population geneticists, of which the third exon exhibits the most extensive polymorphism in allelic variants. This polymorphism occurs as a variable number of tandem repeats that range from 2- to 11-repeats, resulting in a total number of ten different allele variants (Van Tol et al., 1991, 1992; Lichter et al., 1993).[5]

It is worth mentioning that especially the 2- and 7-repeat allele variants of the human DRD4 exon III gene show a blunted response to elevated dopamine levels in comparison to the ancestral 4-repeat allele variant (Asghari et al., 1995; Wang et al., 2004). It has been suggested that this difference in neurobiological functioning of the 2- and 7-repeat allele variants might result in human personality outcomes related to important economic attitudes such as novelty-seeking, risk-taking, entrepreneurial activity, and creativity (Benjamin et al., 1996; Dreber et al., 2009; Galor and Michalopoulos, 2012; Mayseless et al., 2013). Thus, differences in the frequency of the DRD4 exon III 2- and 7-repeat allele variants across countries might indicate differences in the extent of these beneficial economic attitudes in society. The study in Gören (2017) provides DRD4 exon III 2- and 7-repeat allele frequencies in a large sample of 181 countries across the world, which I use as the key explanatory variable in the empirical analysis on the prevalence of novelty-seeking traits in society.

3.3 Additional Control Variables Related to Biogeography, History, Social, and Economic Factors

In this section, I outline the choice of the set of additional control variables employed across the various model specifications in the empirical analysis. The selection of these control variables is based on their relative predictive power in the field of development economics in explaining differences in standards of living across countries that appear equally important in the creation of knowledge in society. These variables refer to a full set of biogeographic, socioeconomic, and historical factors.

Biogeographic Factors. An empirical analysis of the determinants of scientific knowledge creation should include a full set of microgeographic, land productivity, climatic, health, and regional factors. For example, geographic proximity to technologically advanced countries might facilitate knowledge diffusion across national borders, which in turn could positively affect the country's own rate of technological progress. Furthermore, it has been reported that the country's natural endowment with mineral resources might result in widespread corruption and rent-seeking activities that ultimately discourage the accumulation of individual human capital in society (Sachs and Warner, 2001). Additionally, unfavorable climatic conditions (e.g., temperature and precipitation) are associated with higher physiological thermal stress, which undermines the individual's cognitive performance. The heavy disease burden

[5] In biology, genetic information in every cell is present in the form of chromosomes, which are made up of long threads of relatively small nucleotides. These long threads are the carriers of genetic information and are called DNA. A particular gene corresponds to a shorter segment of the DNA thread along a specific chromosome. The genes are actually responsible for specific activities in the cells. Each gene can take many different forms, which are called alleles of that gene. For example, the specific gene that is responsible for different blood types contains different allele variants (A, B, and O). In molecular genetics, a gene is called *polymorphic* if population geneticists detected more than two different allele variants in populations. The interested reader is referred to Cavalli-Sforza et al. (1994) for a more detailed discussion of the definitions and methods employed in molecular genetics.

associated with tropical areas (e.g., prevalence of malaria falciparum) is another important determinant of cross-country differences in standards of living due to its potential detrimental impact on labor productivity (Gallup et al., 1999).

Besides providing a source of cross-country variation in scientific knowledge creation, biogeographic controls fulfill another important purpose in the empirical analysis. Recent evidence suggests that local biogeographic conditions contribute significantly to between-population variation of DRD4 exon III allele variants (Gören, 2016). Therefore, the inclusion of a large set of biogeographical controls in the regression model effectively rules out endogeneity concerns that the proposed association between knowledge creation and the prevalence of novelty-seeking traits solely reflects the issue of geographic proximity.

Given the fact that the distribution of DRD4 exon III allele variants differs substantially across continents, the estimated impact associated with the country-level DRD4 exon III 2- and 7-repeat measure could simply reflect differences in knowledge creation across continents rather than countries. I control for this kind of endogeneity concern through the inclusion of a full set of regional fixed effects (i.e., indicator variables for countries located in North America, Central and South America, Sub-Saharan Africa, East Asia, Central Asia, South Asia, the Middle East, Europe, or Oceania) across all model specifications. It is worth mentioning that the inclusion of regional fixed effects further controls for unobserved heterogeneity related to any kind of systematic measurement error in the construction of the key country-level DRD4 exon III 2- and 7-repeat allele frequency measures and the set of historical, cultural, and socioeconomic controls across continents, which might undermine the interpretation of the estimated regression coefficients. Following the relevant literature, I further include a set of indicator variables for countries that are either islands or landlocked, as these may perform differently in terms of innovation activity.

Diversity Factors. The issue of ethnic, linguistic, religious, and even genetic diversity is the subject of intense scientific debate on the potential costs and benefits of a diverse society on socioeconomic outcomes such as redistribution, provision of public goods, economic growth, and the incidence of conflict (Alesina et al., 2003; Desmet et al., 2012). On the one hand, a more diverse society may exhibit a wide range of human skills that might be complementary in an increasingly complex production environment. In theory, teams comprised of diverse members may outperform homogenous ones due to increased problem-solving capabilities of the former group (Hong and Page, 2001, 2004). In contrast, the potential costs of a diverse society on aggregate productivity are associated with higher communication costs, lower interpersonal trust, and conflicts. These factors appear equally important for the level of innovation activity as it may hamper the diffusion of knowledge in society.

Recent evidence on the potential benefits and costs of genetic diversity on aggregate productivity suggests that this issue affects economic development quite differently than, for example, ethnic diversity (Ashraf and Galor, 2013). Several mechanisms have been reported that might be consistent with the non-monotonic inverted-U relationship between genetic diversity and contemporary economic development. It has been suggested that the potential costs of genetic diversity are associated with lower interpersonal trust and higher conflict incidence in society (Ashraf and Galor, 2013; Arbatli et al., 2015). More importantly, in regard to the research focus of this paper, the study in Ashraf and Galor (2013) shows that the potential benefits of genetic diversity work through increased innovation activity, as indicated by the number of scientific and technical journal articles per capita in a cross-section of countries.

Historical Factors. The enduring legacy of historical factors has attracted considerable attention in the field of development economics. Specifically, it has been reported that the percentage of Europeans in countries during the early stages of colonization contributed significantly to local economic development (Acemoglu et al., 2001, 2002). This finding is interpreted to suggest Europeans developed specific kinds of political institutions, human capital factors, technology, and culture that were particularly conducive to the process of economic development (Glaeser et al., 2004; Easterly and Levine, 2016).

Additional mechanisms by which a country's colonial history might affect scientific knowledge creation relate to the issue of knowledge diffusion between former colonized countries and the former colonizing powers. For example, countries that were once British colonies might find it easier to absorb knowledge and technology from Great Britain. Related to this aspect is the issue of whether the country's official language is English. This definition covers former colonies that adopted English as their official language in school, business, and public transactions. The inclusion of this variable effectively controls for two distinct but important aspects in the empirical analysis. First, this variable is intended to capture the country's linguistic proximity to the language of scientific communication (e.g., English) which may be particularly capable of absorbing global knowledge stocks, among other things. Second, bibliographic information from the SCI and SSCI databases may be biased toward scientific and technical articles published in English-speaking journals. This may underestimate the extent of scientific knowledge creation in non-English-speaking countries.

Another focus of historical research is on the legal traditions characterizing a country's political system, which capture the extent of the government's interventions into political and social life (La Porta et al., 1999). The protection of individual rights is certainly an important prerequisite for the freedom of expression and individual self-realization, which substantially affect the creation and diffusion of knowledge in society. It is worth mentioning that British common law is the least interventionist of the set of different national legal traditions, emphasizing the protection of individual rights against state authorities (La Porta et al., 1999). In contrast, the socialist legal tradition exerts a strong influence on the content of public education and even limits vocational freedom.

To consider cultural factors as another aspect of history, the baseline specification further includes a full set of a country's major religions (i.e., the percentage of Protestants, Catholics, and Muslims). It has been argued for quite some time that the kind of cultural norms and beliefs frequently associated with adherents to Protestantism (e.g., higher literacy rate) resulted in the so-called 'Protestant work ethic', which was particularly beneficial for the process of economic development (Weber, 1958). The promotion of literacy among Protestants and their relative emancipation from the church may have contributed significantly to the rise of secular knowledge in society. Accordingly, La Porta et al. (1999) point out that the culture of intolerance and closed-mindedness propagated by hierarchically organized religions (e.g., Catholicism and Islam) during the process of industrialization may have restricted the inflow of ideas into society, resulting in economic backwardness relative to countries whose major religion was Protestantism.

Technological Frontier Factors. Observed disparities in standards of living have been attributed to barriers in the adoption of technological innovations across countries (Parente and Prescott, 1994). According to this theory, barriers in the diffusion of general or scientific knowledge across societies result in differences in the level of technological and thus economic development. Therefore, the identification of barriers to the diffusion of knowledge across countries is key to our understanding of contemporary income differences.

In research on the human barriers to knowledge diffusion, Spolaore and Wacziarg (2009) have shown that measures of relative genetic distance to the technological frontier (e.g., the United States) appear to have a statistically significant and economically sizeable effect on cross-country differences in income per capita. The authors interpret their results as evidence that genetic distance between populations captures differences in culture, norms, and beliefs that affect the diffusion of technology across societies. In other words, countries that are genetically similar to the frontier society may find it easier to adopt recent technological advances that in turn positively affect their own economic development (Spolaore and Wacziarg, 2012). This finding is consistent with micro-level evidence on the determinants of international patent citations, highlighting the importance of knowledge diffusion across countries through ethnic scientific networks (Kerr, 2008).

In the empirical analysis, I consider various measures of genetic, linguistic, and religious distance from the technological frontier (i.e., United States, OECD, and Neo-Europe). Even though the correlation between the various technological distance measures is quite high, its implications for cross-country knowledge diffusion through long-term genealogical, linguistic, and religious differences may differ. For example, cultural differences in habits, norms, and beliefs may result in lower trust and communication between populations that in turn prevent the diffusion of knowledge and ideas across countries, even though the two populations are genetically indistinguishable from each other (e.g., German and French).

Technological and Human Capability Factors. The creation and adoption of new scientific knowledge requires specific investments in a country's technological and human capability resource base. This definition is similar to the notion of absorptive capacity initially proposed by Cohen and Levinthal (1990). In theory, models of endogenous economic growth already suggest that cross-country differences in the rate of technological progress are attributable to differences in the country's human capital and research and development resource base (Romer, 1990). Even though the identification and measurement of the relevant technological and human capability factors is a difficult task to accomplish in empirical work, the survey presented in Fagerberg et al. (2010) provides a useful guideline for identification of the relevant factors. Specifically, a country's level of economic development might be used as an overall measure for standards of living that are highly correlated with other proximate factors of economic growth, such as investments in human capital, physical capital, and infrastructure quality, among others. Besides the use of GDP per capita in the empirical analysis, I employ additional variables to identify the specific channels that facilitate scientific knowledge creation in society (i.e., various proxy variables related to national R&D efforts, infrastructure quality, and the level of human capital skills).

Social Value Factors. In addition to the importance of pecuniary incentives in the process of knowledge creation, Abramovitz (1986) emphasizes that differential rates of convergence in income per capita across countries may be attributed to 'social capability' factors that define a country's potential to either create or adopt new technological innovations. This concept includes socially defined values that facilitate knowledge creation through beneficial, co-operative social activities (Fagerberg et al., 2010). These social values may encourage interpersonal communication over the exchange of ideas, knowledge, and technology. Although the quantitative measurement of social capability factors appears difficult, variables such as openness to trade or international migration and civil liberties (e.g., freedom of expression and belief and freedom of speech) could to some extent reflect latent social values related to openness of society and pluralism. For example, it has been argued that the attainment of higher social status and the discovery of new knowledge together are a main driving force in the scientific research process (Merton, 1973). Consistent with this hypothesis, Gorodnichenko and Roland (2017) report a positive and statistically significant link between a culture's tendency toward individualism and the choice of research-oriented occupations among the population, based on U.S. census data from the years 1970 and 2000.

Quality of Governance Factors. Differences in the forms of political institutions are considered one of the main determinants of differences in the level of economic development across countries (Acemoglu and Johnson, 2005). The reason is that the type of institutional settings affects people's economic incentives, influencing the accumulation of knowledge and technology in the aggregate economy. Factors such as freedom from corruption, intellectual property rights protection, the enforcement of law and order, the independence of the judiciary, and the constraint of executive power of political parties and leaders may substantially affect the individual's rate of return of accumulating knowledge or human capital in general.

4 Econometric Specification

The theoretical model predicts that the prevalence of novelty-seeking traits in society is conducive for innovation activity in the aggregate economy, which is consistent with the hypothesis that such traits facilitate the creation of knowledge through beneficial economic attitudes related to risk-taking, creativity, and explorative behavior. In order to investigate this hypothesis empirically, the following cross-country regression equation between scientific knowledge creation and the prevalence of novelty-seeking traits is estimated

$$STP_c = \beta_0 + \beta_1 DRD4_c^{R2R7} + \boldsymbol{\beta}_2'\boldsymbol{\Gamma}_c + \boldsymbol{\beta}_3'\boldsymbol{\Delta}_c + \boldsymbol{\beta}_4'\boldsymbol{\Theta}_c + \boldsymbol{\beta}_5'\boldsymbol{\Lambda}_c + \boldsymbol{\beta}_6'\boldsymbol{\Xi}_c + \boldsymbol{\beta}_7'\boldsymbol{\Phi}_c + \boldsymbol{\beta}_8'\boldsymbol{\Psi}_c + \boldsymbol{\beta}_9'\mathbf{R}_c + \varepsilon_c, \quad (22)$$

where STP_c refers to the average number of scientific and technical publications per 1,000 people in country c during the period 1981 to 2013, and $DRD4_c^{R2R7}$ is the country-level DRD4 exon III 2- and 7-repeat allele frequency measure. To maintain the robustness of the main empirical results, the regression model also includes a full set of biogeographic, historical, and socioeconomic control variables. Again, the choice of relevant variables is intended to rule out two main considerations in the empirical analysis related to the possibility of confounding and the omission of key determinants of knowledge creation. Specifically, the baseline specification includes a vector $\boldsymbol{\Gamma}$ that refers to a full set of biogeographic control variables (e.g., microgeographic, climatic, and land productivity factors). The inclusion of biogeographic controls rules out the possibility that the relationship between knowledge creation and the country-level measure of novelty-seeking traits simply reflects the issue of geographic proximity. Moreover, differences in biogeographic conditions might itself explain variations in knowledge creation across countries. For example, the factors responsible for low labor productivity in areas located in the tropics (e.g., health outcomes, life expectancy, and the formation of human capital in malaria-prone regions) appear to be directly important for the creation of knowledge in society. Additionally, geographic factors affect knowledge creation indirectly, through past institutional, cultural, and economic events and processes whose effects persist to this day. Furthermore, the vector $\boldsymbol{\Delta}$ refers to a full set of historical variables (e.g., colonial heritage, cultural background, and legal tradition) that might be equally important for the accumulation and diffusion of knowledge in society. In addition to this standard set of country-level controls, I further assess the sensitivity of the main results to the inclusion of diversity ($\boldsymbol{\Theta}$), technological frontier ($\boldsymbol{\Lambda}$), technological and human capability ($\boldsymbol{\Xi}$), social value ($\boldsymbol{\Phi}$), and quality of governance ($\boldsymbol{\Psi}$) factors. The line of reasoning for the inclusion of these factors in the regression model was discussed above in the data description section. The regression equation further includes a full set of regional fixed effects (e.g., indicator variables for the main world regions, landlocked, and island countries), as summarized in the vector $\mathbf{R}$. Finally, the term ε_c refers to a country-specific error term.

A positive and statistically significant coefficient associated with the country-level $DRD4^{R2R7}$ measure would be consistent with the hypothesis that the prevalence of novelty-seeking traits in society increases the level of scientific knowledge through beneficial economic attitudes related to risk-taking, explorative activities, and creativity.

5 Empirical Results

This section presents the first results on the relationship between the number of scientific and technical journal articles per 1,000 people and the country-level DRD4 exon III 2- and 7-repeat allele frequency measure. I first present coefficient estimates that include various biogeographic controls. Afterwards, I test the sensitivity of the main findings to the issue of genetic, ethnic, linguistic, and religious diversity. This step of the empirical analysis is of particular importance to rule out concerns that the estimated coefficient associated with the country-level $DRD4^{R2R7}$ measure simply captures unobserved effects related to various aspects of diversity in society. Once the robustness of the main findings to the inclusion of the various diversity controls has been established, I proceed in presenting the baseline specification, which also includes a full set of historical controls in the regression model. This model specification then forms the basis for the subsequent analysis of the robustness of the main results to the inclusion of additional social, economic, and institutional controls.

Including Biogeographic Factors. Table 1 presents coefficient estimates for the set of biogeographic controls. The results shown in column (1) assess the relationship between the number of scientific and technical journal articles per 1,000 people and the country-level $DRD4^{R2R7}$ measure conditional on a full set of regional, island, landlocked, and OPEC fixed effects. Even though the estimated coefficient associated with the country-level $DRD4^{R2R7}$ measure is positive, it is not statistically significant at conventional significance levels. This result suggests that the influence of the country-level $DRD4^{R2R7}$ measure on scientific knowledge creation in society might be conditional on additional country-specific controls.

Hence, the results presented in column (2) include a set of microgeographic factors in the regression model. Once conditioning on this set of microgeographic controls, the estimated coefficient associated with the country-level $DRD4^{R2R7}$ variable increases substantially in magnitude and turns highly statistically significant at the 1% significance level. Regarding the estimated magnitude, the result suggests that a one standard deviation increase in the country-level $DRD4^{R2R7}$ measure in the 123-country sample ($\triangle DRD4^{R2R7} = 0.0658$) would, ceteris paribus, raise the number of scientific and technical journal articles per 1,000 people by about 0.1133 scientific articles. The magnitude of the estimated effect is quite substantial, corresponding approximately to a 32% increase of the standard deviation of the number of scientific and technical journal articles per 1,000 people. The remaining coefficient estimates are all of the expected signs. For example, the positive coefficient associated with the country's approximate absolute latitude in decimal degrees is consistent with productivity-enhancing climate conditions usually found in higher latitude regions (Gallup et al., 1999). Moreover, geographic proximity to the world's three major technological regions (e.g., the United States, continental Europe, and Japan) appear of considerable importance for the creation of knowledge in society through, for example, knowledge diffusion.

In column (3), I include the logarithm of the country's proven mineral resources (i.e., oil and natural gas reserves expressed in British thermal units per capita in the year 1993) in the regression equation. It is generally acknowledged that countries rich in mineral resources perform differentially in economic terms due, for example, to widespread corruption and rent-seeking activities that might hinder the accumulation of knowledge in society (Sachs and Warner, 2001). As expected, the estimated coefficient associated with the log of hydrocarbons per capita is negative, but statistically insignificant.

The estimates presented in column (4) examine the sensitivity of the main results to the inclusion of various land productivity factors. Neither of these variables alter the main findings substantially. The estimated coefficient associated with the country-level $DRD4^{R2R7}$ measure increases both in magnitude and statistical significance, suggesting that the previous model specification was partially confounded by the omission of land productivity factors. It is worth mentioning that the coefficient associated with land suitability for agriculture enters with a negative sign in the regression equation. This result is consistent with the hypothesis that the extent of scientific knowledge creation is on average lower in countries that place a greater emphasis on agricultural practices.

In column (5), the relationship between scientific knowledge creation and the country-level $DRD4^{R2R7}$ measure is also robust to the inclusion of climatic factors. The positive coefficient associated with the country's mean precipitation value is consistent with the notion that knowledge creation is higher in temperate climatic zones. Specifically, the cognitive performance of individuals might be negatively affected by higher thermal stress in regions with high temperatures and lower precipitation.

Finally, the results presented in column (6) examine the sensitivity of the previous findings to the inclusion of health-related factors in the regression model. Again, the estimated coefficient associated with the country-level $DRD4^{R2R7}$ measure remains highly statistically significant at the 1% significance level. Furthermore, the coefficient estimate regarding the share of a country's population in areas with malaria enters with the expected negative sign, consistent with the heavy disease burden and the resulting detrimental effects on the accumulation of knowledge in society (Gallup et al., 1999).

The Issue of Diversity. In the following, I assess the sensitivity of the previous findings to the issue of genetic, ethnic, linguistic, and religious diversity. Specifically, given the fact that the country-level DRD4 exon III 2- and 7-repeat allele frequency measure is positively correlated with migratory distance from East Africa, it is conceivable that the estimated regression coefficient simply captures the impact of genetic diversity on scientific knowledge creation (Gören, 2017). Indeed, Ashraf and Galor (2013) show that the extent of knowledge creation (measured by the number of scientific and technical journal articles) is positively affected by the level of overall genetic diversity in society. Furthermore, I examine the sensitivity of the main results to the inclusion of ethnic, linguistic, and religious diversity to rule out concerns that the empirical findings might be prone to unobserved factors related to various definitions of diversity in society. The corresponding results are shown in Table 2.

The estimates presented in column (1) correspond to the model specification in Table 1, column (6), and are shown for comparison purposes.

In column (2), I include the measure of genetic diversity, predicted by migratory distance from East Africa, in the regression equation. To account for the fact that most countries today have populations from various regions of the world, an ancestry-adjusted version of this measure is employed based on population flow data since 1500 AD (Putterman and Weil, 2010). The coefficient estimate of predicted genetic diversity is positive but statistically insignificant. More importantly, the regression coefficient on the country-level $DRD4^{R2R7}$ measure remains relatively robust and precisely estimated at the 1% significance level.

The estimates presented in column (3) examine the sensitivity of the main results to the inclusion of ethnic diversity. The estimated coefficient associated with ethnic diversity enters with a negative sign in the regression model. This result is consistent with the notion that ethnically diverse societies are subject to higher communication costs and greater conflict potential, which may hamper the accumulation of knowledge in the aggregate economy. Again, the estimated coefficient associated with the country-level $DRD4^{R2R7}$ measure remains rather robust and precisely estimated.

In columns (4) and (5), I subsequently include measures of linguistic and religious diversity in the regression model. None of the included measures were statistically significant at conventional significance levels. Furthermore, the regression coefficient associated with the key explanatory variable remains virtually unchanged and statistically significant at the 1% significance level.

Finally, the results shown in column (6) test the sensitivity of the main findings to the simultaneous inclusion of the various diversity measures in the regression model. Only the regression coefficient associated with ethnic diversity remains statistically significant at the 5% significance level. Nevertheless, the main results regarding the country-level $DRD4^{R2R7}$ measure are unaffected even in this augmented model specification. Overall, the main findings from this empirical exercise reveal that the relationship between knowledge creation and the country-level $DRD4^{R2R7}$ measure is not confounded by unobserved factors related to various definitions of diversity in society.

Including Historical Factors. In this section, I examine the sensitivity of the main findings to the inclusion of country-specific historical factors. The corresponding estimates are reported in Table 3. I first present the core findings, which include the full set of biogeographic and regional factors in the regression model, as shown in column (1). In the subsequent analysis, I examine the sensitivity of the regression coefficient associated with the country-level $DRD4^{R2R7}$ measure to the inclusion of various country-specific historical controls (e.g., colonial heritage, cultural background, and legal tradition), as shown in columns (2) to (6).

I start the empirical analysis by including the country's share of the European population in the regression equation, as shown in column (2). It has been argued that the kind of norms, values, and beliefs frequently attached to European populations might be particularly favorable to the process of economic development by facilitating the accumulation of human capital in society (Glaeser et al., 2004). Consistent with this idea, the estimated regression coefficient associated with the share of European population is positive and highly statistically significant at the 1% significance level. The point estimate suggests that increasing the share of Europeans in Cameroon (i.e., = 0%) to the median value of Europeans in the 123-country sample (e.g., Turkmenistan = 5.5%) would, ceteris paribus, raise the number of scientific and technical publications per 1,000 people by about 0.0534 scientific articles. This estimate corresponds roughly to a 15% increase in the standard deviation of the number of scientific and technical publications per 1,000 people. However, the estimated coefficient associated with the country-level $DRD4^{R2R7}$ measure remains qualitatively unaffected by the inclusion of the share of European population in the regression model.

The next two model specifications assess the sensitivity of the previous findings to the country's linguistic and colonial heritage, respectively. The corresponding estimates are shown in columns (3) and (4). As discussed in the data description section of this paper, the inclusion of an English language indicator controls for the ease of knowledge acquisition and a potential bias in the number of scientific journal articles in English-speaking countries, respectively. Even though the coefficient of the English indicator enters with a positive sign in the regression model, it is not statistically different from zero, as shown in column (3). Furthermore, the robustness of the estimated coefficient associated with the country-level $DRD4^{R2R7}$ measure suggests that the main results are not additionally confounded by a possible publication bias toward English-speaking journals. In addition, the main results presented in column (4) remain qualitatively unaffected by the inclusion of a former British colony indicator. The positive and statistically significant coefficient on the British colony indicator suggests potential knowledge spillovers between Britain and its former colonies that facilitate the latter countries' scientific knowledge bases.

In column (5), I examine the importance of the legal tradition in the creation of scientific knowledge in society. Interestingly, the regression coefficient on the British legal tradition indicator is positive and statistically significant at the 5% significance level. This result is consistent with the idea that a legal tradition that protects individual rights against arbitrary state interventions may facilitate freedom of expression, which appears quite important for the creation of knowledge in society. The positive and highly statistically significant coefficient associated with the Scandinavian legal tradition may reflect the high scientific productivity observed in Nordic countries (i.e., Denmark, Finland, Norway, and Sweden) triggered, for example, by high levels of research and development expenditures (OECD, 2002; Sihvonen and Vähämaa, 2015). Even though the regression coefficient of the country-level $DRD4^{R2R7}$ measure falls by 26% in this model specification, it remains highly statistically significant at the 1% significance level.

In column (6), I investigate the cultural hypothesis of scientific knowledge creation by including the country's major religions (i.e., Protestants, Catholics, and Muslims) as percentages of the population in the regression model. The estimated coefficient associated with the country-level $DRD4^{R2R7}$ measure falls by an additional 26%, but remains precisely estimated at the 1% significance level. With regard to economic magnitude, the coefficient estimates in this final model specification suggest that a one standard deviation increase in the country-level $DRD4^{R2R7}$ measure would, ceteris paribus, raise the number of scientific and technical publications per 1,000 people by about 0.0678 scientific articles. This estimate corresponds approximately to a 19% increase of the standard deviation of the number of scientific and technical publications per 1,000 people. It is worth mentioning that only the coefficient associated with the country's share of Muslims in the population enters with a negative and statistically significant sign in the regression model. The positive but insignificant coefficient associated with the share of Protestants may at first glance seem to contradict the 'Protestant work ethic' hypothesis proposed by Weber (1958). However, the share of Protestants in the overall population is strongly correlated with the share of Europeans and the British legal tradition indicator, which would explain its insignificant effect on the number of scientific and technical publications. In other words, once conditioning on the share of European population and the legal tradition, there is no separate effect of the share of Protestants on scientific knowledge creation in society.

6 Robustness Analysis

In this section, I examine the robustness of the baseline estimates to the inclusion of technological frontier, technological and human capability, social value, and quality of governance factors that might be important for scientific knowledge creation in society. In addition, I present results from an instrumental variables estimation approach to rule out endogeneity concerns in the relationship between knowledge creation and the country-level $DRD4^{R2R7}$ measure. I elaborate on these issues further below in the empirical analysis.

Technological Frontier Factors. In the following, I investigate the hypothesis that countries that are closer to the technological frontier (whether in genetic, linguistic, or religious terms) might find it easier to adopt advanced knowledge stocks and use them more efficiently for the accumulation of scientific knowledge. The corresponding estimates are shown in Table 4. I consider three different entities that define the technological frontier at a global level: the United States, member countries of the OECD, and Neo-European countries. For each of these regions, I calculated the country's genetic, linguistic, and religious distance from the technological frontier, each capturing different aspects of human barriers to the diffusion of knowledge across countries. Again, I assessed the sensitivity of the estimated coefficient regarding the country-level $DRD4^{R2R7}$ measure relative to the inclusion of the various technological distance variables.

In columns (1) to (4), I present coefficient estimates for the country-level $DRD4^{R2R7}$ measure and the various technological distance variables relative to the United States. First of all, the main results regarding the key country-level $DRD4^{R2R7}$ measure remains qualitatively unaffected by the inclusion of these technological distance factors. Additionally, the estimates in column (4) suggest that linguistic barriers to the United States as the technological frontier have a negative and statistically significant impact on scientific knowledge creation, consistent with the idea of linguistic barriers to the diffusion of knowledge across countries.

The results presented in columns (5) to (8) employ the member countries of the OECD as the global technological frontier. However, the main results remain qualitatively unaffected by this definition.

Finally, in model specifications (9) to (12), I define Neo-European countries (i.e., Australia, Canada, the United States, and New Zealand) as the global technological frontier. Again, the estimated coefficient associated with the country-level $DRD4^{R2R7}$ measure remain relatively robust and precisely estimated.

Technological and Human Capability Factors. The creation of knowledge might depend on the country's technological and human resource base. Specifically, technologically advanced countries with a well-educated workforce might be quite efficient in the creation and absorption of scientific knowledge. To investigate this issue in more detail, I present coefficient estimates that control for a set of technological and human capability factors in the regression equation. The corresponding estimates are shown in Table 4. Notice that the baseline sample is reduced to 99 countries due to missing observations on the set of technological and human capability factors. Nevertheless, the main results remain qualitatively unaffected when restricting the baseline specification to the reduced 99-country sample, as shown in column (1).

In a first exercise, I include the logarithm of GDP per capita in the regression model. The level of economic development can be interpreted as a summary measure capturing the standards of living in society (e.g., quality of public services, political institutions, and technology). The main results remain unaffected by the inclusion of GDP per capita in the regression model, as shown in column (2). In a similar vein, the estimates presented in column (3) show that the coefficient associated with the country-level $DRD4^{R2R7}$ measure is not sensitive to the inclusion of a human capital index. More importantly, the estimates in column (4) demonstrates very clearly that the level of resources devoted to research and development activities (measured as R&D expenditure over total GDP) is significantly correlated with the creation of scientific knowledge. Even though the regression coefficient associated with the key $DRD4^{R2R7}$ measure falls substantially in magnitude, its impact remains statistically different from zero at the 5% significance level. In columns (5) to (7), I examine the influence of infrastructure quality controls (measured as the number of internet users, fixed telephone, and mobile cellular subscriptions per 100 people) on the creation of scientific knowledge in society. All estimated coefficients associated with the various infrastructure quality measures are positive and statistically significant in all but one model specification. It is worth mentioning that internet access may also reflect openness to new ideas and people that appear beneficial for the creation of knowledge in society (Fagerberg and Srholec, 2008). The regression coefficient associated with the country-level $DRD4^{R2R7}$ measure is virtually unaffected by the inclusion of infrastructure quality controls.

Finally, the estimates presented in column (8) assess the sensitivity of the main results to the simultaneous inclusion of the various technological and human capability factors in a single model specification. The relationship between scientific knowledge creation and the country-level $DRD4^{R2R7}$ measure remains positive and statistically significant at conventional significance levels.

Social Value Factors. In this part of the empirical analysis, I examine the sensitivity of the coefficient associated with the key $DRD4^{R2R7}$ measure relative to the inclusion of various social value factors. The relevant estimates are shown in Table 6. Again, the estimates presented in column (1) correspond to the baseline results in the 123-country sample and are shown for comparison purposes. In column (2), I include a measure of trade openness in the regression model. The corresponding regression coefficient is negative and highly statistically significant at the 1% significance level. I interpret this finding to indicate that international trade may reduce national efforts in the accumulation of scientific knowledge through the exchange of foreign technology embodied in high-tech commercial goods and services. Next, I consider the effect of international migration on scientific knowledge creation. The movement of people from one country or region to another may facilitate the diffusion of ideas and therefore might contribute positively to the creation of knowledge in society. The corresponding regression coefficient is positive but not statistically different from zero at conventional significance levels, as shown in column (3). In the following two model specifications, I consider a set of social value factors that might appear important for the creation of knowledge in society. These variables refer to the freedom of expression and belief and to a composite index of civil liberties. The corresponding estimates are reported in columns (4) and (5), respectively. Finally, I report coefficient estimates that simultaneously include all social value factors in one regression equation, as shown in column (6). Neither of these model specifications alter the main findings for the key country-level $DRD4^{R2R7}$ measure substantially.

Quality of Governance Factors. In Table 7, I examine the robustness of the main results to quality of governance factors. To summarize the main findings from this empirical exercise, the results clearly illustrate that freedom from corruption is the most important political factor in the creation of scientific knowledge in society. This result is consistent with the notion that widespread corruption is associated with pervasive rent-seeking activities that undermine private returns to knowledge accumulation. Moreover, executive constraints on government authorities and the extent of institutional democracy are positively correlated with scientific knowledge creation when entered individually into the regression equation. More importantly, the relationship between the number of scientific and technical publications per 1,000 people and the country-level $DRD4^{R2R7}$ measure remain qualitatively unaffected by the set of quality of governance factors.

Instrumental Variables Estimation Approach. It might be argued that the results reported here are prone to endogeneity problems. Specifically, a country's technological attractiveness could result in a disproportionate inflow of novelty-seeking individuals, resulting in the positive correlation identified between scientific knowledge creation and the country-level $DRD4^{R2R7}$ measure. Although it is difficult to demonstrate that international migration flows since 1500 AD are the result of different countries' technological attractiveness in former times (Spolaore and Wacziarg, 2012), I present additional results from an instrumental variables estimation approach.

I consider the following list of instruments employed in the $DRD4^{R2R7}$ equation:

$$DRD4_c^{R2R7} = \delta_0 + \delta_1 MigDist_c + \boldsymbol{\delta}'_P \boldsymbol{P}_c + \boldsymbol{\delta}'_E \boldsymbol{E}_c + \boldsymbol{\delta}'_L \boldsymbol{L}_c + \nu_c, \tag{23}$$

where $MigDist$ is the migratory distance from East Africa (in 1,000 km) to the country's capital city, $\boldsymbol{P}$ refers to a vector of pasture land controls (i.e., pasture and pasture squared), $\boldsymbol{E}$ is a vector of elevation controls (i.e., elevation and elevation squared), $\boldsymbol{L}$ refers to a vector of land suitability for agriculture controls (i.e., agricultural suitability and agricultural suitability squared), and ν_c is a country-specific error term. The set of excluded country-level instruments were ancestry-adjusted to account for post-1500 AD migration flows across countries. The selection of excluded instruments was motivated by the recent study in Gören (2016), providing a detailed empirical analysis of the biogeographic origins of the human DRD4 exon III allele variants in a large sample of indigenous populations across the world.

The results from the proposed IVs estimation approach are shown in Table 8. A formal test of the relevance of the excluded instruments is performed using the Kleibergen and Paap (2006) rk LM statistic. The null hypothesis of this test is that the set of excluded instruments are uncorrelated with the endogenous variable $DRD4^{R2R7}$. Furthermore, I examine the validity of the excluded instruments using the Hansen J statistic. The null hypothesis of this test is that the set of excluded instruments is valid, and that it is uncorrelated with the error term in the scientific knowledge regression equation (22). I present coefficient estimates for the complete 123-country sample in columns (1) to (2) and the reduced 99-country sample in columns (3) to (11). First of all, the relevance and validity of the IVs estimation approach are satisfied, as indicated by the corresponding $p-values$ of the Kleibergen and Paap (2006) rk LM and Hansen J statistic, respectively. Overall, the regression coefficient associated with the key country-level $DRD4^{R2R7}$ measure remains qualitatively unaffected by this alternative estimation methodology.

7 Conclusion

This paper establishes, both theoretically and empirically, the beneficial effects of novelty-seeking traits on the level of scientific knowledge creation in society. The key explanatory variable refers to the human DRD4 exon III 2- and 7-repeat ($DRD4^{R2R7}$) allele variants that candidate gene studies of personality have linked to the human phenotype of novelty-seeking behavior. The issue of novelty-seeking has been linked to beneficial economic attitudes related to risk-taking, creativity, and entrepreneurship that constitute the primary source of technological progress in Schumpeterian-inspired endogenous growth models. The baseline estimates suggest that a one standard deviation increase in the prevalence of novelty-seeking traits would, ceteris paribus, increase the number of scientific and technical journal articles per 1,000 people by about 0.0678 scientific articles. This result is robust to the inclusion of potentially confounding microgeographic, land productivity, climatic, health, legal, cultural, colonial, and regional factors. The estimated magnitude is substantial as it corresponds approximately to a 19% increase in the standard deviation of the number of scientific and technical publications per 1,000 people. The main findings remain qualitatively robust to the inclusion of additional determinants in cross-country differences of scientific knowledge creation (e.g., technological, human, and institutional factors) and to the use of an IVs estimation methodology to address potential endogeneity concerns in the empirical analysis.

This research contributes significantly to understanding the importance of some deep-rooted historical factors in the creation of scientific knowledge in society. The findings should not be misinterpreted to suggest that the prevalence of novelty-seeking traits is the only factor explaining differences in scientific knowledge creation across countries. Rather, the main findings show that after controlling for a large set of technological, human capital, socioeconomic, cultural, and institutional determinants, there remains a positive and statistically significant effect of the country-level $DRD4^{R2R7}$ measure on the number of scientific and technical journal articles per 1,000 people. This observation highlights the importance of socioeconomic factors that might help to facilitate the creation and diffusion of knowledge across borders.

References

Abramovitz, M. (1986). Catching Up, Forging Ahead, and Falling Behind. *Journal of Economic History*, 46:386–406.

Acemoglu, D. and Johnson, S. (2005). Unbundling Institutions. *Journal of Political Economy*, 113:949–995.

Acemoglu, D., Johnson, S., and Robinson, J. A. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *The American Economic Review*, 91:1369–1401.

Acemoglu, D., Johnson, S., and Robinson, J. A. (2002). Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. *The Quarterly Journal of Economics*, 117:1231–1294.

Aghion, P. and Howitt, P. (1992). A Model of Growth Through Creative Destruction. *Econometrica*, 60:323–351.

Aghion, P. and Howitt, P. (1998). *Endogenous Growth Theory*. The MIT Press, Cambridge, Massachusetts.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, 8:155–194.

Arbatli, E., Ashraf, Q. H., Galor, O., and Klemp, M. (2015). The Nature of Conflict. *mimeo*.

Asghari, V., Sanyal, S., Buchwaldt, S., Paterson, A., Jovanovic, V., and Van Tol, H. H. M. (1995). Modulation of Intracellular Cyclic AMP Levels by Different Human Dopamine D4 Receptor Variants. *Journal of Neurochemistry*, 65:1157–1165.

Ashraf, Q. and Galor, O. (2013). The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development. *The American Economic Review*, 103:1–46.

Barro, R. J. and Lee, J. W. (2013). A New Data Set of Educational Attainment in the World, 1950–2010. *Journal of Development Economics*, 104:184–198.

Basu, S. and Weil, D. N. (1998). Appropriate Technology and Growth. *The Quarterly Journal of Economics*, 113:1025–1054.

Benjamin, J., Li, L., Patterson, C., Greenberg, B. D., Murphy, D. L., and Hamer, D. H. (1996). Population and Familial Association Between the D4 Dopamine Receptor Gene and Measures of Novelty Seeking. *Nature*, 12:81–84.

Cann, H. M., de Toma, C., Cazes, L., Legrand, M. F., Bodmer, V. M. L. P. J., Bodmer, W. F., Bonne-Tamir, B., Cambon-Thomsen, A., Chen, Z., Chu, J., Carcassi, C., Contu, L., Du, R., Excoffier, L., Ferrara, G. B., Friedlaender, J., Groot, H., Gurwitz, D., Jenkins, T., Herrera, R. J., Kidd, X. H. J., Kidd, K. K., Langaney, A., Lin, A. A., Mehdi, S. Q., Parham, P., Piazza, A., Pistillo, M. P., Qian, Y., Shu, Q., Xu, J., Zhu, S., Weber, J. L., Greely, H. T., Feldman, M. W., Thomas, G., Dausset, J., and Cavalli-Sforza, L. L. (2002). A Human Genome Diversity Cell Line Panel. *Science*, 296:261–262.

Cavalli-Sforza, L. L., Menozzi, P., and Piazza, A. (1994). *The History and Geography of Human Genes*. Princeton University Press, Princeton, NJ.

Cingranelli, D. L., Richards, D. L., and Clay, K. C. (2014). The CIRI Human Rights Dataset. *http://www.humanrightsdata.com*.

Cohen, W. M. and Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, 35:128–152.

Cook, C. J. (2014). The Role of Lactase Persistence in Precolonial Development. *Journal of Economic Growth*, 19:369–406.

Cook, C. J. (2015). The Natural Selection of Infectious Disease Resistance and its Effect on Contemporary Health. *The Review of Economics and Statistics*, 97:742–757.

Desmet, K., Ortuño-Ortín, I., and Wacziarg, R. (2012). The Political Economy of Linguistic Cleavages. *Journal of Development Economics*, 97:322–338.

Dreber, A., Apicella, C. L., Eisenberg, D. T. A., Garcia, J. R., Zamore, R. S., Lum, J. K., and Campbell, B. (2009). The 7R Polymorphism in the Dopamine Receptor D_4 Gene (DRD4) is Associated With Financial Risk Taking in Men. *Evolution and Human Behavior*, 30:85–92.

Easterly, W. and Levine, R. (2016). The European Origins of Economic Development. *Journal of Economic Growth*, 21:225–257.

Fagerberg, J. and Srholec, M. (2008). National Innovation Systems, Capabilities and Economic Development. *Research Policy*, 37:1417–1435.

Fagerberg, J., Srholec, M., and Verspagen, B. (2010). Innovation and Economic Development. In Arrow, K. J. and Intriligator, M. D., editors, *Handbook of the Economics of Innovation*, chapter 20, pages 833–872. Elsevier.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The Next Generation of the Penn World Table. *American Economic Review*, 105:3150–3182.

Freedom House (2016). Reports. *https://freedomhouse.org/*.

Gallup, J. L., Sachs, J. D., and Mellinger, A. D. (1999). Geography and Economic Development. *International Regional Science Review*, 2:179–232.

Galor, O. and Michalopoulos, S. (2012). Evolution and the Growth Process: Natural Selection of Entrepreneurial Traits. *Journal of Economic Theory*, 147:759–780.

Glaeser, E. L., Porta, R. L., Lopez-De-Silanes, F., and Shleifer, A. (2004). Do Institutions Cause Growth? *Journal of Economic Growth*, 9:271–303.

Global Mapping International (2010). World Language Mapping System, Version 16. *http://worldgeodatasets.com/language/index.html*.

Gören, E. (2016). The Biogeographic Origins of Novelty-Seeking Traits. *Evolution and Human Behavior*, 37:456–469.

Gören, E. (2017). The Persistent Effects of Novelty-Seeking Traits on Comparative Economic Development. *Journal of Development Economics*, 126:112–126.

Gorodnichenko, Y. and Roland, G. (2017). Culture, Institutions and the Wealth of Nations. *The Review of Economics and Statistics*, Forthcoming.

Grossman, G. M. and Helpman, E. (1991). Quality Ladders in the Theory of Growth. *The Review of Economic Studies*, 58:43–61.

Heritage Foundation (2015). Index of Economic Freedom. *http://www.heritage.org/index/*.

Hong, L. and Page, S. E. (2001). Problem Solving by Heterogenous Agents. *Journal of Economic Theory*, 91:123–163.

Hong, L. and Page, S. E. (2004). Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers. *Proceedings of the National Academy of Sciences*, 101:16385–16389.

Kerr, W. (2008). Ethnic Scientific Communities and International Knowledge Diffusion. *The Review of Economics and Statistics*, 90:518–537.

Kleibergen, F. and Paap, R. (2006). Generalized Reduced Rank Tests Using the Singular Value Decomposition. *Journal of Econometrics*, 133:97–126.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., and Vishny, R. (1999). The Quality of Government. *The Journal of Law, Economics, & Organization*, 15:222–279.

Lichter, J. B., Barr, C. L., Kennedy, J. L., Tol, H. H. V., Kidd, K. K., and Livak, K. J. (1993). A Hypervariable Segment in the Human Dopamine Receptor D_4 (DRD4) Gene. *Human Molecular Genetics*, 2:767–773.

Mannuzza, S. and Klein, R. G. (2000). Long-Term Prognosis in Attention-Deficit/Hyperactivity Disorder. *Child and Adolescent Psychiatric Clinics of North America*, 9:711–726.

Marshall, M. G., Gurr, T. R., and Jaggers, K. (2016). Polity IV Project. *Center for Systemic Peace. http://www.systemicpeace.org/*.

Mayseless, N., Uzefovsky, F., Shalev, I., Ebstein, R. P., and Shamay-Tsoory, S. G. (2013). The Association Between Creativity and 7R Polymorphism in the Dopamine Receptor D4 Gene (DRD4). *Frontiers in Human Neuroscience*, 7:1–7.

Merton, R. K. (1973). *The Sociology of Science: Theoretical and Empirical Investigations*. The University of Chicago Press, Chicago and London.

National Science Board (2014). *Science and Engineering Indicators 2014*. National Science Foundation, Arlington VA.

New, M., Lister, D., Hulme, M., and Makin, I. (2002). A High-Resolution Data Set of Surface Climate Over Global Land Areas. *Climate Research*, 21:1–25.

Nordhaus, W. D. (2006). Geography and Macroeconomics: New Data and New Findings. *Proceeding of the National Academy of Sciences*, 103:3510–3517.

Oak, J. N., Oldenhof, J., and Tol, H. H. V. (2000). The Dopamine D_4 Receptor: One Decade of Research. *European Journal of Pharmacology*, 405:303–327.

OECD (2002). OECD Science, Technology, and Industry Outlook 2002. *OECD Publications Service. http://www.oecd-ilibrary.org/industry-and-services/oecd-science-technology-and-industry-outlook-2002_sti_outlook-2002-en*.

OECD (2015). *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities*. OECD Publishing. http://www.oecd-ilibrary.org/science-and-technology/frascati-manual-2015_9789264239012-en, Paris.

Parente, S. L. and Prescott, E. C. (1994). Barriers to Technology Adoption and Development. *Journal of Political Economy*, 102:298–321.

Psacharopoulos, G. (1994). Returns to investment in education: A global update. *World Development*, 22:1325–1343.

Putterman, L. and Weil, D. N. (2010). Post-1500 Population Flows and the Long Run Determinants of Economic Growth and Inequality. *The Quarterly Journal of Economics*, 125:1627–1682.

Ramachandran, S., Deshpande, O., Roseman, C. C., Rosenberg, N. A., Feldman, M. W., and Cavalli-Sforza, L. L. (2005). Support from the Relationship of Genetic and Geographic Distance in Human Populations for a Serial Founder Effect Originating in Africa. *Proceedings of the National Academy of Sciences*, 44:15942–15947.

Ramankutty, N., Foley, J. A., Norman, J., and McSweeney, K. (2002). The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate change. *Global Ecology and Biogeography*, 11:377–392.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, 98:71–102.

Sachs, J. D. and Warner, A. M. (2001). The Curse of Natural Resources. *European Economic Review*, 45:827–838.

Sihvonen, J. and Vähämaa, S. (2015). Business Research in the Nordic Countries: An Analysis of Research Output across Countries, Disciplines, and Institutions. *Nordic Journal of Business*, 64:266–296.

Spolaore, E. and Wacziarg, R. (2009). The Diffusion of Development. *The Quarterly Journal of Economics*, 124:469–529.

Spolaore, E. and Wacziarg, R. (2012). The Long-Term Barriers to the International Diffusion of Innovations. *National Bureau of Economic Research*, pages 11–46.

Spolaore, E. and Wacziarg, R. (2013). How Deep Are the Roots of Economic Development. *Journal of Economic Literature*, 51:325–369.

Spolaore, E. and Wacziarg, R. (2016). Ancestry, Language and Culture. In Ginsburgh, V. and Weber, S., editors, *The Palgrave Handbook of Economics and Language*, chapter 6, pages 174–211. Palgrave Macmillan.

Teorell, J., Dahlberg, S., Holmberg, S., Rothstein, B., Khomenko, A., and Svensson, R. (2017). *The Quality of Government Standard Dataset, version Jan17*. The Quality of Government Institute, University of Gothenburg. http://qog.pol.gu.se/.

Van Tol, H. H. M., Bunzow, J. R., Guan, H.-C., Sunahara, R. K., Seeman, P., Niznik, H. B., and Civelli, O. (1991). Cloning of the Gene for a Human Dopamine D_4 Receptor with High Affinity for the Antipsychotic Clozapine. *Nature*, 350:610–614.

Van Tol, H. H. M., Wu, C. M., Guan, H.-C., Ohara, K., Bunzow, J. R., Civelli, O., Kennedy, J., Seeman, P., Niznik, H. B., and Jovanovic, V. (1992). Multiple Dopamine D4 Receptor Variants in the Human Population. *Nature*, 358:149–152.

Wang, E., Ding, Y.-C., Flodman, P., Kidd, J. R., Kidd, K. K., Grady, D. L., Ryder, O. A., Spence, M. A., Swanson, J. M., and Moyzis, R. K. (2004). The Genetic Architecture of Selection at the Human Dopamine Receptor D4 (DRD4) Gene Locus. *American Journal of Human Genetics*, 74:931–944.

Weber, M. (1958). *The Protestant Ethic and the Spirit of Capitalism*. Charles Scribner's Sons, New York.

A Figures

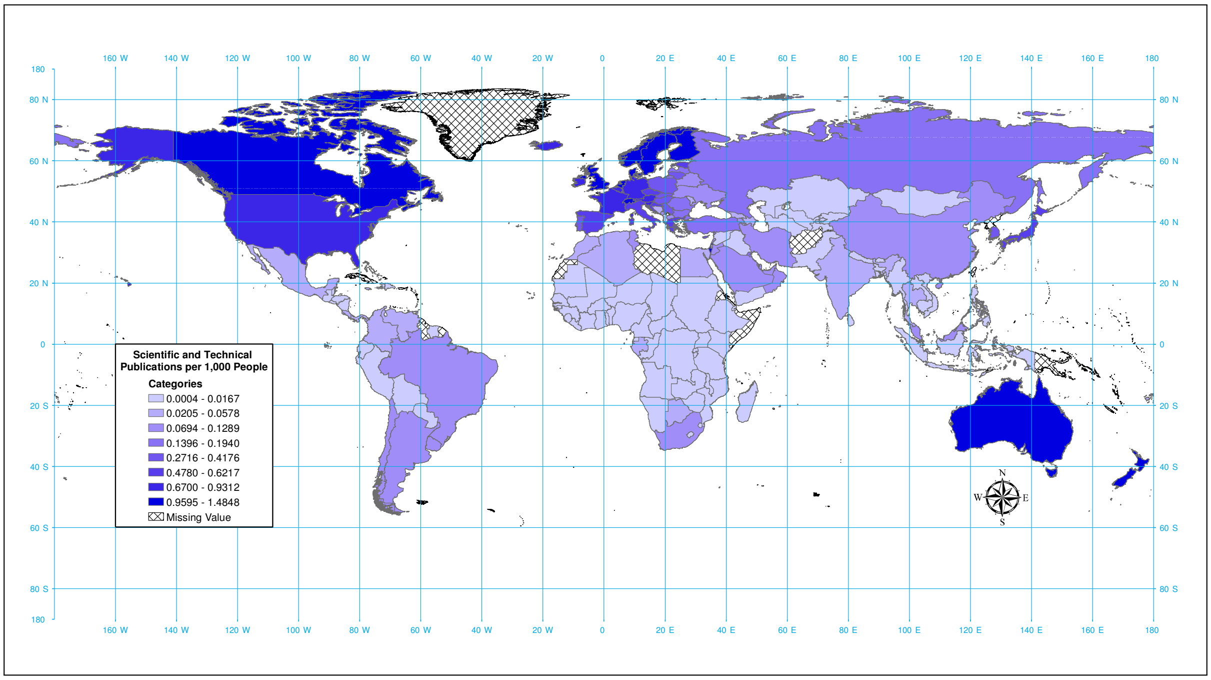

Figure 1: Number of Scientific and Technical Publications per 1,000 People Across Countries

Notes: This map shows the worldwide distribution of the number of scientific and technical publications per 1,000 people in the years 1981 to 2013 across countries. It refers to scientific publications that have been published in the fields of physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences. See the main text for additional details regarding data construction and sources.

B Regression Tables

Table 1: Novelty-Seeking Traits and Number of Scientific and Technical Publications (Biogeographic Factors)

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable: Number of Scientific and Technical Publications per 1,000 People					
$DRD4^{R2R7}$	0.5917	1.7213***	1.7237***	1.8501***	1.7512***	2.0045***
(Ethnicity-Weighted)	(0.5281)	(0.5702)	(0.5824)	(0.5616)	(0.5537)	(0.6201)
Absolute Latitude		0.0108***	0.0107***	0.0097***	0.0080*	0.0065
		(0.0031)	(0.0030)	(0.0029)	(0.0042)	(0.0047)
ln Distance to Major Markets		-0.1399***	-0.1399***	-0.1457***	-0.1482***	-0.1518***
		(0.0417)	(0.0419)	(0.0423)	(0.0392)	(0.0419)
ln Distance to Coast or River		-0.0211	-0.0206	-0.0516*	-0.0493	-0.0346
		(0.0258)	(0.0283)	(0.0270)	(0.0325)	(0.0331)
ln Hydrocarbons per Person			-0.0003	-0.0022	-0.0019	-0.0038
			(0.0065)	(0.0064)	(0.0064)	(0.0067)
ln Percentage Arable Land Area				-0.0096	-0.0041	0.0046
				(0.0258)	(0.0250)	(0.0267)
Agricultural Suitability				-0.2878**	-0.3182**	-0.3690***
				(0.1254)	(0.1266)	(0.1298)
Terrain Roughness				0.1531	0.0126	-0.0457
				(0.1860)	(0.2176)	(0.2293)
Temperature					-0.0090	-0.0056
					(0.0067)	(0.0064)
Precipitation					0.0012**	0.0016***
					(0.0006)	(0.0006)
Population in Tropics (Share)						-0.0529
						(0.0816)
Population in Area with Malaria (Share)						-0.2022*
						(0.1021)
Number of Countries	123	123	123	123	123	123
R^2	0.54	0.63	0.63	0.66	0.67	0.69
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Island Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Landlock Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
OPEC Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the average number of scientific and technical publications per 1,000 people during 1981 to 2013.
Independent variables: $DRD4^{R2R7}$ refers to the ethnicity-weighted DRD4 exon III 2- and 7-repeat allele frequency measure on the country-level. *Absolute Latitude* is the absolute value of a country's approximate centroid latitude in decimal degrees. ln *Distance to Major Markets* is the log of the minimum great-circle distance (in 1,000 km) from the country's capital city to New York, Rotterdam, or Tokyo. ln *Distance to Coast or River* is the log of mean distance (in 1,000 km) to the nearest ice-free coastline or sea-navigable river. ln *Hydrocarbons per Person* is the log of British thermal units per person of proven crude oil and natural gas reserves in 1993. ln *Percentage Arable Land Area* is the log of a country's percentage arable land area. *Agricultural Suitability* is a geospatial indicator, ranging from 0 to 1, of land suitability for agriculture across 0.5 decimal degrees latitude × longitude grid cells in the area covered by each country. *Terrain Roughness* is a geospatial indicator that indicates the the average absolute change in elevation values across contiguous 1 decimal degrees latitude × longitude grid cells in the area covered by each country. *Temperature* is the mean country's temperature (in degree celsius) during the period 1960 and 1990. *Precipitation* is the mean country's precipitation (in total millimeters per month) during the period 1960 and 1990. *Population in Tropics (Share)* is the share of a country's population in 1995 residing in tropics. *Population in Area with Malaria (Share)* is the share of a country's population in 1995 residing in areas contracting with malaria falciparum. *Region Fixed Effects* refer to region dummies for Sub-Saharan Africa, North America, Central and South America, East Asia, Central Asia, South Asia, Middle East, and Europe. *Island Dummy* takes value 1 for countries being an island. *OPEC Dummy* takes value 1 for member countries belonging to the Organization of the Petroleum Exporting Countries (OPEC). *Landlock Dummy* takes value 1 for countries being landlocked. Constant term included but not shown.
Robust standard errors are reported between parenthesis.
*: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level.

Table 2: Novelty-Seeking Traits and Number of Scientific and Technical Publications (Diversity Factors)

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable: Number of Scientific and Technical Publications per 1,000 People					
$DRD4^{R2R7}$	2.0045***	2.2631***	2.0963***	1.9826***	2.0639***	2.3051***
(Ethnicity-Weighted)	(0.6201)	(0.7808)	(0.6063)	(0.6162)	(0.6150)	(0.6313)
Predicted Genetic Diversity		1.4945				1.5088
(Ancestry-Adjusted)		(1.3520)				(1.4492)
Ethnic Diversity			-0.2425*			-0.4062**
			(0.1386)			(0.1942)
Linguistic Diversity				0.0307		0.2229
				(0.1216)		(0.1663)
Religious Diversity					0.1542	0.1176
					(0.1352)	(0.1407)
Number of Countries	123	123	123	123	123	123
R^2	0.69	0.69	0.70	0.69	0.69	0.71
Microgeographic Factors	Yes	Yes	Yes	Yes	Yes	Yes
Land Productivity Factors	Yes	Yes	Yes	Yes	Yes	Yes
Climatic Factors	Yes	Yes	Yes	Yes	Yes	Yes
Health Factors	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Island Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Landlock Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
OPEC Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the average number of scientific and technical publications per 1,000 people during 1981 to 2013.
Independent variables: $DRD4^{R2R7}$ refers to the ethnicity-weighted DRD4 exon III 2- and 7-repeat allele frequency measure on the country-level. *Predicted Genetic Diversity* refers to the country's overall genetic diversity (expressed as expected heterozygosity) predicted by migratory distance from East Africa. *Ethnic Diversity* refers to the country's ethnic diversity (0=low and 1=high). *Linguistic Diversity* refers to the country's linguistic diversity (0=low and 1=high). *Religious Diversity* refers to the country's religious diversity (0=low and 1=high). *Microgeographic Factors* include *Absolute Latitude*, ln *Distance to Major Markets*, ln *Distance to Coast or River*, and ln *Hydrocarbons per Person*. *Land Productivity Factors* include ln *Percentage Arable Land Area*, *Agricultural Suitability*, and *Terrain Roughness*. *Climatic Factors* include *Temperature* and *Precipitation*. *Health Factors* include *Population in Tropics (Share)* and *Population in Area with Malaria (Share)*. *Region Fixed Effects* refer to region dummies for Sub-Saharan Africa, North America, Central and South America, East Asia, Central Asia, South Asia, Middle East, and Europe. *Island Dummy* takes value 1 for countries being an island. *OPEC Dummy* takes value 1 for member countries belonging to the Organization of the Petroleum Exporting Countries (OPEC). *Landlock Dummy* takes value 1 for countries being landlocked. Constant term included but not shown.
Robust standard errors are reported between parenthesis.
*: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level.

Table 3: Novelty-Seeking Traits and Number of Scientific and Technical Publications (Historical Factors)

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable: Number of Scientific and Technical Publications per 1,000 People					
$DRD4^{R2R7}$	2.0045***	1.8410***	1.8548***	1.8741***	1.3853***	1.0305***
(Ethnicity-Weighted)	(0.6201)	(0.4772)	(0.4645)	(0.4563)	(0.4029)	(0.3864)
European Population (Share)		0.9702***	0.9617***	0.9778***	0.9247***	0.8298***
		(0.1940)	(0.1845)	(0.1782)	(0.1536)	(0.1374)
Official Language is English			0.0665	-0.0316	-0.0993	-0.1543*
			(0.0441)	(0.0612)	(0.0951)	(0.0893)
Former Colonizer is British				0.1474**	0.0977*	0.1084**
				(0.0623)	(0.0549)	(0.0544)
British Legal Origin					0.2408**	0.2565**
					(0.0975)	(0.1030)
French Legal Origin					0.0900	0.1266
					(0.0722)	(0.0857)
German Legal Origin					0.3226	0.3224*
					(0.2002)	(0.1835)
Scandinavian Legal Origin					0.6832***	0.5495***
					(0.0934)	(0.1499)
Percentage of Protestants						0.0016
						(0.0015)
Percentage of Catholics						0.0002
						(0.0008)
Percentage of Muslims						-0.0018**
						(0.0009)
Number of Countries	123	123	123	123	123	123
R^2	0.69	0.77	0.78	0.79	0.87	0.89
Microgeographic Factors	Yes	Yes	Yes	Yes	Yes	Yes
Land Productivity Factors	Yes	Yes	Yes	Yes	Yes	Yes
Climatic Factors	Yes	Yes	Yes	Yes	Yes	Yes
Health Factors	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Island Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Landlock Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
OPEC Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the average number of scientific and technical publications per 1,000 people during 1981 to 2013.
Independent variables: $DRD4^{R2R7}$ refers to the ethnicity-weighted DRD4 exon III 2- and 7-repeat allele frequency measure on the country-level. *European Population (Share)* is the share of country's current population with European ancestry. *Official Language is English* refers to an indicator variable that takes a value of one if the country's official language is English and zero otherwise. *Former Colonizer is British* refers to an indicator variable that takes a value of one if the country's former colonizer was British and zero otherwise. *Legal Origin Factors* refer to a set of indicator variables that takes a value of one if the country's legal tradition is British, French, German, or Scandinavian, and zero otherwise. *Major Religion Shares* refer to the share of the country's major religions (i.e. Protestants, Catholics, and Muslims) in %. *Microgeographic Factors* include *Absolute Latitude*, ln *Distance to Major Markets*, ln *Distance to Coast or River*, and ln *Hydrocarbons per Person*. *Land Productivity Factors* include ln *Percentage Arable Land Area*, *Agricultural Suitability*, and *Terrain Roughness*. *Climatic Factors* include *Temperature* and *Precipitation*. *Health Factors* include *Population in Tropics (Share)* and *Population in Area with Malaria (Share)*. *Region Fixed Effects* refer to region dummies for Sub-Saharan Africa, North America, Central and South America, East Asia, Central Asia, South Asia, Middle East, and Europe. *Island Dummy* takes value 1 for countries being an island. *OPEC Dummy* takes value 1 for member countries belonging to the Organization of the Petroleum Exporting Countries (OPEC). *Landlock Dummy* takes value 1 for countries being landlocked. Constant term included but not shown.
Robust standard errors are reported between parenthesis.
*: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level.

Table 4: Novelty-Seeking Traits and Number of Scientific and Technical Publications (Technological Frontier Factors)

	Technological Frontier: USA				Technological Frontier: OECD				Technological Frontier: Neo-European			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Dependent Variable: Number of Scientific and Technical Publications per 1,000 People											
$DRD4^{R2R7}$	1.0232**	1.0789***	1.0382***	1.1337***	1.0226**	1.0655***	1.0148***	1.0435***	1.0122**	1.1332***	0.9942***	1.0930***
(Ethnicity-Weighted)	(0.3953)	(0.3783)	(0.3918)	(0.3969)	(0.3927)	(0.3786)	(0.3841)	(0.3799)	(0.3969)	(0.3744)	(0.3758)	(0.3844)
Genetic Distance to the Frontier	-0.1100			-0.1094	-0.1894			-0.1069	-0.2109			-0.2395
	(0.7872)			(0.7533)	(0.7095)			(0.6738)	(0.6740)			(0.6224)
Linguistic Distance to the Frontier		-0.3474		-0.4969**		-0.5351		-0.4795		-0.6059***		-0.5751**
		(0.2397)		(0.2446)		(0.9848)		(0.9588)		(0.2165)		(0.2418)
Religious Distance to the Frontier			0.0748	0.3993			-0.2508	-0.2226			-0.5706	-0.2220
			(0.2459)	(0.3012)			(0.3924)	(0.3846)			(0.3854)	(0.3907)
Number of Countries	123	123	123	123	123	123	123	123	123	123	123	123
R^2	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.90	0.89	0.90
Microgeographic Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Land Productivity Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Climatic Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Health Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Island Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landlock Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
OPEC Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal Origin Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major Religion Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Colonial Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the average number of scientific and technical publications per 1,000 people during 1981 to 2013.

Independent variables: $DRD4^{R2R7}$ refers to the ethnicity-weighted DRD4 exon III 2- and 7-repeat allele frequency measure on the country-level. *Genetic Distance to the Frontier* refers to the country's relative genetic distance to the technological frontier (i.e. U.S., OECD, and Neo-European countries). *Linguistic Distance to the Frontier* refers to the country's relative linguistic distance to the technological frontier (i.e. U.S., OECD, and Neo-European countries). *Religious Distance to the Frontier* refers to the country's relative religious distance to the technological frontier (i.e. U.S., OECD, and Neo-European countries). *Microgeographic Factors* include *Absolute Latitude*, ln *Distance to Major Markets*, ln *Distance to Coast or River*, and ln *Hydrocarbons per Person*. *Land Productivity Factors* include ln *Percentage Arable Land Area*, *Agricultural Suitability*, and *Terrain Roughness*. *Climatic Factors* include *Temperature* and *Precipitation*. *Health Factors* include *Population in Tropics (Share)* and *Population in Area with Malaria (Share)*. *Region Fixed Effects* refer to region dummies for Sub-Saharan Africa, North America, Central and South America, East Asia, Central Asia, South Asia, Middle East, and Europe. *Island Dummy* takes value 1 for countries being an island. *OPEC Dummy* takes value 1 for member countries belonging to the Organization of the Petroleum Exporting Countries (OPEC). *Landlock Dummy* takes value 1 for countries being landlocked. *Legal Origin Factors* refer to a set of indicator variables that takes a value of one if the country's legal tradition is British, French, German, or Scandinavian, and zero otherwise. *Major Religion Shares* refer to the share of the country's major religions (i.e. Protestants, Catholics, and Muslims) in %. *Colonial Factors* include *European Population (Share)*, *Official Language is English*, and *Former Colonizer is British*. Constant term included but not shown.

Robust standard errors are reported between parenthesis.

*: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level.

Table 5: Novelty-Seeking Traits and Number of Scientific and Technical Publications (Technological and Human Capability Factors)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent Variable: Number of Scientific and Technical Publications per 1,000 People							
$DRD4^{R2R7}$	1.2205***	1.2574***	1.1143**	0.7081**	1.1554***	1.2276***	1.2119***	0.7884**
(Ethnicity-Weighted)	(0.4402)	(0.4456)	(0.4396)	(0.3266)	(0.4286)	(0.3255)	(0.4287)	(0.3165)
ln GDP per Capita		0.0384						-0.0541
		(0.0506)						(0.0390)
Human Capital			0.0898					0.0602
			(0.0758)					(0.0572)
Research and Development				0.2361***				0.1538***
(Percentage of GDP)				(0.0355)				(0.0395)
Internet Users					0.0059*			-0.0011
(Per 100 People)					(0.0034)			(0.0026)
Fixed Telephone Subscriptions						0.0166***		0.0136***
(Per 100 People)						(0.0026)		(0.0031)
Mobile Cellular Subscriptions							0.0020	-0.0015
(Per 100 People)							(0.0021)	(0.0025)
Number of Countries	99	99	99	99	99	99	99	99
R^2	0.90	0.90	0.91	0.94	0.91	0.95	0.90	0.96
Microgeographic Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Land Productivity Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Climatic Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Health Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Island Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landlock Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
OPEC Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal Origin Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major Religion Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Colonial Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the average number of scientific and technical publications per 1,000 people during 1981 to 2013.
Independent variables: $DRD4^{R2R7}$ refers to the ethnicity-weighted DRD4 exon III 2- and 7-repeat allele frequency measure on the country-level. *ln GDP per Capita* is the log of a country's GDP per capita averaged across the years 1981 to 2013. *Human Capital* is the country's human capital measure averaged across the years 1981 to 2013. *Research and Development* is the country's flow of resources devoted to research and development activities (as percentage of total GDP) averaged across the years 1981 to 2013. *Internet Users* is the the number of internet users per 100 people averaged across the years 1981 to 2013 *Fixed Telephone Subscriptions* is the the number of fixed telephone subscriptions per 100 people averaged across the years 1981 to 2013. *Mobile Cellular Subscriptions* is the the number of mobile telephone cellular subscriptions per 100 people averaged across the years 1981 to 2013. *Microgeographic Factors* include *Absolute Latitude*, ln *Distance to Major Markets*, ln *Distance to Coast or River*, and ln *Hydrocarbons per Person*. *Land Productivity Factors* include ln *Percentage Arable Land Area*, *Agricultural Suitability*, and *Terrain Roughness*. *Climatic Factors* include *Temperature* and *Precipitation*. *Health Factors* include *Population in Tropics (Share)* and *Population in Area with Malaria (Share)*. *Region Fixed Effects* refer to region dummies for Sub-Saharan Africa, North America, Central and South America, East Asia, Central Asia, South Asia, Middle East, and Europe. *Island Dummy* takes value 1 for countries being an island. *OPEC Dummy* takes value 1 for member countries belonging to the Organization of the Petroleum Exporting Countries (OPEC). *Landlock Dummy* takes value 1 for countries being landlocked. *Legal Origin Factors* refer to a set of indicator variables that takes a value of one if the country's legal tradition is British, French, German, or Scandinavian, and zero otherwise. *Major Religion Shares* refer to the share of the country's major religions (i.e. Protestants, Catholics, and Muslims) in %. *Colonial Factors* include *European Population (Share)*, *Official Language is English*, and *Former Colonizer is British*. Constant term included but not shown.
Robust standard errors are reported between parenthesis.
*: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level.

Table 6: Novelty-Seeking Traits and Number of Scientific and Technical Publications (Social Value Factors)

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Variable: Number of Scientific and Technical Publications per 1,000 People					
$DRD4^{R2R7}$	1.0305***	1.1880***	1.0213**	0.9672**	0.9493**	1.1176***
(Ethnicity-Weighted)	(0.3864)	(0.4032)	(0.3952)	(0.3833)	(0.3857)	(0.4194)
Trade Openness		-0.0012***				-0.0012**
(Percentage of GDP)		(0.0004)				(0.0005)
International Migrant Stock			0.0004			-0.0001
(Percentage of Population)			(0.0021)			(0.0022)
Freedom of Expression and Belief				0.0055		-0.0029
(0 = low to 16 = high)				(0.0055)		(0.0104)
Civil Liberties					-0.0292	-0.0344
(1 = low to 7 = high)					(0.0215)	(0.0424)
Number of Countries	123	123	123	123	123	123
R^2	0.89	0.89	0.89	0.89	0.89	0.89
Microgeographic Factors	Yes	Yes	Yes	Yes	Yes	Yes
Land Productivity Factors	Yes	Yes	Yes	Yes	Yes	Yes
Climatic Factors	Yes	Yes	Yes	Yes	Yes	Yes
Health Factors	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Island Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Landlock Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
OPEC Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Legal Origin Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Major Religion Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Colonial Factors	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the average number of scientific and technical publications per 1,000 people during 1981 to 2013.
Independent variables: $DRD4^{R2R7}$ refers to the ethnicity-weighted DRD4 exon III 2- and 7-repeat allele frequency measure on the country-level. *Trade Openness* is the amount of exports and imports over total GDP averaged across the years 1981 to 2013. *International Migrant Stock* is the country's international migrant stock averaged across the years 1981 to 2013. *Freedom of Expression and Belief* is the country's freedom of expression and belief (0 = low to 16 = high) averaged across the years 1981 to 2013. *Civil Liberties* is the country's extent of civil liberties (1 = low to 7 = high) averaged across the years 1981 to 2013. *Microgeographic Factors* include *Absolute Latitude*, ln *Distance to Major Markets*, ln *Distance to Coast or River*, and ln *Hydrocarbons per Person*. *Land Productivity Factors* include ln *Percentage Arable Land Area*, *Agricultural Suitability*, and *Terrain Roughness*. *Climatic Factors* include *Temperature* and *Precipitation*. *Health Factors* include *Population in Tropics (Share)* and *Population in Area with Malaria (Share)*. *Region Fixed Effects* refer to region dummies for Sub-Saharan Africa, North America, Central and South America, East Asia, Central Asia, South Asia, Middle East, and Europe. *Island Dummy* takes value 1 for countries being an island. *OPEC Dummy* takes value 1 for member countries belonging to the Organization of the Petroleum Exporting Countries (OPEC). *Landlock Dummy* takes value 1 for countries being landlocked. *Legal Origin Factors* refer to a set of indicator variables that takes a value of one if the country's legal tradition is British, French, German, or Scandinavian, and zero otherwise. *Major Religion Shares* refer to the share of the country's major religions (i.e. Protestants, Catholics, and Muslims) in %. *Colonial Factors* include *European Population (Share)*, *Official Language is English*, and *Former Colonizer is British*. Constant term included but not shown.
Robust standard errors are reported between parenthesis.
*: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level.

Table 7: Novelty-Seeking Traits and Number of Scientific and Technical Publications (Quality of Governance Factors)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent Variable: Number of Scientific and Technical Publications per 1,000 People							
$DRD4^{R2R7}$	1.0305***	0.8886**	0.9943**	1.0298**	0.9647**	0.9413**	0.9718**	0.8430**
(Ethnicity-Weighted)	(0.3864)	(0.3919)	(0.3900)	(0.3929)	(0.3993)	(0.3787)	(0.3836)	(0.3816)
Freedom from Corruption		0.0043***						0.0073***
(0 = low to 100 = high)		(0.0015)						(0.0026)
Rule of Law			0.0069					-0.0096
(0 = low to 16 = high)			(0.0058)					(0.0080)
Independence of the Judiciary				0.0578				-0.0133
(0 = low to 2 = high)				(0.0520)				(0.0550)
Property Rights Protection					0.0016			-0.0029
(0 = low to 100 = high)					(0.0012)			(0.0019)
Executive Constraints						0.0288*		0.0190
(1 = low to 7 = high)						(0.0165)		(0.0423)
Institutionalized Democracy							0.0168*	0.0162
(0 = low to 10 = high)							(0.0092)	(0.0224)
Number of Countries	123	123	123	123	123	123	123	123
R^2	0.89	0.90	0.89	0.89	0.89	0.89	0.89	0.90
Microgeographic Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Land Productivity Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Climatic Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Health Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Island Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landlock Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
OPEC Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal Origin Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major Religion Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Colonial Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the average number of scientific and technical publications per 1,000 people during 1981 to 2013.
Independent variables: $DRD4^{R2R7}$ refers to the ethnicity-weighted DRD4 exon III 2- and 7-repeat allele frequency measure on the country-level. *Freedom from Corruption* is the country's freedom from corruption averaged across the years 1981 to 2013. *Rule of Law* is the country's rule of law averaged across the years 1981 to 2013. *Independence of the Judiciary* is the country's independence of the judiciary averaged across the years 1981 to 2013. *Property Rights Protection* is the country's property rights protection averaged across the years 1981 to 2013. *Executive Constraints* is the country's constraint of executive power of political parties and leaders averaged across the years 1981 to 2013. *Institutionalized Democracy* is the country's extent of institutionalized democracy averaged across the years 1981 to 2013. *Microgeographic Factors* include *Absolute Latitude*, ln *Distance to Major Markets*, ln *Distance to Coast or River*, and ln *Hydrocarbons per Person*. *Land Productivity Factors* include ln *Percentage Arable Land Area*, *Agricultural Suitability*, and *Terrain Roughness*. *Climatic Factors* include *Temperature* and *Precipitation*. *Health Factors* include *Population in Tropics (Share)* and *Population in Area with Malaria (Share)*. *Region Fixed Effects* refer to region dummies for Sub-Saharan Africa, North America, Central and South America, East Asia, Central Asia, South Asia, Middle East, and Europe. *Island Dummy* takes value 1 for countries being an island. *OPEC Dummy* takes value 1 for member countries belonging to the Organization of the Petroleum Exporting Countries (OPEC). *Landlock Dummy* takes value 1 for countries being landlocked. *Legal Origin Factors* refer to a set of indicator variables that takes a value of one if the country's legal tradition is British, French, German, or Scandinavian, and zero otherwise. *Major Religion Shares* refer to the share of the country's major religions (i.e. Protestants, Catholics, and Muslims) in %. *Colonial Factors* include *European Population (Share)*, *Official Language is English*, and *Former Colonizer is British*. Constant term included but not shown.
Robust standard errors are reported between parenthesis.
*: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level.

Table 8: Novelty-Seeking Traits and Number of Scientific and Technical Publications (Instrumental Variables Estimation)

	Full Sample		Technological and Human Capability Sample								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	OLS	IV	OLS	IV	IV	IV	IV	IV	IV	IV	IV
	Dependent Variable: Number of Scientific and Technical Publications per 1,000 People										
$DRD4^{R2R7}$	1.0305***	1.9532***	1.2205***	1.8927***	1.9066***	1.7224***	1.2919***	1.8237***	1.4684***	1.8974***	1.0447**
(Ethnicity-Weighted)	(0.3864)	(0.5618)	(0.4402)	(0.5455)	(0.5694)	(0.5900)	(0.4718)	(0.5147)	(0.3558)	(0.5517)	(0.4315)
ln GDP per Capita					0.0460						-0.0486
					(0.0452)						(0.0309)
Human Capital						0.0693					0.0491
						(0.0668)					(0.0472)
Research and Development							0.2211***				0.1449***
(Percentage of GDP)							(0.0327)				(0.0370)
Internet Users								0.0055**			-0.0014
(Per 100 People)								(0.0027)			(0.0021)
Fixed Telephone Subscriptions									0.0166***		0.0139***
(Per 100 People)									(0.0021)		(0.0024)
Mobile Cellular Subscriptions										0.0020	-0.0014
(Per 100 People)										(0.0017)	(0.0019)
Number of Countries	123	123	99	99	99	99	99	99	99	99	99
R^2	0.89	0.88	0.90	0.90	0.90	0.90	0.94	0.90	0.95	0.90	0.96
Kleibergen-Paap *rk* LM Statistic		19.2620		19.3448	19.1855	19.5791	20.1405	19.7482	19.7232	19.7673	18.9300
$p-value$		0.0074		0.0072	0.0076	0.0066	0.0053	0.0061	0.0062	0.0061	0.0084
Hansen J Statistic		3.3469		3.1794	3.2098	3.4991	5.0646	3.1436	5.5922	3.7051	6.7405
$p-value$		0.7642		0.7860	0.7821	0.7441	0.5356	0.7906	0.4704	0.7165	0.3455
Microgeographic Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Land Productivity Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Climatic Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Health Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Island Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Landlock Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
OPEC Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Legal Origin Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major Religion Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Colonial Factors	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the average number of scientific and technical publications per 1,000 people during 1981 to 2013.

Independent variables: $DRD4^{R2R7}$ refers to the ethnicity-weighted DRD4 exon III 2- and 7-repeat allele frequency measure on the country-level. *ln GDP per Capita* is the log of a country's GDP per capita averaged across the years 1981 to 2013. *Human Capital* is the country's human capital measure averaged across the years 1981 to 2013. *Research and Development* is the country's flow of resources devoted to research and development activities (as percentage of total GDP) averaged across the years 1981 to 2013. *Internet Users* is the the number of internet users per 100 people averaged across the years 1981 to 2013 *Fixed Telephone Subscriptions* is the the number of fixed telephone subscriptions per 100 people averaged across the years 1981 to 2013. *Mobile Cellular Subscriptions* is the the number of mobile telephone cellular subscriptions per 100 people averaged across the years 1981 to 2013. *Microgeographic Factors* include *Absolute Latitude*, ln *Distance to Major Markets*, ln *Distance to Coast or River*, and ln *Hydrocarbons per Person*. *Land Productivity Factors* include ln *Percentage Arable Land Area*, *Agricultural Suitability*, and *Terrain Roughness*. *Climatic Factors* include *Temperature* and *Precipitation*. *Health Factors* include *Population in Tropics (Share)* and *Population in Area with Malaria (Share)*. *Region Fixed Effects* refer to region dummies for Sub-Saharan Africa, North America, Central and South America, East Asia, Central Asia, South Asia, Middle East, and Europe. *Island Dummy* takes value 1 for countries being an island. *OPEC Dummy* takes value 1 for member countries belonging to the Organization of the Petroleum Exporting Countries (OPEC). *Landlock Dummy* takes value 1 for countries being landlocked. *Legal Origin Factors* refer to a set of indicator variables that takes a value of one if the country's legal tradition is British, French, German, or Scandinavian, and zero otherwise. *Major Religion Shares* refer to the share of the country's major religions (i.e. Protestants, Catholics, and Muslims) in %. *Colonial Factors* include *European Population (Share)*, *Official Language is English*, and *Former Colonizer is British*. The set of excluded instruments refer to a vector of ancestry-adjusted country-level controls, namely migratory distance from East Africa, pasture land (i.e. pasture and pasture squared), elevation (elevation and elevation squared), and land suitability for agriculture (agricultural suitability and agricultural suitability squared) factors. Constant term included but not shown.

Robust standard errors are reported between parenthesis.

*: Significant at the 10% level. **: Significant at the 5% level. ***: Significant at the 1% level.

C Descriptive Statistics

Table 9: Number of Scientific and Technical Publications per 1,000 People Across Continents

Variable	N	Mean	SD	Minimum	Maximum
Scientific Publications per 1,000 People	171	0.1831	0.3259	0.0004	1.4848
Africa	50	0.0131	0.0221	0.0004	0.1156
Americas	33	0.0904	0.2379	0.0008	1.0704
Asia	45	0.1106	0.2381	0.0005	1.2784
Europe	40	0.5157	0.3899	0.0139	1.4848
Oceania	3	0.6915	0.5641	0.0434	1.0717

Notes: This table shows basic summary statistics for the number of scientific and technical publications per 1,000 people in the years 1981 to 2013 across continents. It refers to scientific publications that have been published in the fields of physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences. See the main text for additional details regarding data construction and sources.

Table 10: Summary Statistics for the 123-Country Sample

Variable	N	Mean	SD	Minimum	Maximum
Main Dependent Variable					
Scientific Publications per 1,000 people	123	0.2144	0.3586	0.0004	1.4848
Main Explanatory Variable					
$DRD4^{R2R7}$	123	0.2335	0.0658	0.0980	0.4583
Biogeographic Variables					
Absolute Latitude	123	28.6804	17.2200	0.4221	67.4700
ln Distance to Major Markets	123	1.0987	0.9207	-1.9661	2.2322
ln Distance to Coast or River	123	-1.7751	1.2219	-4.8343	0.8694
ln Hydrocarbons per Person	123	0.9175	4.6078	-4.6052	10.5947
ln Percentage Arable Land Area	123	2.2827	1.1215	-2.3010	4.1612
Agricultural Suitability	123	0.4425	0.2517	0	0.9538
Terrain Roughness	123	0.1839	0.1363	0.0127	0.6022
Temperature	123	17.4371	8.5674	-7.9294	28.6391
Precipitation	123	86.9544	58.9073	2.9106	241.7184
Population in Tropics (Share)	123	0.2700	0.3935	0	1
Population in Area with Malaria (Share)	123	0.3373	0.4228	0	1
Diversity Variables					
Predicted Genetic Diversity	123	0.7255	0.0266	0.6279	0.7743
Ethnic Diversity	123	0.4474	0.2488	0.0020	0.9302
Linguistic Diversity	123	0.3918	0.2852	0.0021	0.9227
Religious Diversity	123	0.4245	0.2339	0.0035	0.8603
Historical Variables					
European Population (Share)	123	0.3536	0.4262	0	1
Official Language is English	123	0.1951	0.3979	0	1
Former Colonizer is British	123	0.2439	0.4312	0	1
Percentage of Protestants	123	10.652	19.9559	0	97.8000
Percentage of Catholics	123	29.9122	36.3416	0	96.9000
Percentage of Muslims	123	24.7246	36.3853	0	99.4000
British Legal Origin	123	0.2602	0.4405	0	1
French Legal Origin	123	0.4390	0.4983	0	1
German Legal Origin	123	0.0325	0.1781	0	1
Scandinavian Legal Origin	123	0.0325	0.1781	0	1
Technological Frontier Variables					
Genetic Distance to the Frontier (U.S.)	123	0.0902	0.0515	0	0.2057
Linguistic Distance to the Frontier (U.S.)	123	0.9413	0.1404	0	1
Religious Distance to the Frontier (U.S.)	123	0.8098	0.1462	0	0.9990
Genetic Distance to the Frontier (OECD)	123	0.0870	0.0589	0.0256	0.2137
Linguistic Distance to the Frontier (OECD)	123	0.9673	0.0412	0.8637	1
Religious Distance to the Frontier (OECD)	123	0.8160	0.1272	0.5739	0.9917
Genetic Distance to the Frontier (Neo-Europe)	123	0.0857	0.0616	0.0173	0.2264
Linguistic Distance to the Frontier (Neo-Europe)	123	0.9503	0.1181	0.3407	1
Religious Distance to the Frontier (Neo-Europe)	123	0.8307	0.1146	0.6395	0.9979
Social Value Variables					
Trade Openness	123	73.6411	32.5446	6.5773	163.2497
International Migrant Stock	123	7.0488	10.8809	0.0395	73.9963
Freedom of Expression and Belief	123	11.2919	4.1409	1	16
Civil Liberties	123	3.6614	1.6488	1	6.9545
Quality of Governance Variables					
Freedom from Corruption	123	39.834	21.8374	11	95
Rule of Law	123	8.1593	4.4230	0	16
Independence of the Judiciary	123	1.0873	0.6165	0	2
Property Rights Protection	123	47.8069	21.6783	10	91.3889
Executive Constraints	123	4.7393	1.7953	1	7
Institutionalized Democracy	123	5.1494	3.4784	0	10

Table 11: Pairwise Correlations for the 123-Country Sample

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)
(1) Scientific Publications per 1,000 people	1.0000																						
(2) $DRD4^{R2R7}$	0.0397	1.0000																					
(3) Absolute Latitude	0.6322	-0.3807	1.0000																				
(4) ln Distance to Major Markets	-0.6674	0.0830	-0.6527	1.0000																			
(5) ln Distance to Coast or River	-0.2802	-0.2762	-0.0646	0.3654	1.0000																		
(6) ln Hydrocarbons per Person	0.0568	-0.0522	0.1252	-0.0478	0.1011	1.0000																	
(7) ln Percentage Arable Land Area	0.1564	0.0329	0.2310	-0.3550	-0.2571	-0.2026	1.0000																
(8) Agricultural Suitability	-0.0713	0.1888	-0.0250	-0.1472	-0.3966	-0.2854	0.6168	1.0000															
(9) Terrain Roughness	-0.0411	-0.0512	0.0354	0.0068	-0.0537	-0.0487	0.0872	0.2965	1.0000														
(10) Temperature	-0.5647	0.3257	-0.9023	0.5756	-0.0910	-0.0847	-0.2059	-0.0169	-0.2073	1.0000													
(11) Precipitation	-0.1658	0.5596	-0.5254	0.1302	-0.4095	-0.2021	0.1023	0.3810	0.1500	0.3959	1.0000												
(12) Population in Tropics (Share)	-0.3791	0.4097	-0.7215	0.3386	-0.2227	-0.1856	0.0586	0.1789	-0.0794	0.6284	0.7574	1.0000											
(13) Population in Area with Malaria (Share)	-0.4547	0.3117	-0.7065	0.4767	0.2015	-0.2649	-0.0735	-0.0536	-0.2623	0.6625	0.3393	0.5910	1.0000										
(14) Predicted Genetic Diversity	0.0782	-0.5022	0.1811	-0.0215	0.1625	-0.0394	0.0972	-0.1355	-0.1742	-0.1050	-0.5031	-0.1778	0.0517	1.0000									
(15) Ethnic Diversity	-0.4259	0.1870	-0.5719	0.4630	0.3144	-0.0452	-0.2259	-0.2233	-0.1486	0.4524	0.1352	0.4059	0.5904	0.1159	1.0000								
(16) Linguistic Diversity	-0.2639	0.0480	-0.3948	0.3498	0.3580	-0.1523	-0.0428	-0.1834	-0.2186	0.3156	0.0136	0.3180	0.5910	0.2083	0.7247	1.0000							
(17) Religious Diversity	0.0908	0.0244	-0.0613	0.0292	0.0055	-0.0146	0.0622	0.0934	-0.0530	-0.0572	0.0515	0.0879	0.1122	0.3068	0.2233	0.2742	1.0000						
(18) European Population (Share)	0.6507	0.1170	0.6811	-0.5929	-0.3484	0.0287	0.3014	0.2333	-0.0242	-0.6303	-0.0938	-0.3919	-0.5796	-0.0285	-0.4641	-0.4345	-0.0091	1.0000					
(19) Official Language is English	0.0543	0.1057	-0.1765	0.1288	0.0181	-0.0921	0.0080	-0.1034	-0.1258	0.1188	0.0955	0.1732	0.1966	0.1826	0.2180	0.2780	0.4459	-0.1325	1.0000				
(20) Former Colonizer is British	0.0256	-0.0410	-0.2062	0.2203	0.0028	0.0200	-0.0527	-0.1736	-0.1476	0.2200	0.0258	0.1548	0.1137	0.1981	0.2076	0.1888	0.3189	-0.2498	0.6758	1.0000			
(21) Percentage of Protestants	0.5274	0.0775	0.3001	-0.1932	-0.0609	-0.0239	-0.0466	-0.2163	-0.1901	-0.2891	-0.0564	-0.1060	-0.1027	0.1491	-0.0943	-0.0211	0.2571	0.2680	0.3107	0.1347	1.0000		
(22) Percentage of Catholics	0.0826	0.4777	-0.1204	-0.1071	-0.2538	-0.0293	0.0497	0.3277	0.1530	0.0345	0.3373	0.1143	-0.1169	-0.4193	-0.1104	-0.2819	-0.0774	0.3906	-0.0512	-0.1943	-0.1176	1.0000	
(23) Percentage of Muslims	-0.3167	-0.4531	-0.0971	0.2093	0.2752	0.2524	-0.2229	-0.4038	-0.1179	0.2634	-0.3877	-0.1132	0.1048	0.2694	0.1773	0.1624	-0.3336	-0.4828	-0.1871	0.0575	-0.3160	-0.5008	1.0000
(24) British Legal Origin	0.0446	0.0042	-0.2186	0.2105	0.0084	-0.0228	-0.0046	-0.1505	-0.1237	0.2014	0.0892	0.1911	0.1786	0.1640	0.2258	0.2765	0.3304	-0.2298	0.7835	0.7852	0.2010	-0.2056	-0.0433
(25) French Legal Origin	-0.2569	0.2709	-0.3628	0.1483	-0.0539	0.0430	-0.1917	0.0290	0.0003	0.4149	0.0881	0.1321	0.1786	-0.1751	0.1406	-0.0418	-0.3079	-0.1295	-0.3942	-0.3880	-0.3024	0.4131	0.1938
(26) German Legal Origin	0.2926	-0.0640	0.1420	-0.3452	-0.1388	-0.0990	0.0584	0.0597	0.1267	-0.1749	0.0646	-0.1263	-0.1469	-0.0993	-0.2105	-0.1373	0.1035	0.0599	-0.0903	-0.1041	0.0466	0.0335	-0.1240
(27) Scandinavian Legal Origin	0.4856	-0.0150	0.3636	-0.1992	-0.0498	0.0032	-0.0169	-0.2087	-0.1293	-0.3089	-0.0840	-0.1263	-0.1469	0.0300	-0.2697	-0.1705	-0.1520	0.2776	-0.0903	-0.1041	0.7193	-0.1485	-0.1246
(28) Genetic Distance to the Frontier (U.S.)	-0.4835	0.1658	-0.6180	0.5418	0.2922	-0.2749	-0.1832	-0.0484	-0.1580	0.4613	0.2087	0.4329	0.6446	0.0835	0.5103	0.5197	0.3004	-0.6076	0.3617	0.1938	0.1019	-0.0852	-0.0692
(29) Linguistic Distance to the Frontier (U.S.)	-0.4092	-0.1245	-0.1504	0.2817	0.1613	-0.1399	-0.0310	0.0615	-0.0589	0.1610	-0.0782	0.0345	0.2626	0.1005	0.1092	0.2318	-0.2707	-0.2980	-0.4056	-0.2747	-0.2793	-0.1328	0.2424
(30) Religious Distance to the Frontier (U.S.)	-0.4377	-0.3445	-0.2086	0.3937	0.2316	-0.0296	-0.1274	-0.2154	-0.1191	0.2519	-0.1287	0.1152	0.3424	0.2793	0.2362	0.3587	-0.0031	-0.5882	-0.1223	0.0612	-0.2863	-0.6965	0.4511
(31) Genetic Distance to the Frontier (OECD)	-0.4570	0.2174	-0.6573	0.5100	0.2596	-0.2900	-0.1677	-0.0543	-0.2058	0.5133	0.2361	0.4887	0.6881	0.1206	0.5536	0.5441	0.3323	-0.5970	0.4176	0.2304	0.1126	-0.0807	-0.0647
(32) Linguistic Distance to the Frontier (OECD)	-0.2521	-0.2655	-0.0764	0.1871	0.3007	-0.0915	0.0021	-0.1428	-0.2168	0.1113	-0.2752	0.0012	0.3633	0.4138	0.1950	0.4883	0.0703	-0.5088	-0.0131	0.0697	-0.0853	-0.6113	0.3856
(33) Religious Distance to the Frontier (OECD)	-0.3192	-0.3997	-0.1381	0.2630	0.3025	-0.0080	-0.1001	-0.2838	-0.1699	0.1734	-0.2044	0.0782	0.3416	0.3914	0.2815	0.4206	0.1524	-0.6019	0.0207	0.1966	-0.1257	-0.8708	0.4709
(34) Genetic Distance to the Frontier (Neo-Europe)	-0.4583	0.1654	-0.6225	0.5165	0.2934	-0.2821	-0.1802	-0.0539	-0.1823	0.4676	0.1970	0.4380	0.6535	0.1250	0.5359	0.5314	0.3357	-0.5928	0.4037	0.2269	0.1274	-0.0808	-0.0728
(35) Linguistic Distance to the Frontier (Neo-Europe)	-0.4320	-0.0969	-0.1725	0.2473	0.2069	-0.1363	-0.0300	0.0896	-0.0576	0.1699	-0.0741	0.0288	0.2756	0.0535	0.1258	0.2238	-0.2910	-0.2981	-0.4372	-0.2926	-0.2976	-0.0882	0.2385
(36) Religious Distance to the Frontier (Neo-Europe)	-0.4175	-0.3771	-0.2079	0.3165	0.3087	-0.0036	-0.1228	-0.2566	-0.1432	0.2441	-0.1688	0.1110	0.3689	0.3205	0.2912	0.4058	0.0671	-0.6535	-0.0433	0.1605	-0.2496	-0.8101	0.5015
(37) Trade Openness	0.0613	-0.0425	0.1910	-0.1203	-0.2291	-0.0940	-0.1706	-0.0748	-0.0927	-0.1538	-0.0694	-0.0459	-0.1844	0.0956	-0.0631	0.0125	0.0586	0.1038	-0.0095	-0.0280	0.1046	-0.0324	0.0122
(38) International Migrant Stock	0.2641	-0.1922	0.1809	-0.0622	-0.1254	0.2682	-0.3430	-0.2855	-0.0985	-0.0587	-0.3201	-0.2252	-0.2418	0.2193	0.0278	-0.0245	0.1056	0.0640	-0.0334	0.2654	0.0122	-0.1350	0.2347
(39) Freedom of Expression and Belief	0.5007	0.3335	0.2353	-0.3614	-0.4009	-0.2266	0.1573	0.2270	0.0387	-0.2535	0.1663	-0.0656	-0.2668	-0.1676	-0.1943	-0.2135	0.1039	0.5862	0.2164	-0.0140	0.3839	0.4712	-0.5971
(40) Civil Liberties	0.5630	-0.7067	-0.2153	-0.4436	0.5361	0.4528	0.1230	-0.1917	-0.1713	-0.0581	0.4200	-0.0778	0.2069	0.4544	0.1467	0.3592	0.3348	-0.0613	-0.7184	-0.1621	-0.0083	-0.4259	-0.4249
(41) Freedom from Corruption	0.8433	0.0593	0.5206	-0.5551	-0.3333	0.1357	-0.0903	-0.1906	-0.1078	-0.4359	-0.1591	-0.3511	-0.4780	0.0129	-0.4207	-0.3350	0.0863	0.5605	0.1326	0.1133	0.5510	0.1231	-0.2776
(42) Rule of Law	0.7060	0.1351	0.4904	-0.5299	-0.3846	-0.1411	0.1555	0.1401	-0.0212	-0.4641	-0.0113	-0.2810	-0.4196	-0.0158	-0.3658	-0.3002	0.1131	0.6862	0.1592	0.0048	0.4485	0.3048	-0.5158
(43) Independence of the Judiciary	0.6659	0.1003	0.3525	-0.4388	-0.3977	-0.0578	0.0110	-0.0135	-0.0456	-0.2790	-0.0360	-0.2316	-0.3932	0.0278	-0.3063	-0.2918	0.1072	0.5220	0.2997	0.2410	0.4894	0.2556	-0.3918
(44) Property Rights Protection	0.7770	0.0806	0.4781	-0.5288	-0.3469	0.0776	0.0067	-0.1200	-0.0753	-0.4127	-0.1352	-0.3165	-0.4473	0.0086	-0.3863	-0.3203	0.0832	0.5262	0.1896	0.1691	0.4890	0.1682	-0.3030
(45) Executive Constraints	0.6107	0.2279	0.4106	-0.4742	-0.4571	-0.1732	0.2375	0.2621	0.2174	-0.4221	0.1456	-0.1927	-0.4664	-0.2341	-0.3672	-0.3595	0.0538	0.6607	0.1429	0.0205	0.3584	0.3943	-0.6291
(46) Institutionalized Democracy	0.6347	0.2353	0.4133	-0.5012	-0.4526	-0.1381	0.2508	0.2608	0.1713	-0.4176	0.1661	-0.1690	-0.4587	-0.2204	-0.3896	-0.3755	0.0402	0.6970	0.1491	-0.0266	0.3778	0.4289	-0.6195

Table 11: *Continued*

Variable	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)	(33)	(34)	(35)	(36)	(37)	(38)	(39)	(40)	(41)	(42)	(43)	(44)	(45)	(46)
(1) Scientific Publications per 1,000 people																							
(2) $DRD4^{R2R7}$																							
(3) Absolute Latitude																							
(4) ln Distance to Major Markets																							
(5) ln Distance to Coast or River																							
(6) ln Hydrocarbons per Person																							
(7) ln Percentage Arable Land Area																							
(8) Agricultural Suitability																							
(9) Terrain Roughness																							
(10) Temperature																							
(11) Precipitation																							
(12) Population in Tropics (Share)																							
(13) Population in Area with Malaria (Share)																							
(14) Predicted Genetic Diversity																							
(15) Ethnic Diversity																							
(16) Linguistic Diversity																							
(17) Religious Diversity																							
(18) European Population (Share)																							
(19) Official Language is English																							
(20) Former Colonizer is British																							
(21) Percentage of Protestants																							
(22) Percentage of Catholics																							
(23) Percentage of Muslims																							
(24) British Legal Origin	1.0000																						
(25) French Legal Origin	-0.5246	1.0000																					
(26) German Legal Origin	-0.1087	-0.1622	1.0000																				
(27) Scandinavian Legal Origin	-0.1087	-0.1622	-0.0336	1.0000																			
(28) Genetic Distance to the Frontier (U.S.)	0.2452	0.0086	-0.0512	-0.1337	1.0000																		
(29) Linguistic Distance to the Frontier (U.S.)	-0.3241	0.1398	0.0211	0.0131	0.2815	1.0000																	
(30) Religious Distance to the Frontier (U.S.)	0.0620	-0.1495	-0.0008	-0.1946	0.3172	0.5658	1.0000																
(31) Genetic Distance to the Frontier (OECD)	0.2889	0.0330	-0.0751	-0.1434	0.9803	0.1919	0.2596	1.0000															
(32) Linguistic Distance to the Frontier (OECD)	0.0472	-0.1998	0.0138	-0.0214	0.4241	0.6135	0.6856	0.3891	1.0000														
(33) Religious Distance to the Frontier (OECD)	0.1994	-0.2815	0.0022	-0.1482	0.2991	0.2853	0.8817	0.2891	0.7192	1.0000													
(34) Genetic Distance to the Frontier (Neo-Europe)	0.2805	0.0115	-0.0613	-0.1342	0.9918	0.2124	0.2621	0.9928	0.4000	0.2878	1.0000												
(35) Linguistic Distance to the Frontier (Neo-Europe)	-0.3551	0.1817	0.0126	0.0031	0.2594	0.9557	0.4421	0.1901	0.5882	0.2467	0.2107	1.0000											
(36) Religious Distance to the Frontier (Neo-Europe)	0.1567	-0.2084	-0.0073	-0.2240	0.3013	0.3633	0.9103	0.2861	0.7184	0.9837	0.2836	0.3351	1.0000										
(37) Trade Openness	0.0248	-0.2408	-0.0408	-0.0111	-0.0057	0.1550	0.0797	-0.0211	0.1639	0.0357	-0.0190	0.1109	0.0216	1.0000									
(38) International Migrant Stock	0.1233	-0.0838	0.0329	-0.0102	-0.2897	-0.0861	0.0335	-0.2763	0.0187	0.0785	-0.2754	-0.1020	0.0546	0.2729	1.0000								
(39) Freedom of Expression and Belief	0.0524	-0.0078	0.1482	0.2093	-0.1557	-0.3462	-0.5707	-0.1191	-0.4700	-0.5687	-0.1253	-0.3473	-0.6127	0.0187	-0.0928	1.0000							
(40) Civil Liberties	-0.0672	0.1010	-0.2300	-0.2910	0.3843	0.4366	0.6133	0.3545	0.5369	0.5876	0.3568	0.4411	0.6433	-0.0386	-0.0462	-0.8964	1.0000						
(41) Freedom from Corruption	0.1430	-0.1854	0.2767	0.4427	-0.4133	-0.4245	-0.4360	-0.3896	-0.3304	-0.3468	-0.3914	-0.4496	-0.4294	0.1360	0.3633	0.5596	-0.7429	1.0000					
(42) Rule of Law	0.0555	-0.1514	0.2629	0.3118	-0.3130	-0.3354	-0.5025	-0.2868	-0.3781	-0.4617	-0.2848	-0.3420	-0.5258	0.0675	0.0470	0.8876	-0.9247	0.7745	1.0000				
(43) Independence of the Judiciary	0.2789	-0.1377	0.2147	0.2725	-0.2668	-0.4052	-0.4557	-0.2441	-0.3813	-0.4008	-0.2370	-0.4274	-0.4623	0.0196	0.1846	0.6816	-0.8227	0.7989	0.7913	1.0000			
(44) Property Rights Protection	0.1972	-0.1719	0.3090	0.3466	-0.3723	-0.4281	-0.4520	-0.3504	-0.3177	-0.3593	-0.3483	-0.4466	-0.4381	0.1059	0.3048	0.6410	-0.7990	0.9281	0.8070	0.8331	1.0000		
(45) Executive Constraints	0.0758	-0.1546	0.1918	0.2318	-0.3610	-0.4056	-0.5393	-0.3566	-0.5218	-0.5308	-0.3498	-0.4162	-0.5749	0.0122	-0.0562	0.8450	-0.9081	0.6001	0.8248	0.7409	0.6666	1.0000	
(46) Institutionalized Democracy	0.0677	-0.1204	0.2063	0.2567	-0.3800	-0.4352	-0.5893	-0.3651	-0.5625	-0.5796	-0.3614	-0.4448	-0.6259	-0.0053	-0.0853	0.8582	-0.9317	0.6296	0.8475	0.7721	0.6984	0.9685	1.0000

Table 12: Summary Statistics for the 99-Country Sample

Variable	N	Mean	SD	Minimum	Maximum
Main Dependent Variable					
Scientific Publications per 1,000 people	99	0.2610	0.3855	0.0005	1.4848
Main Explanatory Variable					
$DRD4^{R2R7}$	99	0.2393	0.0669	0.0980	0.4583
Biogeographic Variables					
Absolute Latitude	99	30.1741	17.5611	0.4221	67.4700
ln Distance to Major Markets	99	1.0147	0.9895	-1.9661	2.2322
ln Distance to Coast or River	99	-1.8420	1.2287	-4.8343	0.8694
ln Hydrocarbons per Person	99	1.1374	4.4913	-4.6052	10.5947
ln Percentage Arable Land Area	99	2.3196	1.0535	-0.5777	4.0291
Agricultural Suitability	99	0.4386	0.2484	0	0.9538
Terrain Roughness	99	0.1880	0.1310	0.0127	0.6022
Temperature	99	16.6067	8.7334	-7.9294	28.6391
Precipitation	99	85.8136	57.0680	2.9106	241.7184
Population in Tropics (Share)	99	0.2418	0.3738	0	1
Population in Area with Malaria (Share)	99	0.2872	0.3982	0	1
Historical Variables					
European Population (Share)	99	0.4060	0.4370	0	1
Official Language is English	99	0.2020	0.4036	0	1
Former Colonizer is British	99	0.2626	0.4423	0	1
Percentage of Protestants	99	11.9485	21.6261	0	97.8000
Percentage of Catholics	99	32.6859	37.5080	0	96.9000
Percentage of Muslims	99	20.0638	34.6720	0	99.4000
British Legal Origin	99	0.2727	0.4476	0	1
French Legal Origin	99	0.4141	0.4951	0	1
German Legal Origin	99	0.0404	0.1979	0	1
Scandinavian Legal Origin	99	0.0404	0.1979	0	1
Technological and Human Capability Variables					
ln GDP per Capita	99	8.9598	1.1225	6.5467	11.4765
Human Capital	99	2.3822	0.6754	1.0768	3.5365
Research and Development (Percentage of GDP)	99	0.7616	0.8702	0.0371	3.8256
Internet Users	99	18.3261	15.2321	0.2774	53.0951
Fixed Telephone Subscriptions	99	18.0103	17.7603	0.2342	62.5126
Mobile Cellular Subscriptions	99	32.2877	16.7262	0.8136	73.9805

Table 13: Pairwise Correlations for the 99-Country Sample

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)
(1) Scientific Publications per 1,000 people	1.0000																												
(2) $DRD4^{R2R7}$	-0.0050	1.0000																											
(3) Absolute Latitude	0.6552	-0.3761	1.0000																										
(4) ln Distance to Major Markets	-0.6610	0.1186	-0.6524	1.0000																									
(5) ln Distance to Coast or River	-0.2817	-0.2746	-0.0660	0.3581	1.0000																								
(6) ln Hydrocarbons per Person	0.0375	-0.0100	0.1110	-0.0528	0.0538	1.0000																							
(7) ln Percentage Arable Land Area	0.1724	0.0428	0.2213	-0.3880	-0.2928	-0.1939	1.0000																						
(8) Agricultural Suitability	-0.0768	0.2332	-0.0316	-0.1568	-0.3725	-0.2842	0.6182	1.0000																					
(9) Terrain Roughness	-0.0820	-0.0050	-0.0743	0.0707	-0.0111	-0.0724	0.0468	0.2181	1.0000																				
(10) Temperature	-0.5726	0.3459	-0.8977	0.5657	-0.1235	-0.0618	-0.1642	0.0105	-0.1194	1.0000																			
(11) Precipitation	-0.1782	0.6244	-0.5255	0.1285	-0.3898	-0.1925	0.0474	0.3526	0.2538	0.4019	1.0000																		
(12) Population in Tropics (Share)	-0.4030	0.4395	-0.7340	0.3528	-0.1945	-0.1676	0.0385	0.1377	0.0287	0.6346	0.7390	1.0000																	
(13) Population in Area with Malaria (Share)	-0.4685	0.3340	-0.6818	0.4524	0.1938	-0.2563	0.0200	0.0108	-0.1331	0.6293	0.3612	0.6122	1.0000																
(14) European Population (Share)	0.6524	0.0702	0.7020	-0.5638	-0.3282	0.0373	0.3229	0.2674	-0.0795	-0.6315	-0.0965	-0.4194	-0.5673	1.0000															
(15) Official Language is English	0.0558	0.0673	-0.1587	0.1120	0.0243	-0.0164	-0.0028	-0.1458	-0.1264	0.1189	0.0278	0.1977	0.1759	-0.1381	1.0000														
(16) Former Colonizer is British	0.0064	-0.0648	-0.2310	0.2301	0.0271	0.0727	-0.1450	-0.2679	-0.1594	0.2536	-0.0577	0.1893	0.1279	-0.2936	0.6716	1.0000													
(17) Percentage of Protestants	0.5284	0.0298	0.3474	-0.2049	-0.0698	-0.0259	-0.0742	-0.2569	-0.2185	-0.3121	-0.0835	-0.1153	-0.1396	0.2788	0.2822	0.0986	1.0000												
(18) Percentage of Catholics	0.0472	0.4669	-0.1253	-0.0823	-0.2332	-0.0785	0.0530	0.3381	0.1463	0.0483	0.3816	0.1287	-0.0692	0.3778	-0.0488	-0.2266	-0.1699	1.0000											
(19) Percentage of Muslims	-0.2983	-0.3803	-0.1181	0.1867	0.2324	0.2887	-0.2012	-0.4276	-0.0971	0.2600	-0.4223	-0.1239	0.0625	-0.4651	-0.1576	0.1249	-0.2877	-0.4619	1.0000										
(20) British Legal Origin	0.0384	-0.0235	-0.2215	0.2105	0.0472	0.0456	-0.0642	-0.2343	-0.1142	0.2160	-0.0214	0.1954	0.1789	-0.2568	0.7652	0.7684	0.1658	-0.2268	0.0008	1.0000									
(21) French Legal Origin	-0.2650	0.2703	-0.3272	0.1425	-0.0566	0.0431	-0.1397	0.0749	0.0547	0.3732	0.1354	0.0828	0.1009	-0.0958	-0.3720	-0.3620	-0.3157	0.4431	0.2136	-0.5149	1.0000								
(22) German Legal Origin	0.2800	-0.0882	0.1384	-0.3423	-0.1434	-0.1238	0.0625	0.0710	0.1414	-0.1726	0.0788	-0.1334	-0.1487	0.0407	-0.1032	-0.1225	0.0358	0.0211	-0.1180	-0.1257	-0.1725	1.0000							
(23) Scandinavian Legal Origin	0.4811	-0.0343	0.3819	-0.1902	-0.0442	-0.0064	-0.0273	-0.2337	-0.1572	-0.3199	-0.0930	-0.1334	-0.1487	0.2786	-0.1032	-0.1225	0.7312	-0.1764	-0.1188	-0.1257	-0.1725	-0.0421	1.0000						
(24) ln GDP per Capita	0.7071	-0.1147	0.6219	-0.6019	-0.3881	0.3659	-0.0506	-0.0944	-0.0849	-0.4941	-0.1921	-0.4341	-0.7316	0.6399	-0.0662	-0.0271	0.2907	0.1394	-0.1310	-0.0059	-0.1570	0.2416	0.2702	1.0000					
(25) Human Capital	0.7094	-0.1168	0.6981	-0.5930	-0.2758	0.1560	0.1068	0.0818	0.0729	-0.7234	-0.1207	-0.4169	-0.7083	0.7020	-0.0027	-0.0750	0.3351	0.1014	-0.3934	-0.0397	-0.4160	0.2595	0.2501	0.7807	1.0000				
(26) Research and Development (Percentage of GDP)	0.8749	-0.1062	0.5742	-0.6538	-0.2068	-0.0346	0.2007	-0.0534	-0.1111	-0.5297	-0.1794	-0.3814	-0.4298	0.4912	-0.0224	-0.0236	0.4345	-0.0302	-0.2734	-0.0074	-0.2636	0.4656	0.4582	0.6323	0.6259	1.0000			
(27) Internet Users	0.8602	-0.0459	0.6838	-0.6927	-0.3937	0.1502	0.0981	-0.0770	-0.1158	-0.5941	-0.1265	-0.4139	-0.6253	0.6912	-0.0219	-0.0410	0.4827	0.0897	-0.2528	-0.0276	-0.2675	0.3247	0.4382	0.8508	0.7904	0.7786	1.0000		
(28) Fixed Telephone Subscriptions	0.9206	-0.0861	0.7237	-0.7322	-0.3469	0.1023	0.1827	0.0329	-0.0278	-0.6421	-0.1772	-0.4609	-0.6193	0.7352	-0.0229	-0.0813	0.4256	0.0914	-0.3091	-0.0527	-0.2215	0.3415	0.3949	0.8355	0.7951	0.8414	0.9160	1.0000	
(29) Mobile Cellular Subscriptions	0.6403	-0.1501	0.6592	-0.5299	-0.3848	0.1752	0.0849	0.0020	-0.0918	-0.5378	-0.2283	-0.4140	-0.6684	0.6800	-0.1618	-0.1613	0.2682	0.0786	-0.1883	-0.0885	-0.2729	0.1691	0.2524	0.8342	0.7873	0.5808	0.7555	0.7344	1.0000

D Data Description and Sources

Dependent Variable

Scientific Publications per 1,000 people. This variable refers to the country-level measure of scientific and technical journal articles per 1,000 people during the period 1981 to 2013. It refers to scientific articles that have been published in the fields of physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences. This data is part of the *Science & Technology* statistics of the World Bank *World Development Indicators* database. It provides information on scientific publication counts during the period 1981 to 2013 for a large number of countries across the world. In order to facilitate comparison of scientific productivity across countries, I normalize total scientific output by population size based on population data from the Penn World Tables Version 9.0 database (Feenstra et al., 2015).

The raw data on scientific and technical publication counts is prepared by the National Science Foundation's National Center for Science and Engineering Statistics (NCSES) as part of the Science and Engineering Indicators (SEI) database (National Science Board, 2014). It aims to provide a wide range of scientific and engineering indicators regarding scientific education (e.g., workers with scientific and engineering skills), research outputs (e.g., research and development performance and research publications), the creation and use of intellectual property (e.g., patent applications), and the economic activity of knowledge-intensive industries (e.g., production of high-technology manufacturing goods) both for the U.S. and other countries across the world, respectively, that can be employed for cross-country comparisons on the performance of scientific knowledge creation.

Data on cross-country publication counts are from peer-reviewed scientific and technical articles from journals with bibliographical information assigned by the Science Citation Index (SCI) and Social Science Citation Index (SSCI), respectively. Scientific articles are classified by year and institutional rather than individual authorship. The latter point is relevant for the identification of potential co-authors from scientists that differ with respect to institutional address and country of origin. In particular, scientific articles written by authors hosted within the same department or institution are classified as single institutionalized authorship. Article counts are assigned to countries based on the location of the institutional address rather than the nationality of the corresponding author. Articles written in international co-authorship are allocated in proportion to the respective institutional country of origin.

Genetic Variables

DRD4 Exon III 2- and 7-Repeat Allele Frequency. This variable refers to the country-level DRD4 exon III 2- and 7-repeat allele frequency measure. Based on existing evidence from molecular genetics on the relationship between the DRD4 exon III gene and novelty-seeking behaviour, the study in Gören (2017) provides DRD4 exon III 2- and 7-repeat allele frequencies for a large number of countries across the world. The raw data on DRD4 exon III allele frequencies is based on a worldwide sample of 120 populations in 59 countries with population genome data compiled from an extensive list of various molecular, population, and candidate gene studies of the DRD4 exon III gene (Gören, 2016). To account for the fact that contemporary countries consist of various ethnic groups that have partly undergone genetic admixture since the 1500 AD period, the methodology developed in Gören (2017) proposes the calculation of an ethnicity-weighted DRD4 exon III 2- and 7-repeat allele frequency measure. Specifically, this methodology matches the population shares data in the Alesina et al. (2003) ethnicity data to the DRD4 exon III population genome data in Gören (2016) utilizing information on linguistic groups from the *Ethnologue* database as the identifying feature. It is worth mentioning that in cases where the Alesina et al. (2003) ethnicity data refers to universal ethno-linguistic groups (e.g., 'white') or groups of mixed ancestry (e.g., 'Amerindian' or 'Mestizo'), the proposed methodology utilizes combinations of different populations in the Gören (2016) population genome data to infer the genetic composition of ethno-linguistic groups with respect to DRD4 exon III allele frequencies. In summary, the study in Gören (2017) provides DRD4 exon III 2- and 7-repeat allele frequencies in a sample of 181 countries across the world. These measures are used as a simple *proxy* variable indicating the prevalence of novelty-seeking behaviour in society.

Predicted Genetic Diversity (Ancestry-Adjusted). This variable refers to a country-level measure of overall genetic diversity. It is worth mentioning that this variable employs the predictive power of migratory distance from East Africa to the country's modern capital city, restricting the pre-historic migration routes to five intermediate land points, e.g., Cairo, Egypt (30N, 31E); Istanbul, Turkey (41N, 28E); Phnom Penh, Cambodia (11N, 104E); Anadyr, Russia (64N, 177E); and Prince Rupert, Canada (54N, 130W). The narrative of this approach is based on existing evidence in population genetics regarding the potential influence of a serial founder effect on genetic diversity (calculated as expected heterozygosity) since the pre-historic exodus of anatomically modern humans from East Africa (Ramachandran et al., 2005). This association has been uncovered based on a sample of 53 indigenous populations of the Human Genome Diversity Cell Line Panel from the Human Genome Diversity Project-Centre d'Etude du Polymorphisme Humain (HGDP-CEPH) (Cann et al., 2002).
It is exactly this strong statistical association between population-level expected heterozygosity and migratory distance from East Africa that Ashraf and Galor (2013) employ in the construction of their country-level measure of genetic diversity. Specifically, the authors use the regression coefficient associated with migratory distance from East Africa in the 53 HGDP-CEPH population sample and combine this estimate with the migratory distance from East Africa to the country's respective capital city. In order to control for the large immigration flows since the post-Columbian era (i.e., 1500 AD), I use the ancestry-adjusted version of the country-level measure of genetic diversity. This variable equals the weighted average of the various country-specific predicted genetic diversity measures, using as weights the corresponding population shares of a country's contemporary population that can trace their ancestry back to countries in the pre-Columbian era. The corresponding pre-Columbian population shares data is calculated from the *World Migration Matrix* database of Putterman and Weil (2010).

Diversity Variables

Ethnic Diversity. This variable refers to a country-level measure of ethnic diversity, ranging from 0 to 1, indicating the probability that two randomly chosen individuals within the same country will belong to different ethnic groups. A higher value of this measure corresponds to a more ethnically diverse country. See Alesina et al. (2003) for additional details on data construction and sources.

Linguistic Diversity. This variable refers to a country-level measure of linguistic diversity, ranging from 0 to 1, indicating the probability that two randomly chosen individuals within the same country will belong to different linguistic groups. A higher value of this measure corresponds to a more linguistically diverse country. See Alesina et al. (2003) for additional details on data construction and sources.

Religious Diversity. This variable refers to a country-level measure of religious diversity, ranging from 0 to 1, indicating the probability that two randomly chosen individuals within the same country will belong to different religious groups. A higher value of this measure corresponds to a more religiously diverse country. See Alesina et al. (2003) for additional details on data construction and sources.

Biogeographic Variables

Absolute Latitude. This variable refers to the country's approximate centroid absolute latitude in decimal degrees. This variable is part of *The Center for International Development* research datasets at Harvard University. The corresponding data is available online at http://www.cid.harvard.edu/ciddata/ciddata.html.

Distance to Major Markets. This variable refers to the minimum geographic distance (in 1,000 km) from the country's capital city to one of the three major capital supplying regions in the world (i.e., U.S., Western Europe, and Japan), as indicated by New York, Rotterdam, or Tokyo. This variable is part of *The Center for International Development* research datasets at Harvard University. The corresponding data is available online at http://www.cid.harvard.edu/ciddata/ciddata.html.

Distance to Coast or River. This variable refers to the minimum distance (in 1,000 km) from 1 decimal degree grid cells to the nearest ice-free coastline or sea-navigable rivers averaged across grid cells within the country's national border. This variable is part of *The Center for International Development* research datasets at Harvard University. The corresponding data is available online at http://www.cid.harvard.edu/ciddata/ciddata.html.

Hydrocarbons per Person. This variable refers to the country's proven oil and natural gas reserves (expressed as British thermal units per capita) in 1993. This variable is part of *The Center for International Development* research datasets at Harvard University. The corresponding data is available online at http://www.cid.harvard.edu/ciddata/ciddata.html.

Percentage Arable Land Area. This variable refers to the percentage of a country's arable land area. This variable is part of the World Bank's *World Development Indicators* database. The corresponding data is available online at http://data.worldbank.org/.

Agricultural Suitability. This variable refers to a geospatial indicator (ranging from 0 = low to 1 = high) of land suitability for agriculture. The raw data is provided by Ramankutty et al. (2002) with a spatial resolution of 0.5 decimal degrees latitude $\times$ longitude (approximately 55 km $\times$ 55 km at the equator). The country-level measures have been constructed by averaging across 0.5 decimal degree grid cells of land suitability of agriculture within the country's national border. The corresponding raw data is available online at the Center for Sustainability and the Global Environment (SAGE), http://www.sage.wisc.edu, at the University of Wisconsin.

Terrain Roughness. This variable refers to the mean absolute change in elevation values across contiguous neighbouring 1 decimal degree grid cells that captures the notion of 'mountainousness' in a particular region. The raw data is provided by the Global Economic Activity (G-ECON) database with a spatial resolution of 1 decimal degree latitude $\times$ longitude (Nordhaus, 2006). The corresponding country-level measures of terrain roughness across 1 decimal degree grid cells within the country's national border have been constructed by Ashraf and Galor (2013).

Temperature. This variable refers to the country's mean temperature value (in Degree Celsius) during the period 1960 to 1990. The raw data is constructed by the G-ECON database with a spatial resolution of 1 decimal degree latitude × longitude from monthly temperature observations during the period 1960 to 1990 provided by the Climatic Research Unit database at the University of East Anglia. See Nordhaus (2006) and New et al. (2002) for additional details on data construction and sources. The corresponding country-level temperature values across 1 decimal degree grid cells within the country's national border have been constructed by Ashraf and Galor (2013).

Precipitation. This variable refers to the mean precipitation value (in total millimeters per month) during the period 1960 to 1990. The raw data is constructed by the G-ECON database with a spatial resolution of 1 decimal degree latitude × longitude from monthly temperature observations during the period 1960 to 1990 provided by the Climatic Research Unit database at the University of East Anglia. See Nordhaus (2006) and New et al. (2002) for additional details on data construction and sources. The corresponding country-level precipitation values across 1 decimal degree grid cells within the country's national border have been constructed by Ashraf and Galor (2013).

Population in Tropics (Share). This variable refers to the share of a country's population in tropical areas. This variable is part of *The Center for International Development* research datasets at Harvard University. The corresponding data is available online at http://www.cid.harvard.edu/ciddata/ciddata.html.

Population in Area with Malaria (Share). This variable refers to the share of a country's population residing in areas subject to malaria falciparum. This variable is part of *The Center for International Development* research datasets at Harvard University. The corresponding data is available online at http://www.cid.harvard.edu/ciddata/ciddata.html.

Region Fixed Effects. This variable refers to a set of indicator variables that takes a value of 1 if the respective country is located in one of the main regional classifications according to the definition of the World Bank (i.e., North America, Central and South America, Sub-Saharan Africa, East Asia, Central Asia, South Asia, Middle East, Europe, or Oceania) and zero otherwise.

Island Effects. This variable refers to an indicator variable that takes a value of 1 for island countries and zero otherwise.

Landlocked Effects. This variable refers to an indicator variable that takes a value of 1 for landlocked countries and zero otherwise. This variable is part of the *Geography* database at the *Centre d'Études Prospectives et d'Informations Internationales (CEPII)*. The corresponding data is available online at http://www.cepii.fr/CEPII/en/bdd_modele/bdd.asp.

Historical Variables

European Population (Share). The variable refers to the share of a country's current population that can trace their ancestry back to European populations in the year 1500 AD. This variable is calculated from the *World Migration Matrix* database of Putterman and Weil (2010).

Official Language is English. This variable refers to an indicator variable that takes a value of 1 if the country's official language is English and zero otherwise. This variable is part of the *Geography* database at the *Centre d'Études Prospectives et d'Informations Internationales (CEPII)*. The corresponding data is available online at http://www.cepii.fr/CEPII/en/bdd_modele/bdd.asp.

Former Colonizer is British. This variable refers to an indicator variable that takes a value of 1 if the country's former colonizer was British and zero otherwise. This variable is part of the *Geography* database at the *Centre d'Études Prospectives et d'Informations Internationales (CEPII)*. The corresponding data is available online at http://www.cepii.fr/CEPII/en/bdd_modele/bdd.asp.

Legal Origin Fixed Effects. This variable refers to the set of indicator variables that takes a value of 1 if the respective country's legal origin is British, French, German, Scandinavian, or Socialist and zero otherwise. The corresponding data is available from the La Porta et al. (1999) study.

Major Religion Fixed Effects. This variable refers to the set of a country's major religion shares for Protestants, Catholics, and Muslims (as % of population in 1980). The corresponding data is available from the La Porta et al. (1999) study.

Technological Frontier Variables

Genetic Distance to the Technological Frontier. This variable refers to the mean population-weighted country-level FST genetic distance measure to various technological frontier countries (i.e., U.S., OECD, or Neo-Europe). The corresponding data is available from the cultural distance database of Spolaore and Wacziarg (2016).

Linguistic Distance to the Technological Frontier. This variable refers to the mean population-weighted country-level linguistic distance measure to various technological frontier countries (i.e., U.S., OECD, or Neo-Europe). The measures of linguistic distance between any pair of languages is based on the phylogenetic language tree approach, which indicates the number of common nodes that any two languages share within a global language tree classification. The corresponding data is available from the cultural distance database of Spolaore and Wacziarg (2016).

Religious Distance to the Technological Frontier. This variable refers to the mean country-level religious distance measure to various technological frontier countries (i.e., U.S., OECD, or Neo-Europe). Religious distance is measured as the number of common nodes between any pair of religions based on a global categorization of religions. The corresponding data is available from the cultural distance database of Spolaore and Wacziarg (2016).

Technological and Human Capability Variables

GDP per Capita. This variable refers to the purchasing power parity (PPP) adjusted output-side real GDP per capita during the years 1981 to 2013 (expressed in 2011 USD). This variable has been constructed from the Penn World Tables database, Version 9.0 (Feenstra et al., 2015). The corresponding data is available online at http://www.rug.nl/ggdc/productivity/pwt/.

Research and Development. This variable refers to research and development expenditures in % of GDP during the years 1981 to 2013. This variable is part of the *Science & Technology* statistics of the World Bank *World Development Indicators* database. The corresponding data is available online at http://data.worldbank.org/.

Human Capital. This variable refers to a human capital index during the years 1981 to 2013. This index has been constructed from the Penn World Tables database, Version 9.0, by combining the Barro and Lee (2013) years of schooling database and information on the assumed rate of return to educational years, as suggested in Psacharopoulos (1994). The corresponding data is available online at http://www.rug.nl/ggdc/productivity/pwt/.

Internet Users. This variable refers to the number of internet users (per 100 people) during the years 1981 to 2013. This variable is part of the World Bank *World Development Indicators* database. The corresponding data is available online from the *Quality of Government* (QoG) database (Teorell et al., 2017).

Fixed Telephone Subscriptions. This variable refers to the number of fixed telephone subscriptions (per 100 people) during the years 1981 to 2013. This variable is part of the World Bank *World Development Indicators* database. The corresponding data is available online from the *Quality of Government* (QoG) database (Teorell et al., 2017).

Mobile Cellular Subscriptions. This variable refers to the number of mobile cellular subscriptions (per 100 people) during the years 1981 to 2013. This variable is part of the World Bank *World Development Indicators* database. The corresponding data is available online from the *Quality of Government* (QoG) database (Teorell et al., 2017).

Social Value Variables

Trade Openness. This variable refers to the sum of exports and imports of goods and services over total GDP (measured in % of GDP) during the years 1981 to 2013. This variable is part of the World Bank *World Development Indicators* database. The corresponding data is available online from the *Quality of Government* (QoG) database (Teorell et al., 2017).

International Migrant Stock. This variable refers to the international migrant stock (measured in % of population) during the years 1981 to 2010. This variable is part of the World Bank *World Development Indicators* database. The corresponding data is available online from the *Quality of Government* (QoG) database (Teorell et al., 2017).

Freedom of Expression and Belief. This variable refers to the country's freedom of media, cultural expression, religion, academia, education, open and private discussion during the years 2004 to 2013. The relevant metric ranges from 0 = low to 16 = high. This variable is part of the *Freedom House* database (Freedom House, 2016). The corresponding data is available online from the *Quality of Government* (QoG) database (Teorell et al., 2017).

Civil Liberties. This variable refers to the country's freedom of expression and belief, associational and organizational rights, rule of law, and personal autonomy and individual rights during the years 1981 to 2013. The relevant metric ranges from 1 = low to 7 = high. This variable is part of the *Freedom House* database (Freedom House, 2016). The corresponding data is available online from the *Quality of Government* (QoG) database (Teorell et al., 2017).

Quality of Governance Variables

Freedom from Corruption. This variable refers to the Transparency International's Corruption Perception Index (CPI), indicating the country's level of freedom from corruption during the years 1995 to 2013. The relevant metric ranges from 0 = low to 100 = high. This variable is part of the *Heritage Foundation* database (Heritage Foundation, 2015). The corresponding data is available online from the *Quality of Government* (QoG) database (Teorell et al., 2017).

Rule of Law. This variable refers to a composite index of a country's rule of law during the years 2004 to 2013. The relevant metric ranges from 0 = low to 16 = high. This variable is part of the *Freedom House* database (Freedom House, 2016). The corresponding data is available online from the *Quality of Government* (QoG) database (Teorell et al., 2017).

Independence of the Judiciary. This variable refers to a country's independence of the judiciary during the years 1981 to 2011. The relevant metric is 0 = *not independent*, 1 = *partially independent*, and 2 = *generally independent*. This variable is part of the Cingranelli et al. (2014) (CIRI) Human Rights database. The corresponding data is available online from the *Quality of Government* (QoG) database (Teorell et al., 2017).

Property Rights Protection. This variable refers to a country's degree of property rights protection during the years 1995 to 2013. The relevant metric ranges from 0 = low to 100 = high. This variable is part of the *Heritage Foundation* database (Heritage Foundation, 2015). The corresponding data is available online from the *Quality of Government* (QoG) database (Teorell et al., 2017).

Executive Constraints. This variable refers to a country's executive constraints during the years 1981 to 2013. The relevant metric ranges from 1 = low to 7 = high. This variable is part of the *Polity IV* dataset (Marshall et al., 2016). The corresponding data is available online from the *Quality of Government* (QoG) database (Teorell et al., 2017).

Institutionalized Democracy. This variable refers to the country's degree of institutionalized democracy during the years 1981 to 2013. The relevant metric ranges from 0 = low to 10 = high. This variable is part of the *Polity IV* dataset (Marshall et al., 2016). The corresponding data is available online from the *Quality of Government* (QoG) database (Teorell et al., 2017).

Instrumental Variables

Migratory Distance from East Africa (Ancestry-Adjusted). This variable refers to the migratory distance from East Africa to the country's modern capital city (in 1,000 km). This variable has been constructed by conditioning the pre-historic migratory paths through five land-restricted way-points: namely Cairo, Egypt (30N, 31E); Istanbul, Turkey (41N, 28E); Phnom Penh, Cambodia (11N, 104E); Anadyr, Russia (64N, 177E); and Prince Rupert, Canada (54N, 130W). The calculation of the country's migratory distance utilizes the great circle or geodesic calculation approach, as outlined in Ramachandran et al. (2005). Geographic coordinates of the country's modern capital city in decimal degrees are provided by the CIA's *The World Factbook*. The corresponding country's capital latitude and longitude coordinates are available at https://www.cia.gov/library/publications/the-world-factbook/. This variable has been ancestry-adjusted to correct for post-Columbian population flows since 1500 AD based on the *World Migration Matrix* data of Putterman and Weil (2010).

Pasture Land (Ancestry-Adjusted). This variable refers to a geospatial indicator (ranging from 0 = low to 1 = high) of the fraction of land allocated to pasture. The raw data has a spatial resolution of 0.5 decimal degrees latitude × longitude grid cells. The country-level measures have been constructed by averaging across 0.5 decimal degree grid cells of pasture land within the country's national border. The corresponding data is distributed by the Center for Sustainability and the Global Environment (SAGE) at the University of Wisconsin and available online at http://nelson.wisc.edu/sage/data-and-models/global-land-use/index.php. The country-level pasture land variable has been ancestry-adjusted to correct for post-Columbian population flows since 1500 AD based on the *World Migration Matrix* data of Putterman and Weil (2010).

Elevation (Ancestry-Adjusted). This variable refers to the mean elevation value (in km) above the sea level. The raw data has a spatial resolution of 0.0833333 decimal degrees latitude × longitude that has been spatially aggregated to 0.5 decimal degree grid cells. The country-level measures have been constructed by averaging across 0.5 decimal degree grid cells of elevation values within the country's national border. The raw data is distributed by the National Oceanic and Atmospheric Administration (NOAA) and U.S. National Geophysical Data Center, TerrainBase, release 1.0 (CD-ROM), Boulder, Colorado, and available online at http://nelson.wisc.edu/sage/data-and-models/atlas/maps.php. The country-level elevation variable has been ancestry-adjusted to correct for post-Columbian population flows since 1500 AD based on the *World Migration Matrix* data of Putterman and Weil (2010).

Agricultural Suitability (Ancestry-Adjusted). This variable refers to a geospatial indicator (ranging from 0 = low to 1 = high) of land suitability for agriculture. See the *Agricultural Suitability* index for additional details on data construction and sources. The country-level land suitability for agriculture variable has been ancestry-adjusted to correct for post-Columbian population flows since 1500 AD based on the *World Migration Matrix* data of Putterman and Weil (2010).

Zuletzt erschienen /previous publications:

V-402-17 Erkan Gören, The Role of Novelty-Seeking Traits in Contemporary Knowledge Creation
V-401-17 **Heinz Welsch, Jan Kühling**, Divided We Stand: Immigration Attitudes, Identity, and Subjective Well-Being
V-400-17 **Christoph Böhringer, Thomas F. Rutherford**, Paris after Trump: An inconvenient insight
V-399-17 **Frank Pothen, Heinz Welsch**, Economic Development and Material Use
V-398-17 **Klaus Eisenack, Marius Paschen,** Designing long-lived investments under uncertain and ongoing change
V-397-16 **Marius Paschen**, The effect of intermittent renewable supply on the forward premium in German electricity markets
V-396-16 **Heinz Welsch, Philipp Biermann**, Poverty is a Public Bad: Panel Evidence from Subjective Well-being Data
V-395-16 **Philipp Biermann**, How Fuel Poverty Affects Subjective Well-Being: Panel Evidence from Germany
V-394-16 **Heinz Welsch**, Electricity Externalities, Siting, and the Energy Mix: A Survey
V-393-16 **Leonhard Kähler, Klaus Eisenack**, Strategic Complements in International Environmental Agreements: a New Island of Stability
V-392-16 **Christoph Böhringer, Xaquin Garcia-Muros, Ignacio Cazcarro, Iñaki Arto**, The Efficiency Cost of Protective Measures in Climate Policy
V-391-16 **Achim Hagen, Juan-Carlos Altamirano-Cabrera, Hans-Peter Weikard**, The Influence of Political Pressure Groups on the Stability of International Environmental Agreements
V-390-16 **Christoph Böhringer, Florian Landis, Miguel Angel Tovar Reaños,** Cost-effectiveness and Incidence of Renewable Energy Promotion in Germany
V-389-16 **Carsten Helm, Mathias Mier,** Efficient diffusion of renewable energies: A roller-coaster ride
V-388-16 **Christoph Böhringer, Jan Schneider, Emmanuel Asane-Otoo,** Trade In Carbon and The Effectiveness of Carbon Tariffs
V-387-16 **Achim Hagen, Leonhard Kähler, Klaus Eisenack,** Transnational Environmental Agreements with Heterogeneous Actors
V-386-15 **Jürgen Bitzer, Erkan Gören, Sanne Hiller,** Absorption of Foreign Knowledge: Firms' Benefits of Employing Immigrants
V-385-15 **Klaus Eisenack, Julien Minnemann, Paul Neetzow, Felix Reutter,** Contributions to the institutional economics of the energy transition
V-384-15 **Christoph Böhringer, Xaquín Garcia-Muros, Mikel Gonzalez-Eguino, Luis Rey,** US Climate Policy: A Critical Assessment of Intensity Standards
V-383-15 **Christoph Böhringer, Edward J. Balistreri, Thomas F. Rutherford**, Carbon policy and the structure of global trade
V-382-15 **Christoph Böhringer, Brita Bye, Taran Fæhn, Knut Einar Rosendahl,** Output-based rebating of carbon taxes in the neighbor's backyard
V-381-15 **Christoph Böhringer, Markus Bortolamedi**, Sense and No(n)-Sense of Energy Security Indicators
V-380-15 **Christoph Böhringer, Knut Einar Rosendahl, Halvor Briseid Storrøsten,** Mitigating carbon leakage:Combining output-based rebating with a consumption tax
V-379-15 **Jan Micha Steinhäuser, Klaus Eisenack**, Spatial incidence of large-scale power plant curtailment costs
V-378-15 **Carsten Helm, Franz Wirl**, Climate policies with private information: The case for unilateral action
V-377-15 **Klaus Eisenack**, Institutional adaptation to cooling water scarcity in the electricity sector under global warming
V-376-15 **Christoph Böhringer, Brita Bye, Taran Fæhn, and Knut Einar Rosendahl**, Targeted carbon tariffs – Carbon leakage and welfare effects
V-375-15 **Heinz Welsch, Philipp Biermann**, Measuring Nuclear Power Plant Externalities Using Life Satisfaction Data: A Spatial Analysis for Switzerland

V-374-15 **Erkan Gören**, The Relationship Between Novelty-Seeking Traits And Comparative Economic Development
V-373-14 **Charlotte von Möllendorff**, Heinz Welsch, Measuring Renewable Energy Externalities: Evidence from Subjective Well-Being Data
V-372-14 **Heinz Welsch, Jan Kühling**, Affective States and the Notion of Happiness: A Preliminary Analysis
V-371-14 **Carsten Helm, Robert C. Schmidt,** Climate cooperation with technology investments and border carbon adjustment
V-370-14 **Christoph Böhringer, Nicholas Rivers, Hidemichi Yonezawa**, Vertical fiscal externalities and the environment
V-369-14 **Heinz Welsch, Philipp Biermann,** Energy Prices, Energy Poverty, and Well-Being: Evidence for European Countries
V-368-14 **Marius Paschen**, Dynamic Analysis of the German Day-Ahead Electricity Spot Market
V-367-14 **Heinz Welsch, Susana Ferreira,** Environment, Well-Being, and Experienced Preference
V-366-14 **Erkan Gören**, The Biogeographic Origins of Novelty-Seeking Traits
V-365-14 **Anna Pechan,** Which Incentives Does Regulation Give to Adapt Network Infrastructure to Climate Change? - A German Case Study
V-364-14 **Christoph Böhringer, André Müller, Jan Schneider**, Carbon Tariffs Revisited
V-363-14 **Christoph Böhringer, Alexander Cuntz, Diemtar Harhoff, Emmanuel A. Otoo,** The Impacts of Feed-in Tariffs on Innovation: Empirical Evidence from Germany
V-362-14 **Christoph Böhringer, Nicholas Rivers, Thomas Ruhterford, Randall Wigle**, Sharing the burden for climate change mitigation in the Canadian federation
V-361-14 **Christoph Böhringer, André Müller**, Environmental Tax Reforms in Switzerland A Computable General Equilibrium Impact Analysis
V-360-14 **Christoph Böhringer, Jared C. Carbone, Thomas F. Rutherford,** The Strategic Value of Carbon Tariffs
V-359-13 **Heinz Welsch, Philipp Biermann**, Electricity Supply Preferences in Europe: Evidence from Subjective Well-Being Data
V-358-13 **Heinz Welsch, Katrin Rehdanz, Daiju Narita, Toshihiro Okubo,** Well-being effects of a major negative externality: The case of Fukushima
V-357-13 **Anna Pechan, Klaus Eisenack**, The impact of heat waves on electricity spot markets
V-356-13 **Heinz Welsch, Jan Kühling,** Income Comparison, Income Formation, and Subjective Well-Being: New Evidence on Envy versus Signaling
V-355-13 **Christoph Böhringer, Knut Einar Rosendahl, Jan Schneider,** Unilateral Climate Policy: Can Opec Resolve the Leakage Problem?
V-354-13 **Christoph Böhringer, Thomas F. Rutherford, Marco Springmann;** Clean-Development Investments:An Incentive-Compatible CGE Modelling Framework
V-353-13 **Erkan Gören,** How Ethnic Diversity affects Economic Development?
V-352-13 **Erkan Gören,** Economic Effects of Domestic and Neighbouring Countries'Cultural Diversity
V-351-13 **Jürgen Bitzer, Erkan Gören**, Measuring Capital Services by Energy Use: An Empirical Comparative Study
V-350-12 **Heinz Welsch, Jan Kühling**, Competitive Altruism and Endogenous Reference Group Selection in Private Provision of Environmental Public Goods
V-349-12 **Heinz Welsch,** Organic Food and Human Health: Instrumental Variables Evidence
V-348-12 **Carsten Helm, Dominique Demougin,** Incentive Contracts and Efficient Unemployment Benefits in a Globalized World
V-347-12 **Christoph Böhringer, Andreas Lange, Thomas F. Rutherford,** Optimal Emission Pricing in the Presence of International Spillovers: Decomposing Leakage and Terms-of-Trade Motives
V-346-12 **Christoph Böhringer, Jared C. Carbone, Thomas F. Rutherford**, Efficiency and Equity Implications of Alternative Instruments to Reduce Carbon Leakage
V-345-12 **Christoph Böhringer, Brita Bye, Taran Fæhn, Knut Einar Rosendahl,** Alternative Designs for Tariffs on Embodied Carbon: A Global Cost-Effectiveness Analysis