"Geography and Regional Economic Growth: The high cost of deviating from nature"

Daniel Albalate, Germà Bel and Ferran A. Mazaira-Font

The Research Institute of Applied Economics (IREA) in Barcelona was founded in 2005, as a research institute in applied economics. Three consolidated research groups make up the institute: AQR, RISK and GiM, and a large number of members are involved in the Institute. IREA focuses on four priority lines of investigation: (i) the quantitative study of regional and urban economic activity and analysis of regional and local economic policies, (ii) study of public economic activity in markets, particularly in the fields of empirical evaluation of privatization, the regulation and competition in the markets of public services using state of industrial economy, (iii) risk analysis in finance and insurance, and (iv) the development of micro and macro econometrics applied for the analysis of economic activity, particularly for quantitative evaluation of public policies.

IREA Working Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. For that reason, IREA Working Papers may not be reproduced or distributed without the written consent of the author. A revised version may be available directly from the author.

Any opinions expressed here are those of the author(s) and not those of IREA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

Abstract

We analyze the role of nature and geography in determining economic and social outcomes. We propose a theoretical model relating geography and nature to economic growth, and examine that model using data from NUTS 2 European regions. By doing this, we identify the predictive power of first-nature variables to explain regional population distribution. Then we analyze the effects of misadjustment between the actual and predicted distribution of populations on economic performance. Our results indicate that deviating from first-nature outcomes has a significant negative effect on economic growth. Furthermore, we find that departing from natural endowment has a negative effect on social cohesion. The main policy implication emerging from our analysis is that strategies that harmonize with nature and geography yield better social welfare and human well-being than those policies that conflict with them.

Daniel Albalate: Department of Econometrics, Statistics and Applied Economics (Public Policy Unit). Universitat de Barcelona. John Keynes 1-11, 08034 Barcelona. Spain. E-mail: albalate@ub.edu

Germà Bel: Department of Econometrics, Statistics and Applied Economics (Public Policy Unit). Universitat de Barcelona. John Keynes 1-11, 08034 Barcelona. Spain. E-mail: gbel@ub.edu

Ferran A. Mazaira-Font: Department of Econometrics, Statistics and Applied Economics (Public Policy Unit). Universitat de Barcelona. John Keynes 1-11, 08034 Barcelona. Spain. E-mail: fmazaifo7@alumnes.ub.edu

Geography and Regional Economic Growth: The high cost of deviating from nature

Throughout history, humanity has made unimaginable progress, not only overcoming the obstacles or conditioning factors of nature and geography, but even putting them to use in the service of its aspirations and interests. Indeed, communities have been able to survive, develop and flourish in settlements with hostile climates, to overcome the limits set by water – the barrier effects of large rivers and oceans – and the terrain, with its harshness and geological complexity. The vast knowledge acquired and applied over time has led to technological advancement and hence to the possibility of mankind adapting to its environment and vice versa.

This progress has inspired several scholars to write of the *end of geography* (O'Brien 1992) and the *death of distance* (Cairncross 1995; 2001), and to describe the new era as that of a *placeless society*, a *shrinking, flat world* (Friedman 2005), all due to advances in transportation, information and communication technologies. At the same time, the most isolated places in the world have become easily accessible. This has allowed the development of prosperous settlements, now integrated into the global economy, that would otherwise have been challenged by the limiting features attributable to their natural environments. Human life has apparently become liberated from the constraints of space and frictional effects of distance (Graham 1998).

Neglecting the ever-present role and restraints of nature and geography may be premature. Regarding the so-called *death of distance*, Rietveld and Vickerman (2004) pointed out that many economic activities have not become that 'footloose', due to transaction costs and other reasons. In fact, proximity to higher-tiered urban centers continues to be an important positive determinant of local job growth, despite the alleged death of distance (Partridge *et al.* 2008). International conflicts, past and current economic dilemmas and challenges worldwide also exhibit a strong relationship with geography and nature, even now (Senese 2005; Starr 2005; Kaplan 2012).

The role of nature and geography has been and still is crucial to understanding many of the social, political and economic outcomes and prospects of human settlements. From a historical perspective, the capacity of the environment to support human life has commonly been considered a major restraint on population growth, density and prosperity. This idea has been discussed, for instance, by Machiavelli (1519), Botero (1588) and Montesquieu (1748), and formed part of the famous demographic theory of Malthus (1798). Even earlier, Plato and Aristotle expressed concerns about overpopulation and limited resources. Thus, thinkers and

scholars alike have long spoken of the importance of place and natural constraints and endowments for the location and density of human settlements and their effects on economic prospects.

In economic geography models, agglomerations are expected to be located and to develop according to a set of first-nature and second-nature determinants (Krugman 1993). Among first-nature determinants, geography and nature play the most crucial role. Thus, economic activity depends on the physical landscape, climate, access to the sea and to navigable rivers, etc. Clearly though, human action and incentives define the second-nature determinants that lead to increasing returns, due to scale and density economies, knowledge spillovers, etc. (Krugman 1991). Labor migrations between regions are responses to market signals and they determine the balance between agglomeration and dispersion forces (Krugman and Venables 1990). Locational advantages – attributable to geography and nature – favor the concentration and mobility of human and economic settlements (Ellisson and Glaeser 1999; Glaeser and Shapiro 2003; Black and Henderson 1998), leading to both the concentration of populations and the growth of productivity (Beeson 2001; Mitchener and McLean 2003).

The most extensive strand in the literature exploring this relationship is the one concerned with the role played by geography in relation to economic growth and development (Diamond 1997; Gallup, Sachs, and Mellinger 1999; Sachs and Warner 2001). A recent article by Mitton (2016) evaluates the determinants of economic development in 1,867 subnational regions of 101 countries, focusing on within-country effects of geography and institutions. Several geographic factors had significant explanatory power for within-country differences in per-capita GDP, including terrain ruggedness, tropical climate, ocean access, temperature range, storm risk and natural resources such as oil, diamonds, and iron. Beyond the constraints imposed on economic prosperity, some authors have also argued that geography may have an impact on institutions, another relevant set of economic determinants (Acemoglu, Johnson, and Robinson 2001, 2002; Easterly and Levine, 2003). Moreover, this strand of the literature points out that even though it is still a determining factor, once institutions are accounted for, the contribution of geography as a determinant of economic growth is partially diminished (Acemoglu, Johnson, and Robinson 2001, 2002; Easterly 2001; Rodrik, Subramanian, and Trebbi 2004).

Geography and nature provide an endowment that may facilitate the location, concentration and growth of some settlements or make these more difficult in the case of others. This dependence can be tempered or even completely reversed by the intervention of human capital accumulation that translates into knowledge and technical advances (Glaeser *et al*. 2004;

Bhattacharyya 2009) and/or by means of institutions (Acemoglu, Johnson, and Robinson 2002), which use a framework of incentives, regulations and investments.

In this article, we propose a theoretical model of the way in which features of geography and nature can account for regional economic growth, due to their effects on population density and distribution. This model is empirically examined using data from comparable European regions. We identify the strong predictive power of first-nature variables in explaining regional population density and capital city location, to the extent that we can estimate the degree of geographic harmonization of the actual distribution compared to the predicted distribution. This allows us to detect deviations produced by the forces of human action, led mainly by institutions, and to evaluate the predicted consequences in terms of relative economic performance. Our main results indicate that deviating from nature's outcomes has a significant negative effect on economic growth and may also produce larger inequalities. This result suggests that societies that opt to accommodate to the provisions of nature, and consequently, to exploit the opportunities of the best locations, rather than forcing a different distribution of the population across regions, perform better.

The remainder of the article is organized as follows. Next, we propose the theoretical model. After that we present the data and the empirical model that allows calculatating deviations by considering actual versus predicted population density and capital city location, across regions and countries. Then, we interpret the results focusing on the countries that deviate the most. The economic cost of deviations is estimated in the subsequent section, where an econometric convergence growth model is estimated. Finally, we discuss our main results and conclude.

Regional economic model

The aim of this section is to provide a conceptual model for understanding how population distribution and geography could impact economic growth.

Let us consider a closed economy formed by M regions. Let us assume that all of them occupy an equal area (equal to 1 to normalize), have equal access to technology A and a neoclassical production function of the form:

$$Y_i = F(g_i, K_i, D_i) = G(g_i)AK^\alpha D^{1-\alpha} \tag{1}$$

where K_i is the capital of a region i, D_i the total population living in the region (since area is equal to one, this is density of population), g_i its geographic endowment and $G(g)$ a function of geographic endowment such that $\frac{\partial G}{\partial g} > 0$. Total production of the economy is

$$Y = \sum_{i=1}^{M} Y_i \qquad (2)$$

Firms' maximization problem

Let us consider that a large number of firms face the classic problem of profit maximization under conditions of competitive labor and capital markets in every region i:

$$\Pi_i = \max_{K,D} F_i(g_i, K, D) - R_i K - w_i D \qquad (3)$$

First-order conditions imply that capital and labor are paid their marginal contributions:

$$w_i = \frac{\partial F_i}{\partial D}, \ R_i = \frac{\partial F_i}{\partial K} \qquad (4)$$

Capital markets

For there to be an equilibrium, assuming that financial markets are competitive and there are no externalities, we need the return on capital to be equal in all regions. Proposition 1 shows that there is equilibrium only in allocation of capital between the different regions.

Proposition 1

Let us note by K the total amount of capital in the economy. Let us also note by $K_i = t_i K$, with $0 < t_i < 1$ and $\sum_{i=1}^{M} t_i = 1$, the capital of the region i. Then, given $D_1, ..., D_M$, there exist unique values $t_1^*, ..., t_M^*$ such that $R_i = R_j$ for all i *and* j. These values correspond to a proportional allocation of capital with respect to production, that is:

$$t_i^* = \frac{Y_i}{Y}$$

Proof:

Let us consider M = 2. By (1) and (4), return on capital is

$$R_i = \alpha \frac{Y_i}{K_i}$$

Let us note that $Y_i(t) = Y_i(g_i, tK, D_i)$. Then,

$$R_1 = R_2 \Leftrightarrow \alpha \frac{Y_1(t_1)}{t_1 K} = \alpha \frac{Y_2(t_2)}{t_2 K} \Leftrightarrow \frac{Y_1(t_1)}{t_1 K} = \frac{Y_2(1 - t_1)}{(1 - t_1)K}$$

Let us note that $X_1(t) = \frac{Y_1(t)}{t}$. Notice that $X_1(t)$ is decreasing, since

$$X_1'(t) = \frac{\frac{\partial Y_i(t)}{\partial K} tK - Y_1(t)}{t^2} = \frac{\alpha Y_1(t) - Y_1(t)}{t^2} = (\alpha - 1)\frac{Y_1(t)}{t^2} < 0$$

5

Using the same reasoning, $X_2(t) = \frac{Y_2(1-t)}{1-t}$ is increasing in t. Then, since $\lim_{t\to 0} X_1(t) = \infty$ and $\lim_{t\to 1} X_2(t) = \infty$, there exists a unique t^* such that $X_1(t^*) = X_2(t^*)$

Moreover,

$$\frac{Y_1(t^*)}{t^*} = \frac{Y_2(1-t^*)}{1-t^*} \Leftrightarrow Y_1(t^*) = t^*\big(Y_1(t^*) + Y_2(1-t^*)\big) \Leftrightarrow t^* = \frac{Y_1(t^*)}{Y}$$

Consider now the case of $M > 2$. By induction hypothesis, let us assume that the results hold for M regions. We want to see whether it holds if we consider an economy with $M+1$. Let us define,

$$X(t) = \sum_{i=1}^{M} \frac{Y_i(tt_i)}{tt_i}$$

Where $t_1, \dots, t_M$ are the unique t_i (which exist by induction hypothesis) such that $\sum_{i=1}^{M} t_i = 1$ and $\beta \equiv \frac{Y_i(tt_i)}{tt_i} = \frac{Y_i(tt_j)}{tt_j}$ for all i and j. Notice that t_i does not depend on t, because the last equality holds for all $t \in (0,1)$.

As before, X is decreasing in t:

$$X'(t) = \sum_{i=1}^{M} \frac{\frac{\partial Y_i(t)}{\partial K} tt_i^2 K - t_i Y_i}{(t_i t)^2} = \sum_{i=1}^{M} (\alpha - 1)\frac{t_i Y_i}{(t_i t)^2} = (\alpha - 1)\frac{\beta M}{t} < 0$$

and $\lim_{t\to 1} X(t) = \infty$. Note that $X_{M+1}(t) = \frac{Y_{M+1}(1-t)}{1-t}$. As in the case of two regions, there exists t^* such that $X(t^*) = X_{M+1}(t^*)$, and by the induction hypothesis, there exist $t_1, \dots, t_M$, such that $X_i(t_i t^*) = X_{M+1}(t^*)$

Household maximization problem

Let us assume that households can choose where to locate and can move without any costs. Let us assume that utility of region i is a function of the form:

$$u_i(w_i, D_i) = f(w_i + \tau_i) + e(D_i) \tag{5}$$

where f is a concave and strictly increasing function, w_i is the net income per capita after taxes in the region i, τ_i represents public transfers per capita at region i and $e(D_i) = aD^2 + bD_i$ represents the externalities associated with density of population. Following theoretical models in urban economics (e.g., O'Sullivan 2007; Duranton and Puga 2020) we assume that increasing

density in lowly populated areas has a positive effect (e.g., accessibility of a greater diversity of goods and services), but at a certain point externalities become negative (e.g., congestion or a rise in land prices). Thus, $a < 0$ and $b > 0$.

The household maximization challenge is to choose the region i that maximizes utility. To illustrate which kind of equilibrium will be reached, let us consider, w.l.g., the case of two regions.

Proposition 2

Let us note by D_i, $i = 1, 2$, the population living in region i. Then, $D = D_1 + D_2$ and $D_1 = tD$, with $t \in [0, 1]$. The necessary and sufficient conditions for the existence of a Nash equilibrium in population distribution between region 1 and 2, given capitals K_1 and K_2 are:

i) $t = 1$ and $u_1(D) > u_2(0)$, or

ii) $t = 0$ and $u_1(0) < u_2(D)$, or

iii) $t \in (0, 1)$ and $\frac{\partial u_i(D_i)}{\partial D_i} < 0$

When one of the two regions is empty, the equilibrium is Pareto optimal if, and only if, household utility of the non-empty region at D is higher than the maximum utility of the other region.

Proof:

For there to be an equilibrium, expected utility gain from moving has to be non-positive. First, for i) and ii), let us consider, w.l.g., that $t = 1$. If $u_1(D) < u_2(0)$, then the whole population would move to region 2, so $t = 1$ is an equilibrium if, and only if, $u_1(D) > u_2(0)$. For iii), let us consider that there is an equilibrium between regions 1 and 2, and people living in both. Since there cannot be any utility gain from moving, utility in both regions has to be the same. Thus,

$$\hat{u} \equiv u_1(tD) = u_2((1 - t)D)$$

Let us consider, w.l.g., that $\frac{\partial u_1(D_1)}{\partial D_1} > 0$. Then, the marginal gain from moving from 2 to 1 is:

$$u_1(D_1 + \varepsilon) - \hat{u} \approx u_1(D_1) + \frac{\partial u_1(D_1)}{\partial D_1}\varepsilon - \hat{u} = \frac{\partial u_1(D_1)}{\partial D_1}\varepsilon < 0$$

or all $\varepsilon > 0$, small enough. Thus, households would be better off moving to region 1, and the distribution would not hold to a Nash equilibrium.

Figures 1 and 2 present examples of the two equilibria for an economy with $\alpha = 0.3$, $g_1 = 1$, $g_2 = 1.1$, $K = 200$, $a = -5 \cdot 10^{-3}$ and $b = 3.75 \cdot 10^{-2}$. For the single region equilibrium, total population is 30. The red dotted line represents the utility people would get if they lived in the worst region.
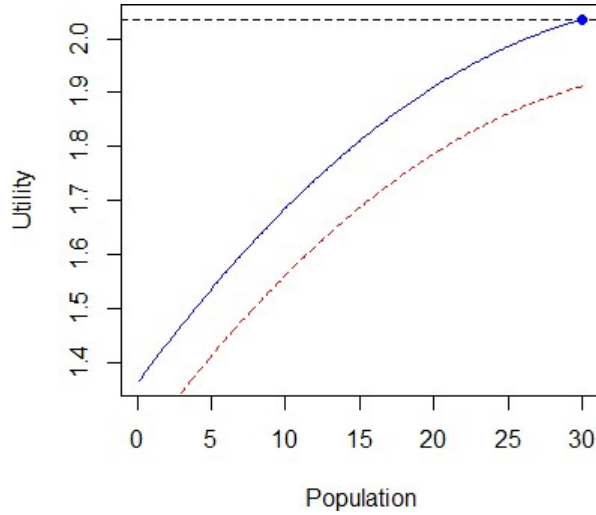
Figure 1: Single region equilibrium where the whole population lives in the best region.

For the two-region equilibrium (Figure 2), total population is 100. The 10 percent relative difference in geographic endowment translates into a 22 percent relative difference in population density.
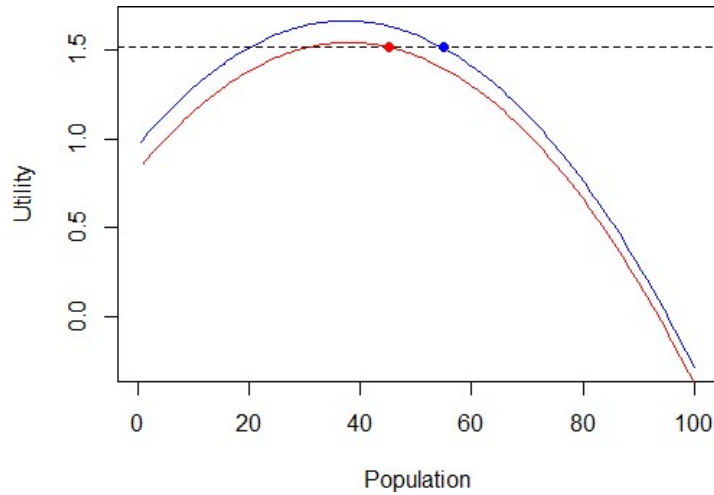

Figure 2: Two-region equilibrium.

Let us now discuss briefly how equilibria would change if there were moving costs c and information asymmetries, so that people from one region could not know exactly what their utility would be if they moved to the other region. Notice that they will remain in their regions if the expected increase in utility is lower than the moving costs:

$$u_2 + \varepsilon_1 - u_1 < c, \quad u_1 + \varepsilon_2 - u_2 < c$$

where ε_i represents the information asymmetry, that is, the error of households in region i when trying to anticipate what their utility would be after moving. Notice that derivatives do not play any role in this case. Moreover, as shown in Figure 3, two-region

equilibrium is no longer guaranteed to be Pareto efficient, since the fact that those who move bear the whole cost of moving could deter people from doing so, even when the social gain from the utility increase produced by reducing the over-population externality could be much higher than the private cost of moving. In this situation, a government intervention could make everybody better off by, for example, subsidizing the moving cost. In other words, a non-Pareto efficient equilibrium with moving cost (left figure) can be Pareto improved (right figure) if moving costs -grey area- are subsidized (or if people are forced to move). Utility gains derived from the movement of people from blue region to red region correspond to red and blue rectangles.
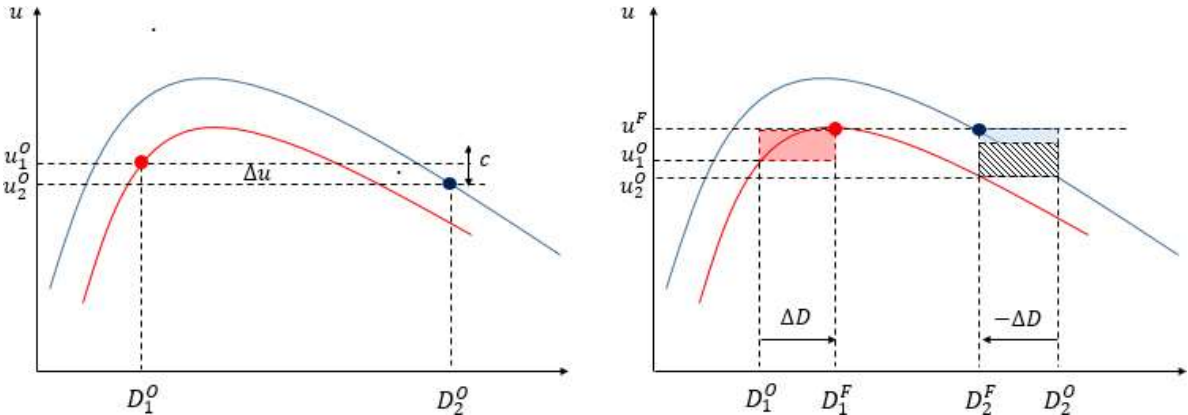

Figure 3. Two-region equilibrium and Pareto efficient equilibrium.

Implications

The model has three main implications. First, population distribution tends toward extreme outcomes and over-population. Geographic differences may lead to empty and over-populated regions, with respect to their optimal level. Moreover, relatively small differences in geographic endowment can also lead to much larger differences in population density. Second, the model predicts that areas with better geographic attributes will have higher densities of population (as in Beeson 2001; Mitchener and McLean 2003), unless there is much more public expenditure in worse areas or a historical legacy that cannot be overcome because of moving costs or incomplete information. In both cases, the result will be a non-Pareto equilibrium with lower utility and higher inequality. Third, more densely populated areas will tend to have higher output and capital per capita (as in Krugman 1991).[1]

---

[1] In Krugman the causality channel is due to externalities in production.

As a final remark, notice that in our model we have not considered any potential density-related externality regarding production function (such as economies of scale). Nonetheless, even with a simple neoclassical function, agglomerations emerge as an equilibrium as long as there are concave externalities with a monotonic change on population well-being. Including production externalities with a concave functional form, as for the households, will add an additional agglomeration force to the equilibrium, further reinforcing the implications of the model.

Data

In this section we present the data used to estimate the model on population density across European regions, which is presented in the following section.

Level of aggregation

As the aim of our article is to identify to what extent geographic drivers explain differences in population density, we need delimited regions to be small enough to consider their geographic attributes as representative of the region, and large enough to preclude specific municipalities or metropolitan areas that are particular agglomerations within a region. Therefore, we choose the level of aggregation denoted by NUTS 2 from the Nomenclature of Territorial Units for Statistics of Eurostat, which have populations of between 800,000 and three million. The current NUTS 2016 classification, valid since 1 January 2018, lists 323 regions at NUTS 2 level.

We consider only EU member states, because other countries included in NUTS classifications, such as Switzerland or Norway, are missing key data from Eurostat. We also exclude EU territories located outside Europe, namely five from France (Overseas France), three from Spain (Ceuta, Melilla and Canary Islands), and two from Portugal (Açores and Madeira). Moreover, we include Bremen (DE50) in Lüneburg (DE93), Hamburg (DE60) in Schleswig-Holstein (DEF0), and we group all five London NUTS regions into one single NUTS 2 area. We end up with 258 NUTS 2 specimens. We consider density of population as at January 1, 2019.

Geographic indicators

We consider several potential geographic drivers of population density (as in Mitton 2016): temperature, rainfall, access to navigable waters and unevenness of the land.

Temperature

We use the daily heating and cooling degree days of each region. Degree days are used as an

indicator of energy demand for the heating or cooling of buildings by comparing the day's average outside temperature against the optimal threshold of 18°C. The heating degree day (HDD) is the number of degrees below the threshold, the cooling degree day (CDD) the number of degrees above it. We compute average HDDs and CDDs during the period 1977 to 2018, using data from Eurostat.

Daily rainfall

Average daily rainfall (l/m^2) has been calculated using Copernicus Project data from 1977 to 2018.

Access to navigable waters

We have considered two main variables to capture access to navigable water. The first variable is the distance in kilometers from the boundary of the region to the nearest sea (or ocean). The second variable is direct accessibility to navigable rivers, that is, the kilometers of navigable river per squared kilometer of the NUTS2. We consider navigable rivers as those rivers with more than 100 segments or with a Strahler index higher than 5, according to the European Environment Agency Spatialite file for rivers. Figure 4 shows the map of the navigable rivers considered.

Figure 4. Main rivers in Europe.

Unevenness

We calculated unevenness as the interquartile range of the height of every LUCAS (Land Use and Coverage Area frame Survey) grid point of each NUTS 2 area. LUCAS is a survey carried out by Eurostat every three years to identify changes in land use and coverage. It contains observations from over 1,000,000 points.

Natural resources

We have considered as natural resources the presence of coal mines or oil refineries within the boundaries of the region. Data have been obtained from the Refineries Sites in Europe Database (Concawe Organitzation) and from the European Commission. Table 1 shows a description of the geographic data.

Table 1: Summary of the data set

Variable	Mean	St. Dev
Daily HDD	7.93	2.35
Daily CDD	0.16	0.24
Daily rain (l/m2)	2.07	0.60
Distance to nearest sea (km)	116.67	145.55
River density (km river/km2)	17.34	17.83
Unevenness (km)	0.2440	0.2434
Natural resources (%)	39.45	48.97
N = 258		

Geographic endowment

Our approach to determining the relationship between geographic variables and population density is to build a linear regression model over the logarithm of the density of population, using the geographic indicators described, that is:

$$GE(i) \equiv \log(Density_i) = \sum_{j=1}^{J} \alpha_j \, GI_{ij} + u_i \qquad (6)$$

where $GE(i)$ refers to the geographic endowment given to the i-th NUTS 2 region, that is defined as the logarithm of its density of population. $GI_{i,j}$ is the value of the j-th geographic indicator of the NUTS 2 region i, and u_i is the residual.

We exclude regions containing capital cities, since their population density can be strongly biased by political intervention. As the theoretical model suggests that small differences in geographic endowment can lead to substantial differences in density of population, the geographic indicators are included in quadratic form. Moreover, we distinguish between warmer and cooler seas by introducing an interaction between distance to the sea and temperature. Table 2 shows the results of a model built by weighting regions by their area size, so smaller regions are less important as the model is about density, and by selecting significant variables at 15 percent or higher. We perform two estimations. First, a Bayesian estimation using the *brms package* available in R (Bürkner 2017), without a prior to avoid introducing any bias. The Bayesian estimation does not assume asymptotic properties of the estimates and

therefore is more suitable for significance tests on small sample sizes (Figure 5). Secondly, we also provide OLS estimates. As can be seen in Table 2, estimates are almost identical.

Table 2: Geography endowment model

Variable name	Bayesian estimation		OLS	
	Estimate	Std. Error	Estimate	Robust Std. Error
Constant	1.65874***	0.42509	1.65620**	0.82357
HDD	0.13837***	0.06235	0.13642	0.11530
HDD^2	-0.01687***	0.00289	-0.01680***	0.00645
Rain	2.01921***	0.31475	2.03330***	0.63571
$Rain^2$	-0.36743***	0.06396	-0.37035***	0.13885
$Uneveness^{-1}$	0.02243***	0.00431	0.02244***	0.00503
$Uneveness^{-2}$	-0.00004***	0.00001	-0.00004***	0.00001
Distance to the sea	-0.01028***	0.00214	-0.01031***	0.00255
HDD: Distance to sea	0.00129***	0.00024	0.00129***	0.00029
Natural resources	0.29930***	0.08413	0.29795***	0.09853

Significance codes: ***: $p < 0.01$; **: $p<0.05$; *: $p< 0.1$.
N: 233 (capital regions not included)
Note: Errors weighted by area. R-squared: 0.7727. F-statistic: 84.2 (p-value: 0.000).
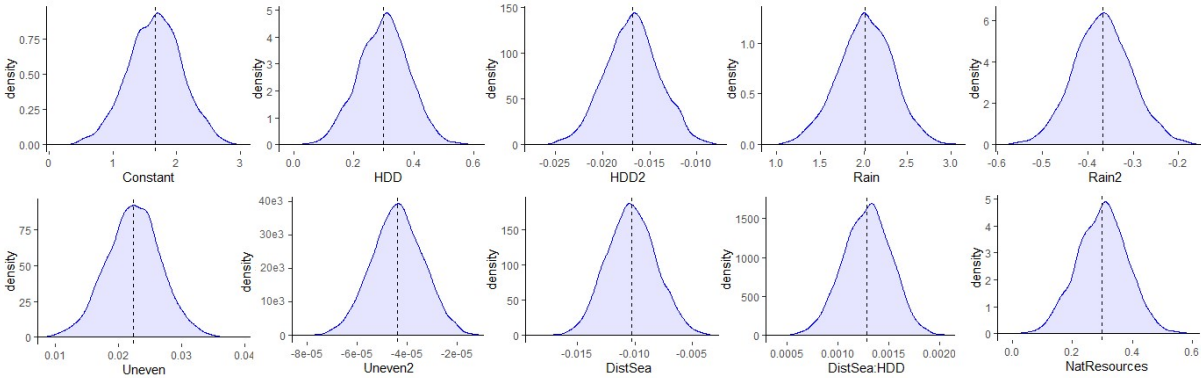

Figure 5: Distribution of the parameters of the model using a Bayesian estimation

Notice that CDDs are not important, and only HDDs matter. This is because Europe is a relatively cold area, with an average of only 0.16 CDD versus 7.93 HDD. Notice also that only distance to the sea is important in terms of access to navigable waters. It may seem counter-intuitive that rivers are not relevant, but the explanation is that most regions have a navigable river (72 percent of the regions, corresponding to 88 percent of the total European area). Since the river network is very dense across Europe, it has no significant impact at a regional level. However, as we will see later, rivers do determine distribution within a region, that is, where to place a city within a region.

Interpretation

Understanding why a model makes a certain prediction can be as crucial as the prediction's accuracy. It provides insight into how a model can be improved and supports understanding of the process being modeled. To do so, we evaluate feature importance and plot the marginal relation estimated by the model between explanatory features and the target (partial dependence functions).

To evaluate feature importance, we estimate SHAP (SHapley Additive exPlanation; Lundberg and Lee 2017, 2019) values. Given an observation $x = (x_1, \ldots, x_J)$, the SHAP value of feature j on the instance x corresponds to how the concrete value of feature j on x modifies the output of the model with respect to other instances that share some of the features of x, but not j. For a parametric model $F(x) = g(\sum_j \alpha_j x_j)$, where g is a function of the weighted features of x, the SHAP value corresponds to: $\varphi_j(x) = \alpha_j(x_j - E(X_j))$, where X is the set of observations and $E(X_j)$ is the average value of the j feature on X. Then, noting by N the total number of observations, we can estimate the relative importance of the feature j in the model as

$$RI_j = \frac{\sum_{i=1}^{N} |\varphi_j(x_i)|}{\sum_{k=1}^{J} \sum_{i=1}^{N} |\varphi_k(x_i)|}$$

Table 3 shows the relative importance of each driver.

Table 3: Relative importance of the geographic factors

Factor	Relative importance
Temperature	44.56%
Rain	19.00%
Access to navigable waters	13.12%
Unevenness	11.81%
Natural resources	11.51%
NOTE: SHAPs weighted by area.	

Temperature is by far the most important geographic factor explaining population density. Access to navigable waters has a relatively low importance, although that might be a singular characteristic of Europe, because most of the continent is made up of small surface peninsulas, so navigable seas and oceans are relatively close everywhere.

Apart from overall feature importance, it is also pertinent to understand whether the relationship between the target and a feature is linear, monotonic or more complex. Partial dependence functions estimate the marginal effect that features have on the predicted outcome

of a model (Friedman, 2001). Therefore, they correspond to SHAP values. As we can see in Figure 6, partial dependence plots show that geographic endowment decreases as the HDD increases; that is, the colder the region, the less attractive it is. In terms of rainfall, the curve suggests the more the better, but with diminishing returns.
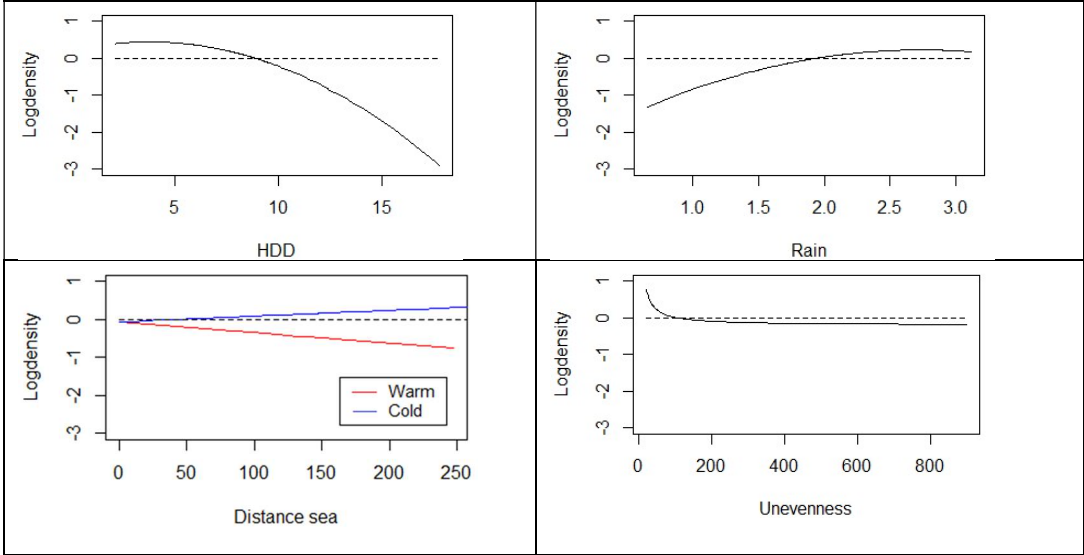

Figure 6. Partial dependence plots for non-binary variables.

The distance to the sea impacts differently according to temperature. For warm regions (in Figure 6, defined as having an HDD of less than 7, with an average of 4.99), such as Mediterranean countries, the effect is more relevant and positive: the closer to the sea, the better. For cold regions (in Figure 6, those regions with an HDD over 8, with an average of 9.61), the effect is almost not relevant and being closer to the sea does not increase geographic endowment. On the one hand, for cold countries with direct access to the sea, such as the UK or Finland, it is not relevant because almost all NUTS 2 areas have access. Moreover, the sea may not be always easily navigable, as it freezes. For cold countries without direct access to the sea, such as Austria or the Czech Republic, because they are crossed by navigable rivers, geographic dynamics are largely determined by the other factors.

Interestingly, as predicted by the theoretical framework presented above, geographic endowment tends to produce extreme population density outcomes, as can be seen in the quadratic relationships with respect to temperature and rainfall, or the inverse relationship with respect to unevenness.

Assuming that the greater the number of people who want to live in an area, the higher is its attractiveness, the geographic endowment (*GE)* can be interpreted as a proxy for the attractiveness of a region (to European people) based on its geographic attributes. Obviously,

there are several other factors that may influence population density, such as historical events (e.g., wars), cultural and religious differences or public investment. However, the *GE* constitutes an ideal framework for assessing to what extent population is distributed according to geography within a country and identifying the cause of misadjustment between current and expected distribution, and which of those, if any, are a consequence of deviating from geography.

Analysis of population distribution by country

Our analysis of population distribution by country comprises two parts. First, we analyze whether the choice of capital city is geographically optimal. Second, we estimate the degree of geographic harmonization for each country by the percentage of the population that would have to move to another region within the same country to achieve the expected distribution of population according to the geographic endowment.

Capital cities

To evaluate the choice of the capital location, in terms of geography, we calculated the relative potential of the NUTS 2 region in which the capital is situated with respect to the maximum potential that could be achieved within the city. The potential of the capital city of the country C is defined as:

$$Potential(C) = 100 \frac{GE(C)}{\max_{NUTS_i \in C} GE(NUTS_i)} \tag{7}$$

Table 4 presents the relative potential of the capitals of those countries considered in the analysis, with the exception of countries that consists of a single NUTS 2 region. It also presents the difference between the geographic factors of the capital and the average of those across the country. Although access to navigable waters is captured in the model by distance, we also included whether the capital city has a river. Rivers play an important role in deciding where to place the city within a region: all but Madrid (Spain) have a navigable river.

As shown by the theoretical model developed above, capital cities, which are highly populated areas with relatively high output per capita, are generally placed in nearly optimal areas, as their endowment almost reaches the maximum potential of the country (91.8 over 100). Madrid in Spain is the most notable exception, scoring only 59.5 over 100. It is placed in the very middle of Spain and on the high Spanish plateau, completely isolated from the sea, in a relatively colder and less rainy area.

Table 4: Capital attractiveness

Country	Potential	HDD vs avg	Rain vs avg	Sea dist vs avg.	Uneven. vs avg.	River
Germany	100	0.8	0.2	53.2	146.1	Yes
Croatia	100	-0.3	-0.4	-110.8	-131.7	Yes
Portugal	100	0.6	-0.5	37.6	-216.9	Yes
Romania	100	-1.6	0.1	-27.8	-152.1	Yes
Italy	100	-1.1	0.1	-57.3	-286.6	Yes
Ireland	99.4	-1.4	-0.3	0.0	-129.0	Yes
Netherlands	97.7	-1.6	-0.5	-15.7	-86.6	Yes
United Kingdom	97.7	-0.3	0.1	-48.2	54.1	Yes
Bulgaria	95.6	0.0	0.2	-20.3	-6.8	Yes
Sweden	95.4	-0.9	0.2	0.0	-154.8	Yes
Slovenia	95.1	-1.1	-0.2	-5.0	-190.7	Yes
Denmark	94.1	-0.1	0.0	0.0	-4.3	Yes
Czechia	92.1	0.0	-0.4	27.7	-219.7	Yes
France	89.4	0.2	-0.1	-22.6	-39.7	Yes
Greece	89.1	0.0	-0.2	0.0	-0.8	Yes
Poland	88.6	-2.4	-1.1	113.2	-496.3	Yes
Austria	88.1	-0.7	-0.1	22.2	-169.8	Yes
Hungary	86.9	-0.1	-0.2	30.7	4.9	Yes
Belgium	86.7	-1.4	0.0	0.0	-18.3	Yes
Finland	85.0	-1.0	-0.2	-87.6	-188.9	Yes
Slovakia	80.2	-0,4	-0,1	-9,7	-32,0	Yes
Spain	59.5	0,4	-0,4	200,5	-171,2	No
Average	91.8	-0.6	-0.2	3.6	-112.3	Yes

Population misadjustment

The GE is an estimation of the logarithm of the density of population. Thus, it can be used to estimate the expected density of population that every region should have if population were distributed according to geographic characteristics. For each country, we estimate the theoretical size of the population that should live in each region, subject to the constraint that total population cannot change. Let us note by *EP(NUTSi)* the expected population of region *i*, and by A_i its area. Then,

$$EP(i) = 100 \frac{e^{GE(NUTS_i)} A_i}{\sum_{k=1}^{N_C} e^{GE(NUTS_k)} A_k} P_C \tag{8}$$

where country C is the country where the i region is located, $k = \{1, ... N_C\}$ are the regions that in country C and P_C is the total population of C.

Given the expected population of every region and the real population, we estimate the percentage of total population that would have to move in order to achieve the expected distribution. We also adjust for the capital bias. Capital bias is the result of differences in how

17

countries classify their capital city. Some countries, such as Belgium, consider the capital city to be a NUTS 2 area. Others, such as Spain or France, consider the capital city with its metropolitan area to comprise the NUTS 2 zone. While others, such as Croatia, include their capital city in a NUTS 2 region that contains not only the city and its metropolitan area, but also a significant part of the country. The larger the territory included in the NUTS 2 sector, the lower the misadjustment due to the capital, since it is diluted among a larger area. To correct for this, we adjust the estimation of population misadjustment by country, delimiting the NUTS 2 area of all capitals according to the boundaries of their metropolitan region, splitting the given NUTS 2 zone into two regions, when needed.

Table 5 presents the results of population misadjustment. On average, 24.4 percent of the population would have to move within their country in order to achieve a purely geography-based distribution of population, in a range that extends from 14.4 percent (Bulgaria) to 35.6 percent (Spain). Without the correction for capital bias, results are very similar, with the exception of Croatia and Ireland, whose misadjustment would be underestimated because they include their capital within a NUTS 2 region far larger that the metropolitan area.

Table 5: Population misadjustment

Country	% misadjustment adjusted by capital effect	% misadjustment not adjusted by capital effect
Bulgaria	14.4	9.6
Slovenia	15.1	10.5
Czechia	16.0	15.9
Romania	16.3	16.4
Poland	16.9	14.8
Finland	19.2	18.7
Croatia	20.6	7.5
Slovakia	22.9	22.9
Sweden	24.0	24.2
Belgium	24.1	23.1
Germany	24.2	26.3
Italy	24.2	21.3
Austria	24.7	23.4
France	25.3	25.3
Netherlands	26.4	27.0
Hungary	26.4	26.1
Denmark	28.9	28.5
Ireland	31.5	30.0
Portugal	32.4	32.4
United Kingdom	34.1	35.3
Greece	34.2	34.2
Spain	35.6	35.6
Average	24.4	23.1

Figure 7 shows the comparison between current distribution of population and the expected distribution according to geographic endowment. Figure 8 shows the necessary change for regions to transition from current to expected distribution. That is, it shows the percentage increase needed for the resident population of the NUTS 2 area to meet expected population distribution. Green regions are those that are underpopulated with respect to the expected distribution. Therefore, more people should live there so there should be a population increase. Red regions are the opposite. They are overpopulated areas, whose population should decrease so as to achieve the expected distribution. In Portugal, for example, we can see that there is a highly overpopulated region, Lisbon (the dark red point), which would need to reduce its current population by more than 50 percent, and two moderately overpopulated regions that would need a decrease in current population of between 25 percent and 50 percent: the northern region (Porto) and the southern region (Algarve).
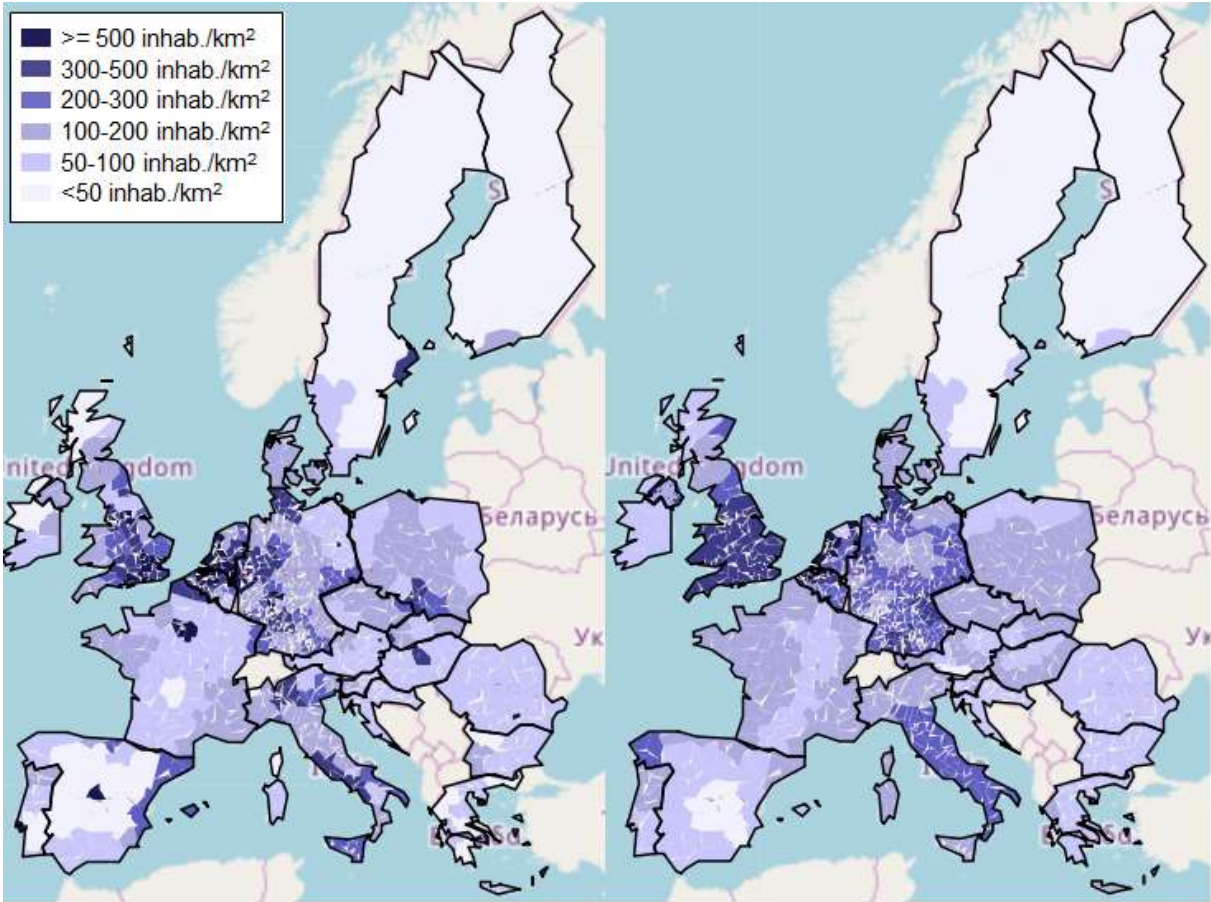

Figure 7. Current population distribution in Europe (left) vs. expected according to geographic endowment (right).
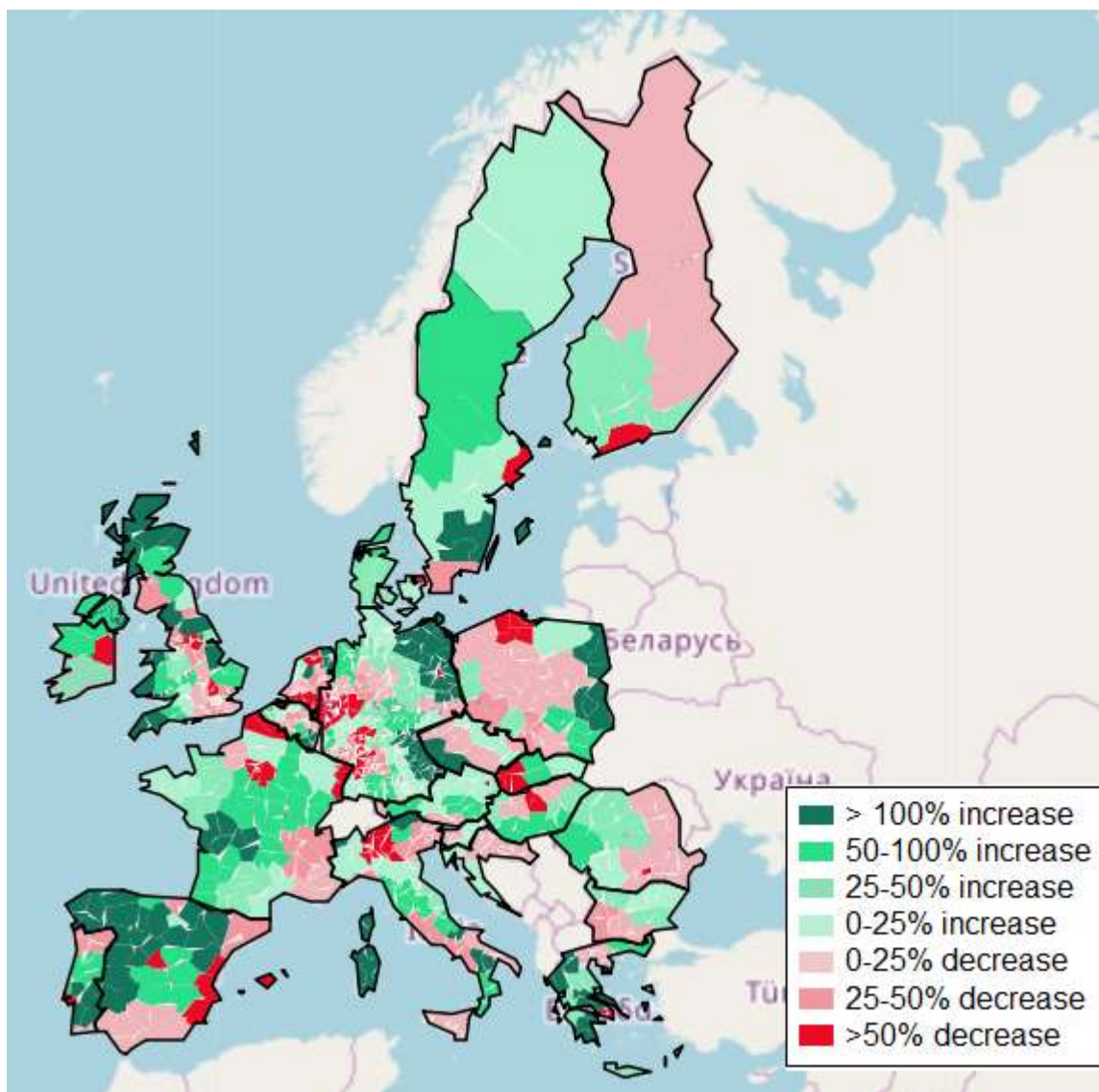
Figure 8. Percentage change in population by NUTS2 that would be needed to adjust actual population distribution with expected.

The theoretical model presented above introduced three potential drivers of greater deviations in population distribution in the long run: information asymmetry, moving costs and policy interventions. From results displayed in Table 5 and from Figures 7 and 8, the latter seems to emerge as the most relevant, at least for those countries in which the misadjustment is particularly high (above 30 percent): Spain, Greece, the United Kingdom, Portugal and Ireland.

These five countries share the trait of being peripheral in the context of the European continent. Hence, they are more isolated, which implies that any policy altering population distribution could be less affected by other countries' policies.

In the case of Ireland and the United Kingdom, parallel processes of consolidation took place as global hubs of large firms developed toward the end of the last century. Many

multinational firms relocated to the Dublin area in the 1980s and 1990s (Gunnigle and McGuire 2001), following Ireland's entry into the European Union (then the EEC) in 1973, and corporate tax reforms introduced in 1997 and 1999 by the minister of finance, Charlie McCreevy, lowered corporate taxes from 32 percent to 12.5 percent and thus laid the framework for Ireland's base erosion and profit shifting tools (BEPS), considered among the world's largest (Torslov, Wier and Zucman 2020). As a result, the Greater Dublin area has experienced an impressive population growth of 46 percent over the last 30 years (source http://www.greaterdublindrainage.com/) while the rest of the state grew by 30 percent, consolidating a secular trend for the preeminence of Dublin that dates back to the end of the Irish War of Independence (1919-1921) and the Civil War (1922-1923).[2]

In the case of the United Kingdom, the deregulation of financial markets in the 1980s (removal of controls on foreign exchange and fixed rate commissions, entry of foreign companies, switch to electronic trading, etc.), helped to kick off a financial transformation, dubbed the 'Big Bang', that cemented London as the global financial capital. Many financial institutions relocated to London, and its metropolitan area experienced a population growth similar to that of Greater Dublin. From 1991 to 2019 it grew by 34 percent, while the rest of the UK population grew by 16 percent.[3] This growth is far larger than that experienced in other important European capital areas such as Paris, Berlin, Rome, or Amsterdam. Among the large capital cities, only Madrid has experienced such a large growth of population (33 percent in Madrid vs. 20 percent in the rest of Spain).

Spain, Greece and Portugal, besides the common trait of being peripheral countries in the EU, share another characteristic: the three of them have French civil law systems and legal origins (La Porta *et al.* 2008). This has determined a type of nation-building based on the French model and a very centralized administrative tradition, which drive territorial policies persistently targeted to reinforce the political and economic role of the capital city in the country.[4] The paradigmatic case is that of Spain, which has been analyzed in Bel (2011, 2012). Other studies have documented the extreme degree of political centralization in Portugal

---

[2] After the Civil War, Dublin was consolidated as the political capital. Its population grew from around 500,000 inhabitants in 1925 to 1,025,000 in 1991. Cork, which was the main city opposing the Anglo-Irish Treaty (1921), lost political influence and economic relevance and its population only grew 112.000 inhabitants during the same period.
[3] Data obtained from https://worldpopulationreview.com/ on June 6, 2020.
[4] It is worth noting that even though it is not a peripheral country, France has a misadjustment rate of 25.3 percent, slightly above the average figure for Europe (24.4 percent). The fact that France is more centrally located in the continental context may have palliated the effect of this type of nation-building, and derived administrative tradition and territorial policies on population misadjustment.

(Magone 2011) and Greece (Ifantis 2004).[5] One of the probable consequences of these nation-building policies has been a particularly strong promotion of concentration of the population in the capital city of the country, which is the most important factor explaining the extremely intense misadjustment of population in these countries.

Economic consequences of population misadjustment

According to the theoretical model presented above, deviations from nature in terms of population distribution lead to non-Pareto allocations of population, which may be perpetuated due to the high cost of moving, incomplete information, and overinvestment in overpopulated areas designed to compensate for overpopulation externalities. Hence, no matter the underlying reason, greater deviations from nature are expected to be associated with lower utility and, potentially, higher inequality. To test these two economic consequences of population misadjustment, in this section we present an empirical estimation of the impact of population misadjustment on economic growth and inequality.

Regional conditional convergence

Conditional convergence theory states that an economy grows faster the further it is from its own steady-state value (Barro and Sala-i-Martin 1995), which is conditioned by different covariates such as the saving rate or human capital (Mankiw, Romer, and Weil 1992). As it allows for different steady states, it is a widely used framework for analysis of the long-term drivers of economic growth (see, among others, Barro 1995; Sala-i-Martin 1997), and for testing convergence between regions located in different countries (see, for example, Cartone, Postiglione, and Hewings 2020).

To test whether population misadjustment has an effect on economic convergence, we estimate a model of the form:

$$g_i = \gamma_0 + \gamma_1 q_i + \gamma_2 s_i + \gamma_3 h_i + \gamma_4 pm_i + \gamma_5 c_i + \varepsilon_i \qquad (9)$$

where g_i is the GDP per capita growth rate of the region i over the period 2000-2018, q_i is the natural log of the initial GDP per capita, h_i is the human capital, s_i is the natural logarithm of the saving rate, pm_i is the population misadjustment of the country in which the region i is situated (as at the year 2000), c_i is a dummy variable equal to 1 when the capital city

[5] In Lundell (2004), Spain, Greece and Portugal appear as the countries with the most centralized systems of party candidate selection for elections in the (then) EU15.

is in the region i, and ε_i is the residual. Human capital is measured as the natural logarithm of the percentage of the population aged between 25 and 64 years that has the highest education level (ISCED level 5–8, corresponding to tertiary levels), and the saving and investment rate as a share of GDP, as at the year 2000. Data have been obtained from Eurostat.

We present three different estimations of equation (9). In the first estimation, all regions are weighted equally in the quadratic error minimization. In the second estimation, we weight regions by total population. Therefore, regions with larger populations are considered to be more representative, to contain more information. Finally, we also present an estimation that weights regions according to their relative population within their respective countries. Hence all countries contribute equally to the model.

Table 6 presents the results for the three estimations of the conditional convergence model. We only provide OLS results because they are almost identical to Bayesian estimates. It can be seen that population misadjustment has a negative impact on economic convergence. This result is consistently statistically significant at 1 percent across models. Countries that deviate farther from nature tend to have less growth and lower steady states. Therefore, the greater the deviation the lower the speed of convergence and the inter-regional inequality decrease, both non-weighted and weighted by population (Concept 1 and Concept 2 inequality, as dubbed by Milanovic 2005), and the lower the total welfare of the population. Concretely, for each 10 percentage points of misadjustment, annual growth is reduced by around 0.5 to 0.7 percentage points. Moreover, as expected from the theoretical model, the positive estimate of the capital city parameter confirms that part of this effect consists of a rent transfer to over-populated regions by means of a public intervention.

Finally, to gain a complete understanding of the convergence model beyond the significance of the parameters, we estimate the relative importance of each variable included in the model. We use again the methodology suggested by Lundberg and Lee (2017, 2019): SHAP (SHapley Additive ExPlanation) values. Table 7 presents the relative importance of each variable in the conditional convergence model. The initial GDP is the most relevant variable and accounts for 50 percent to 60 percent of the total predictive power of the model, depending on the estimation. Remarkably, the population misadjustment and the capital city effect jointly account for around 30 percent of the total importance, being even more relevant than human capital.

Table 6: Regression estimates for conditional convergence model from 2000 to 2018.

Variable	Equal Weights	Weighted by population	Weighted by relative population
Constant	5.17886***	5.61842***	4.46462***
	(0.3598)	(0.2787)	(0.5991)
$logGDP^{2000}$	-0.42146***	-0.47199***	-0.37592***
	(0.0351)	(0.0237)	(0.0519)
Misadjustment	-1.40105***	-1.24990***	-1.07203***
	(0.1706)	(0.2003)	(0.2656)
LogEducation	0.11069***	0.12838***	0.05603+
	(0.0244)	(0.0265)	(0.0381)
LogSavings	0.05482	0.01127	-0.03508
	(0.0541)	(0.0528)	(0.1102)
CapitalCity	0.16702***	0.14164***	0.09247*
	(0.0506)	(0.0319)	(0.0566)
R^2	0.6313	0.6921	0.5811
F	82.18	107.9	66.6

N = 246. 9 regions (6 from Poland, 3 from UK) excluded because of missingness of data.

Robust standard errors in brackets. ***: $p < 0.01$; **: $p < 0.05$; *: $p < 0.1$, +: $p < 0.15$

NOTES:

a) Estimations excluding Ireland, that changed their GDP calculations in 2015 which resulted in a >10% increase that year, lead to results with higher predictive power (especially for the first and third estimation, where the R squared increase up to ~0.72) and a <1% signification of the capital parameter in the third estimation. The estimated parameters are modified by less than 10% (relatively)

b) Including population growth, depreciation and technological progress as a covariate (as in Cartone, Postiglione and Hewings, 2020) does not modify the results (less than 10%, relatively), since the variable is not significant. However, 8 additional regions would be removed because of missing data.

Table 7: Relative importance based on SHAP value

Variable	Equal weights	Weighted By population	Weighted by relative population
$logGDP^{2000}$	52.5%	56.1%	60.6%
Misadjustment	23.4%	19.9%	23.1%
LogEducation	15.4%	17.0%	10.1%
LogSavings	0	0	0
CapitalCity	8.7%	7.0%	6.2%

Income inequality

Apart from the effect on economic growth and regional convergence, we also test whether countries that deviate farther from nature exhibit higher within-country income inequality (Concept 3 inequality, as dubbed by Milanovic 2005). As per the theoretical model, greater deviations are expected to be associated with higher income inequality to compensate (at least partially) for over-population externalities.

Table 8 presents the estimation of a model to test the relationship between inequality (measured as the GINI market income, obtained from Eurostat) and population misadjustment.

The relationship is significant and positive, implying that greater deviation from nature is associated with higher inequality.

Table 8: GINI market income vs population misadjustment

Variable name	Bayesian estimation		OLS	
	Estimate	Std. Error	Estimate	Robust Std. Error
Constant	41.8873***	3.4752	41.9160***	3.5740
HDD	0.3277**	0.1377	0.3263**	0.1383
Significance codes: ***: p< 0.01; **: p<0.05; *: p< 0.1. N: 22 countries R-squared: 0.2321. F-statistic: 6.045				

Discussion

Our results provide empirical evidence that large population misadjustments with respect to geographic endowment come at a cost. As expected from our model, the farther a country deviates from the expected population distribution based on its geographic endowment, the lower its regional convergence and the higher its economic inequality will be. Remarkably, part of the effect consists of a rent transfer to the capital city.

These results, together with those obtained in the section "Analysis of the population distribution by country; Population misadjustment", suggest a potential novel causality channel by which institutions affect economic performance.

Recent economic growth literature has emphasized the role of institutions in economic growth and development (Acemoglu, Johnson, and Robinson 2002; Acemoglu and Robinson, 2005; Rodrik, Subramanian, and Trebbi 2004; Mitton, 2016, among others). 'Institutions' is a broad term that includes the diverse, complex interaction of individuals, firms, states, legislation and social norms which make up a society's social, economic, legal and political organization (see North 1981).[6] According to Acemoglu and Johnson (2005), these institutions are intimately linked to the distribution of political power in society and, as such, regulate the relationship between ordinary private citizens and elites with access to political power. Rodrik, Subramanian, and Trebbi (2004) propose a taxonomy of four categories of institutions that can impact economic performance. Institutions are 1) market-creating; 2) market regulating; 3) market stabilizing, and 4) market-legitimizing.

---

[6] The definition of institutions in North (1981: 201-202) is "a set of rules, compliance procedures, and moral and ethical behavioral norms designed to constrain the behavior of individuals in the interest of maximizing the wealth or utility of principals."

Thus, institutions have enough mechanisms to reverse the outcomes of nature and induce or promote the population distribution that best serves specific societal goals. As we showed, by using a framework of incentives, regulations and investments, institutions are the most relevant drivers of a population distribution equilibrium with extreme deviations from nature (as in Acemoglu, Johnson, and Robinson 2002). In turn, population distribution influences economic growth and income distribution: by privileging certain regions with respect to others, institutions not only transfer rents to the privileged region, but also harm overall economic growth.

On the one hand, for the three European continental countries with the most extreme misadjustments (Spain, Portugal and Greece) the deviation is the result of an intentional political intervention by central government, based on the desire to maintain control over the territory by privileging their capital regions (Madrid, Lisbon, and Athens, respectively). Political intervention in the design of policies such as transportation infrastructure, prioritizing objectives related to the administrative and political concentration of power, and largely neglecting productivity-related objectives, has probably prevented the development of an efficient distribution system in the economy, thus damaging potential economic growth. This is shown in the evidence obtained, for example, in the case of Spain, in Albalate, Bel and Fageda (2012) and Bertoméu-Sánchez and Estache (2017). The result of such intense forced deviations from nature is, according to our econometric results, detrimental to economic performance.

On the other hand, the high concentrations of population in London and Dublin, which explain the greater misadjustment of the United Kingdom and Ireland, seem to be a consequence of public policies intended to promote the development of market forces and private industry located in those regions, taking advantage of the role of agglomeration economies with localized accumulated capital. However, this concentration may have come at the expense of other regions. For instance, in the case of the United Kingdom, as stated by Ronen Palan: "The Bank of England consistently pursued policies that favored the City's position as a world financial center, even when such policies were seen as harmful to the UK's mainland manufacturing needs." (Palan 2010: 165). Inner London's GDP per capita was 328% of the European Union average in 2010, compared with 70 percent in West Wales - the biggest gap in any EU state, according to Eurostat.

Rising inequality and economic performance differences between regions have become a relevant policy debate, and a desire to redress the balance is expressed all the way up to the top. In 2014, even the prime minister, David Cameron, said that for too long the UK economy

had been "too London-focused and too centralised".[7] He had already written, in 2009, that *"Over the last century Britain has become one of the most centralized countries in the developed world."*[8]

Conclusion

The expansion of knowledge and technological innovations in transportation and communication have led to claims of the end of geography; a world in which distance would not play any significant role in decisions about human settlements. In this article, we have analyzed whether the features of nature and geography still play a relevant role in economic and social outcomes, by facilitating or limiting location, concentration and growth of human settlements.

We have proposed a theoretical model to represent the way in which geography and nature can account for regional economic growth, through their effects on population density and distribution. This model has been empirically examined using data from NUTS 2 European regions. This has allowed us to identify a strong predictive power of first-nature variables to explain the regional distribution of populations, and to estimate the degree of geographic harmonization of the actual distribution of population compared to the predicted distribution.

After estimating the misadjustments between actual and predicted regional population distribution, we have analyzed their impact on relative economic performance, together with the impact of institution-related factors, such as the conditional of being the capital of the country. Our main results suggest that deviating from nature's outcomes has a significant negative effect on economic growth and can increase inequalities. Hence, societies that choose to exploit the opportunities of the best locations, according to natural endowment, rather than promoting a different distribution of the population across regions by means of institutional intervention, achieve better economic performance and more social cohesion. That is, policies that harmonize with nature and geography yield better social welfare and human well-being than those policies that conflict with them.

---

[7] This statement was acknowledged in the news article, "Regions to get £6 billion in government funding", *BBC News/Business,* July 7, 2014. Retrieved on June 6, 2020 (https://www.bbc.com/news/business-28190016).
[8] David Cameron, "A radical power shift" *The Guardian,* February 17, 2009. Retrieved online on June 6, 2020 (https://www.theguardian.com/commentisfree/2009/feb/17/cameron-decentralisation-local-government).

Appendix

Variable	Description	Source	Link
Population	Population by NUTS2, in 2000 and 2018	Eurostat	https://ec.europa.eu/eurostat/web/products-datasets/-/demo_r_pjangroup
Area	Area of the NUTS2	Eurostat	https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts
HDD	Average daily number of degrees below 18ºC from 1977 to 2018	Eurostat	https://ec.europa.eu/eurostat/web/products-datasets/-/nrg_chddr2_a
CDD	Average daily number of degrees above 18ºC from 1977 to 2018	Eurostat	https://ec.europa.eu/eurostat/web/products-datasets/-/nrg_chddr2_a
Daily rainfall	Average daily rainfall (l/m2) from 1977 to 2018	Copernicus Project	https://surfobs.climate.copernicus.eu/dataaccess/access_eobs.php
River density	Kilometers of navigable river per km2	EEA	https://www.eea.europa.eu/data-and-maps/data/european-catchments-and-rivers-network/rivers/spatialite-file
Distance to sea	Nearest distance from the boundary of the region to the sea	Google Maps	
Terrain unevenness	Interquartile range of the height (in km) of every LUCAS grid point of each NUTS 2 area	LUCAS-Eurostat	https://ec.europa.eu/eurostat/web/lucas/data/lucas-grid
Natural resources	Dummy variable indicating whether there are coal mines or oil refineries within the region	European Comission and Concawe	https://www.concawe.eu/refineries-map/; https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/eu-coal-regions-opportunities-and-challenges-ahead
Regional GDP per capita	Regional GDP per capita (PPP), in 2000 and 2018	Eurostat	https://ec.europa.eu/eurostat/web/products-datasets/-/tgs00004
Human capital	Natural logarithm of the percentage of the population aged between 25 and 64 years that has the highest education level (ISCED level 5–8), in 2000	Eurostat	https://ec.europa.eu/eurostat/web/products-datasets/product?code=edat_lfse_04
Saving rate	Natural logarithm of the percentage of the gross fixed capital formation (as a % of the regional GDP) by NUTS2, in 2000	Eurostat	https://ec.europa.eu/eurostat/web/products-datasets/product?code=nama_10r_2gfcf

References

Acemoglu, D., and Johnson, S. 2005. Unbundling institutions. *Journal of Political Economy* 113(5), 949-995.

Acemoglu, D., Johnson, S., and Robinson, J. 2001. The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review* 91(5), 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. *Quarterly Journal of Economics* 117(4), 1231-1294

Albalate, D., Bel, G, and Fageda, X. 2012. Beyond the efficiency-equity dilemma: Centralization as a determinant of government investment in infrastructure. *Papers in Regional Science* 91(3), 599-615.

Barro, R. J. 1998. *Determinants of Economic Growth: A Cross-Country Empirical Study*. Cambridge, MA: MIT Press.

Barro, R. J., and Sala-I-Martin, X. 1995. *Economic Growth Theory*. Boston, MA: Mc-Graw Hill.

Beeson, P. E., Dejong, D. N., and Troesken, W. 2001. Population Growth in US Counties, 1840-1990. *Regional Science and Urban Economics* 31, 669-699.

Bel, G. 2011. Infrastructure and nation building: The regulation and financing of network transportation infrastructures in Spain (1720-2010). *Business History* 53(5), 688-705.

Bel, G. 2012. *Infrastructure and the Political Economy of Nation Building in Spain, 1720–2010*. Eastbourne: Sussex Academic Press.

Bertoméu-Sánchez, S. and Estache, A. 2017. Unbundling political and economic rationality: A non-parametric approach tested on transport infrastructure in Spain. *Transport Policy* 59, 181-195.

Bhattacharyya, S. 2009. Unbundled Institutions, Human Capital and Growth. *Journal of Comparative Economics* 37(1), 106–210.

Black, D., and Henderson, V. 1998. *Urban evolution in the USA*. Working Paper No. 98-21, Brown University.

Botero, G. 1588. *The magnificence and greatness of cities*. Translated by R. Peterson, 1606. Reprinted 1979. Amsterdam: Theatrum Orbis Terrarum; Norwood, NJ: Walter J. Johnson.

Bürkner, P.-C. 2017. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80(1), 1-28.

Cairncross, F. 2001. *The Death of Distance 2*.0. London: Texere Publishing Limited.

Cartone, A., Postiglione, P., and Hewings, G. F. D. 2020. Does economic convergence hold? A spatial quantile analysis on European regions. *Economic Modelling* (forthcoming) https://doi.org/10.1016/j.econmod.2020.03.008

Diamond, J. 1997. *Guns, Germs, and Steel*. New York: Norton.

Duranton, G., and Puga, D. 2020. The economics of urban density. *Journal of Economic Perspectives* 34(3), (forthcoming). doi:10.1257/jep.34.3.1

Easterly, W.R. 2001. *The elusive quest for growth: economists' adventures and misadventures in the*

topics. Cambridge MA: MIT Press.

Easterly, W., and Levine, R. 2003. Tropics, Germs, and Crops: How Endowments Influence Economic Development. *Journal of Monetary Economics* 50(1), 3–39.

Ellison, G., and Glaeser, E. L. 1999. The geographic concentration of industry: Does natural advantage explain agglomeration? *American Economic Review* 89(2), 311–316.

Fleming, F. J. 1991. Profits and visions: British capital and railway construction in Argentina, 1854–1886. In *Railway Imperialism,* ed. C. B. Davis, K. E. Wilburn, and R. E. Robinson, 71–84. New York, NY: Greenwood Press

Friedman, J. H. 2001. Greedy function approximation: A gradient boosting machine. *Annals of Statistics* 29(5); 1189-1232.

Friedman, T. L. 2005. *The World is Flat*. New York, NY: Farrar, Straus and Giroux.

Gallup, J. L., and Sachs., J. D. 2001. The Economic Burden of Malaria. *The American Journal of Tropical Medicine and Hygiene* 64(1–2) (Suppl.), 85–96.

Gallup, J. L., Sachs, J. D., and Mellinger, A. D. 1999. Geography and economic development. *International Regional Science Review* 22(2), 179–232.

Glaeser, E. L., La Porta, R., Lopez-de-Silanes, F., and Shleifer, A. 2004. Do institutions cause growth? *Journal of Economic Growth* 9(3), 271-303.

Glaeser, E. L., and Shapiro, J. 2003. Urban Growth in the 1990s: Is city living back? *Journal of Regional Science* 43(1), 139-165.

Graham, S. 1998. The End of Geography or the Explosion of Place? Conceptualizing Space, Place and Information Technology. *Progress in Human Geography* 22, 165–185.

Gunnigle, P., and McGuire, D. 2001. Why Ireland? A Qualitative Review of the Factors Influencing the Location of US Multinationals in Ireland with Particular Reference to Labour Issues. *The Economic and Social Review* 32(1),43-67

Ifantis, K. 2004. State Interests, External Dependency Trajectories and the European Union: Greece, in *European Union Enlargement: A Comparative History,* ed. J. Elvert and W. Kaiser, 70-92. London: Routledge.

Kaplan, R. D. 2012. *The Revenge of Geography. What the Map Tells Us about Coming Conflicts and the Battle against Fate.* New York, NY: Random House.

Krugman, P. 1991. Increasing returns and economic geography. *Journal of Political Economy* 99(3), 483-499.

Krugman, P. 1993. First nature, second nature, and metropolitan location. *Journal of Regional Science* 33(2), 129-144.

Krugman P., and Venables A. 1990. Integration and the competitiveness of peripheral industry. In *Unity with diversity in the European economy: The Community's Southern Frontier,* ed. C. Bliss and J. Braga De Macedo, 56-75. Cambridge: Cambridge University Press

La Porta, R., Lopez-de-Silanes, F., and Shleifer, A. 2008. The economic consequences of legal origins.

Journal of Economic Literature 46(2), 285-332.

Lundberg, S. M., and Lee, S. 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems.*

Lundberg, S. M., and Lee, S. 2019. *Consistent Individualized Feature Attribution for Tree Ensembles*. Working paper, arXiv:1802.03888

Lundell, K. 2004. Determinants of Candidate Selection: The Degree of Centralization in Comparative Perspective. *Party Politics* 10(1), 25–47.

Machiavelli, N. 1519. *Discourses on Livy*. New York, NY: Oxford University Press, 1987.

Magone, J. M. 2011. The difficult transformation of state and public administration in Portugal. Europeanization and the persistence of neo-patrimonialism. *Public Administration* 89(3), 756-782.

Malthus, T. R. 1798. *An essay on the principle of population*. London: J. Johnson.

Mankiw, N. G., Romer, D., Weil D. N. 1992. A Contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics* 107(2), 407-437.

Milanovic, B. 2005. *Worlds Apart: Global and International Inequality 1950-2000*. Princeton, NJ: Princeton University Press.

Mitchener, K. J., McLean, I. W. 2003. The productivity of the US States since 1888. *Journal of Economic Growth* 8(1), 73-114.

Mitton, T. 2016. The wealth of subnations: Geography, institutions, and within-country development. *Journal of Development Economics* 118, 88-111.

Montesquieu, C. de S. 1748. *The Spirit of the Laws*. Paris.

North, D. C. 1981. *Structure and change in economic history*. New York: W.W. Norton & Co.

O'Brien, R. 1992. *Global Financial Integration: The End of Geography*. London: Pinter.

O'Sullivan, A. 2007. *Urban Economics*. McGraw-Hill/Irwin

Palan, R. 2010. International Financial Centers: The British-Empire, City-States and Commercially Oriented Politics. *Theoretical Inquiries in Law* 11(1), 149-176.

Partridge, M. D., Rickman, D. S., Ali, K., and Olfert, M. R. 2008. Employment Growth in the American Urban Hierarchy: Long Live Distance. *The B.E. Journal of Macroeconomics* 8, Article 10

Rietveld, P., Vickerman, R. W. 2004. Transport in Regional Science: The "Death of Distance" is Premature. *Papers in Regional Science* 83, 229–248.

Rodrik, D., Subramanian, A., and Trebbi, F. 2004. Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development. *Journal of Economic Growth* 9(2), 131–165.

Sachs, J., and Warner, A. M. 2001. Natural resources and economic development: the curse of natural resources. *European Economic Review* 45, 827–838.

Sala-I-Martin, X. 1997. I just run two million regressions. *American Economic Review* 87(2), 178-183

Senese, P. D. 2005. Territory, Contiguity, and International Conflict: Assessing a New Joint Explanation. *American Journal of Political Science* 49(4), 769-779.

Starr, H. 2005. Territory, Proximity, and Spatiality: The Geography of International Conflict.

International Studies Review 7(3), 387-406.

Torslov, T., Wier, L., and Zucman, G. 2020. *The missing profits of nations*. NBER Working Paper No. 24701