

Working papers series

WP ECON 21.13

***The spatial distribution of population in
Spain: An anomaly in European
perspective***

Eduardo Gutiérrez
Banco de España

Enrique Moral-Benito
Banco de España

Daniel Oto-Peralías
Universidad Pablo de Olavide

Roberto Ramos
Banco de España

Keywords: Economic Geography, Spain.

JEL Classification: R10.



Department of Economics

The spatial distribution of population in Spain: An anomaly in European perspective*

E. Gutiérrez*, E. Moral-Benito*, D. Oto-Peralías[‡], and R. Ramos*

*Banco de España

[‡]Universidad Pablo de Olavide

July 29, 2021

Abstract

We exploit the GEOSTAT 2011 population grid with a very high 1-km² resolution to document that Spain presents the lowest density of settlements among European countries. Only a small fraction of the Spanish territory is inhabited, particularly in its southern half, which goes hand in hand with a high degree of population concentration. We uncover through standard regression analysis and spatial regression discontinuity that this anomaly cannot be accounted for by adverse geographic and climatic conditions. The second part of the paper takes a historical perspective on Spain's settlement patterns by showing that the spatial distribution of the population has been very persistent in the last two centuries, and that the abnormally low density of settlements with respect to European neighbors was already visible in the 19th century, which indicates that this phenomenon has not emerged recently as a consequence of the transformations associated with industrialization and tertiarization. Using data on ancient sites, we find that Spain did not feature scarcity of settlements in comparison to other countries in pre-medieval times, suggesting that its current anomalous settlement pattern has not always existed and is therefore not intrinsic to its geography.

JEL Codes: R10.

Keywords: Economic Geography, Spain.

*We are very grateful to Francisco Beltrán-Tapia, Alfonso Díez-Minguela, Julio Martínez-Galarraga, Alicia Gómez-Tello and Daniel Tirado for sharing data on 19th century Spain; Sascha O. Becker and Francesco Cinnirella for Prussian data; Luca Salvati for Italian data; and Daniel Alves, Nuno Lima and Pedro Machuqueiro for Portuguese data. We also thank David Cuberes, David Le Bris, Javier Pérez, Diego Puga, Jacopo Timini and seminar participants at Banco de España for useful comments. José A. Torralbo Moreno provided excellent research assistance. Daniel Oto-Peralías gratefully acknowledges financial support by the Spanish Ministry of Science, Innovation and Universities through grant ECO2017-86780-R (AEI/FEDER, UE). The opinions and analyses are the responsibility of the authors and, therefore, do not necessarily coincide with those of the Banco de España or the Eurosystem.

1 Introduction

How human settlements are distributed across space shapes economic interactions and influences economic development (see e.g. Krugman (1991)).¹ This article explores the spatial distribution of population in Spain compared to that of other European countries. Using the GEOSTAT 2011 population grid, we uncover an anomaly in Spanish settlement patterns. We document that Spain presents a much lower share of inhabited areas than its European neighbors, which goes hand in hand with a higher spatial concentration of the population. Importantly, this pattern cannot be accounted for by geographic and climatic factors, and further evidence based on 19th century data suggests that it has not emerged recently but has deep historical roots.

The GEOSTAT 2011 population grid is the main data source that allows us to analyze the distribution of population and settlements at a very granular level. In particular, it provides the figures of population living in each squared kilometer within Europe. This information is crucial to properly measure the distribution of population across space as opposed to the traditional density indicators (based on the ratio of population to surface of administrative areas) that do not appropriately reflect the density actually faced by individuals and firms (see Duranton and Puga (2020)).² We thus consider four alternative indicators to comprehensively measure the spatial distribution of the population across European countries: (i) the ratio of population to surface within each 250-km² grid cell; (ii) the ratio of population to inhabited surface area, to better capture the actual density experienced; (iii) the so-called settlement density, given by the percentage of 10-km² cells that are inhabited within each 250-km² grid cell; and (iv) an indicator of population concentration given by the percentage of the population living in the most populated one percent of the territory within each 250-km² grid cell.

We document through standard regression analysis and spatial regression discontinuity (SRD) an anomaly of Spanish settlement patterns in European perspective that cannot be accounted for by geo-climatic conditions. In particular, the regions with the lowest amount of settlements relative to their surface within Europe are located in Spain, Iceland, Finland, Norway and Sweden. However, while the low settlement density is fully explained by adverse geographic and climatic factors in Scandinavian countries, this is not the case in Spain, which remains the only outlier in terms of low settlement density after accounting for geo-climatic conditions. Using the alternative measures of spatial distribution of the population, we conclude that Spain is characterized by a very low

¹Allen and Arkolakis (2014) show that geographic location alone can explain at least 20% of the spatial variation in income across the United States.

²According to Duranton and Turner (2018), most daily activities take place within an area of 10-km² that does not necessarily coincide with administrative areas. In addition, other dimensions beyond administrative units such as the number of centers within an area or the compactness of development might be relevant in determining urban outcomes (Harari (2020)).

settlement density and a very high concentration of the population in a small share of the territory. These two characteristics render a level of (traditional) population density somehow in line with other European countries, and a high density of population in inhabited areas. We further show that this anomalous settlement pattern is accentuated in the southern half of the country.

The second part of the paper takes a long-term perspective on this phenomenon by exploring its persistence over time and discussing its potential causes. We gather district-level data on several European countries during the second half of the 19th century and show that Spain already featured a very low density of settlements. This indicates that its anomalous pattern of settlements has not emerged recently, and therefore is not the consequence of the transformations associated with industrialization and tertiarization (e.g., rural exodus). We further show that the spatial distribution of the population has been fairly persistent in Spain during the last two centuries. More specifically, we document a high correlation of both population density and settlement density between 1787 and 2011.

As a final piece of evidence, we use ancient sites data to analyze whether Spain was already scarcely settled (in relative terms) before the Middle Ages. The results do not support this hypothesis as the presence of pre-medieval settlements is not abnormally low in Spain as a whole nor in its southern part, thereby suggesting that the current anomalous settlement pattern has not always existed and is therefore not intrinsic to the country's geography. Rather, the possible explanations should be found in the historical processes taking place during the Middle Ages and the Early Modern Period.

According to Oto-Peralías (2020), insecurity characterizing the Reconquest period may contribute to explain the anomalous spatial distribution of Spanish population. Continuous warfare and insecurity heavily conditioned the nature of the colonization process, characterized by the leading role of the military orders as colonizer agents, scarcity of population, and a livestock-oriented economy (González-Jiménez (1992)). This determined a spatial distribution of the population characterized by a low density of settlements and an economy based on ranching, as mobile assets were favored by military conditions (see Bishko (1975)). We further discuss below that other potential explanations such as the sectoral composition of the economy and the land planning regulation are very unlikely to significantly affect the anomaly described.

The rest of the article is structured as follows. Section 2 briefly describes the main data used in the analysis and takes a first glimpse at the data. Section 3 tests for the existence of an anomaly in settlements and population patterns in Spain, both using standard regression analysis and SRD design. We further split Spain into its northern and southern halves to show that the anomaly is particularly accentuated in the south. Section 4 explores the long-run persistence of the phenomenon and discusses its possible causes. Finally, section 5 provides some concluding remarks.

2 Data

2.1 The dataset

The GEOSTAT population grid provides data on population distribution in space at a very high 1-km² resolution (Eurostat (2016)). The sample consists of the territory covered by GEOSTAT 2011 after excluding the overseas regions (see Figure 1), an area slightly larger than 5 million km² with approximately 2.08 million populated 1-km² grid cells.³

We construct four indicators measuring different dimensions of the spatial distribution of population in European countries. First, a settlement density indicator that measures the distribution of settlements along the territory. More specifically, it refers to the percentage of 10-km² grid cells that are inhabited in each 250-km² cell. A 10-km² grid cell is considered to be populated if it contains at least one 1-km² populated cell within it. We choose 10-km² as cell area because it is a meaningful size from an economic point of view. For instance, Duranton and Turner (2018) conclude that most daily activities take place within an area of 10-km². Also, the average size of a commune in France is 15-km² and, typically, each commune has more than one settlement. It is worth noting that, according to this indicator, settlements are identified through the presence of populated 1-km² cells. An important advantage of this way of identifying settlements is its homogeneity across countries. Other alternatives such as data on municipalities or on other local administrative units cannot be used for comparative purposes since they are heterogeneous across countries.

Second, we compute an indicator of population concentration that measures the percentage of the population living in the most populated one percent of the territory within each 250-km² grid cell. It is worth stressing that the level of spatial aggregation used is important for this indicator. Population concentration may be high at the country level but moderate or low at the sub-national level. However, there is in practice a strong correlation (0.86) between the country level value and the average of grid-cell level values.

Third, we consider a traditional indicator of population density given by the ratio of population to surface in each 250-km² grid cell. Fourth, we compute an indicator of “experienced density” as the ratio of population to inhabited surface area in each 250-km² grid cell, aimed at better capturing the actual density experienced by the population.⁴ Regarding the correlation among these indicators,

³The 34 European countries included are Albania (AL), Austria (AT), Belgium (BE), Bosnia and Herzegovina (BI), Bulgaria (BG), Croatia (HR), Czechia (CZ), Denmark (DK), Estonia (EE), Finland (FI), France (FR), Germany (DE), Greece (EL), Hungary (HU), Ireland (IE), Italy (IT), Kosovo (KO), Latvia (LV), Lithuania (LT), Luxembourg (LU), Macedonia (MK), Montenegro (ME), Netherlands (NL), Norway (NO), Poland (PL), Portugal (PT), Romania (RO), Serbia (SR), Slovakia (SK), Slovenia (SI), Spain (ES), Sweden (SE), Switzerland (CH) and the United Kingdom (UK). We do not include Liechtenstein, Iceland, and Malta in our analysis. The first one for being too small and the others because they are outliers.

⁴This is a simplified version of the measure proposed by Roca and Puga (2017).

settlement density is negatively correlated with population concentration (-0.75) and positively with the logarithm of population density (0.77) and experienced density (0.31); population concentration is negatively correlated with population density (-0.58) and experienced density (-0.29); and the two latter are highly positively correlated (0.87).

We also collect data on a wide array of geographical, climatic, and historic variables, including temperature, rainfall, altitude, ruggedness, soil quality, distance to the coast, population density in 1871, etc. The definitions and sources of all the variables used in the analysis as well as their descriptive statistics are provided in Table A1 and Table A2, respectively, in Appendix A in the Supplementary Material.

2.2 A first glimpse at the data

We describe in this section the main features of the population location in Spain in comparison with other European countries. Compared to other European Union countries, population in Spain stands out for two characteristics. First, total population is low. And second, it is highly concentrated. The first two columns of Table 1 show that, despite being the second country by land area, Spain only ranks 5th in terms of total population. For example, the number of inhabitants is 20% lower than in Italy, despite the land area being 65% larger.

The low number of people living in Spain stems from the fact that there is a very high amount of uninhabited areas. Figure 1, which plots in red the inhabited 10-km² grid cells in Europe, shows that this feature stands out in sharp contrast with other countries. Indeed, Spain appears comparable only to areas where geographical and climatological conditions deter population settlements, such as the Scandinavian Peninsula or the Alps. Column 4 of Table 1 shows that only 13% of the Spanish land area is inhabited, which is the lowest value in the European Union.

This exceptionally low settlement density makes population in Spain be very concentrated. Indeed, column 5 of Table 1 reveals that the population density in inhabited areas reaches 737 people per squared km, which is the second largest in Europe (being the largest the small island of Malta). If we compute Lorenz curves estimating the inequality in the spatial location of population along the inhabited 1 km² grid cells, one can see that Spain is the most spatially concentrated country among the selected ones. For example, while in Germany, Poland and Portugal 15% of the most populated cells account for 80% of total population, in Spain they account for 90%. The corresponding figure for France, Italy and the United Kingdom is 85%.

All in all, Table 1 suggests that population in Spain is very concentrated in the space because there is an abnormally large number of uninhabited areas in comparison to other European countries. According to these measures, we are able to identify a Spanish anomaly in settlement patterns.

Table 1: Population Density in EU-28 Countries (2011).

	Surface (km ²) (1)	Population (thousands) (2)	Density (3)	Inhabited area (%) (4)	Density inhabited area (5)
France	549,060	62,765	114	67.8	168
Spain	498,504	46,816	94	12.7	737
Sweden	449,896	9,539	21	25.2	84
Germany	358,327	80,213	224	59.9	374
Finland	337,547	5,340	16	30.0	53
Poland	313,851	38,500	123	62.6	196
Italy	301,291	59,429	197	57.2	345
United Kingdom	247,763	63,154	255	51.7	493
Romania	239,068	20,122	84	29.4	287
Greece	131,912	10,634	81	19.7	409
Bulgaria	110,995	7,365	66	21.3	312
Hungary	93,013	9,938	107	29.8	358
Portugal	88,847	10,562	119	46.6	255
Austria	83,944	8,402	100	51.4	195
Czech Republic	78,874	10,437	132	56.0	236
Ireland	70,601	4,575	65	80.2	81
Latvia	65,519	2,081	32	50.5	63
Lithuania	65,412	3,029	46	54.5	85
Croatia	56,539	4,290	76	43.4	175
Slovakia	49,035	5,399	110	32.4	340
Estonia	45,347	1,294	29	45.6	63
Denkmark	43,162	5,535	128	90.4	142
Netherlands	37,824	16,651	440	81.0	544
Belgium	30,668	10,990	358	83.0	432
Slovenia	20,277	2,049	101	66.0	153
Cyprus	9,249	840	91	37.3	244
Luxembourg	2,595	513	198	65.5	302
Malta	315	417	1325	92.7	1430

Source: Eurostat.

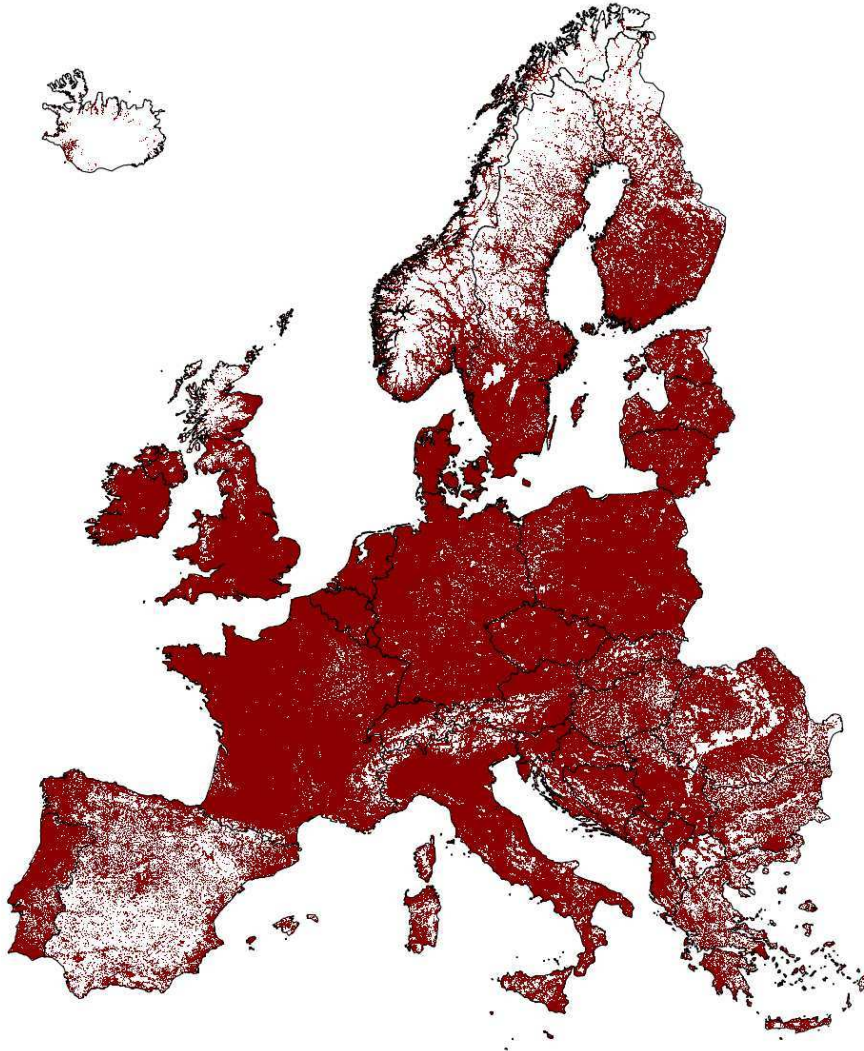
Note however that this anomaly is somehow masked when looking at traditional population density measures at the aggregate level (inhabitants per square km) in which Spain ranks slightly below the mean (see column 3).⁵

3 Testing the Spanish Anomaly

The evidence presented in Table 1 and Figure 1 is suggestive of a Spanish anomaly in population patterns. However, Spain is characterized by high temperatures and large mountainous lands that might account for the high prevalence of uninhabited areas. This section explores whether geographic and climatic factors may explain the differences between Spain and other European countries in the distribution of population across the territory.

⁵Rae (2018) shows inhabited areas density in 39 European countries to reflect that the traditional measure is not complete.

Figure 1: Inhabited 10-km² Grid Cells in Europe (2011).



Source: Eurostat.

Notes: This figure depicts in red (white) the inhabited (uninhabited) 10 km² grid cells in Europe.

3.1 Regression analysis

The following regression allows us to better understand the role of geography and climate in explaining the Spanish anomaly in terms of settlement density and population density:

$$y_{ic} = \eta_c + x'_{ic}\beta + \epsilon_{ic} \quad (1)$$

where y_{ic} refers to one of the four indicators of settlement and population patterns in 250-km² grid cell i belonging to country c . x_{ic} is a vector of geographic and climatic variables for each grid cell including temperature, rainfall, average altitude, ruggedness, soil quality and distance from the coast. In order to allow for non-linear effects of those variables on settlement patterns, quartile dummies are included with the omitted category being the first quartile. An island dummy and a quadratic polynomial in latitude and longitude are also included in the vector x_{ic} .⁶ Our main object of interest in equation (1) is the vector η_c that refers to a full set of country dummies. In order to test the significance of the Spanish anomaly after accounting for geography and climate, we analyze the estimated country dummies ($\hat{\eta}_c$) from equation (1) and their associated standard errors. France is the omitted category in all cases so that the estimated country dummies reflect the average difference in settlement density between each country and France.

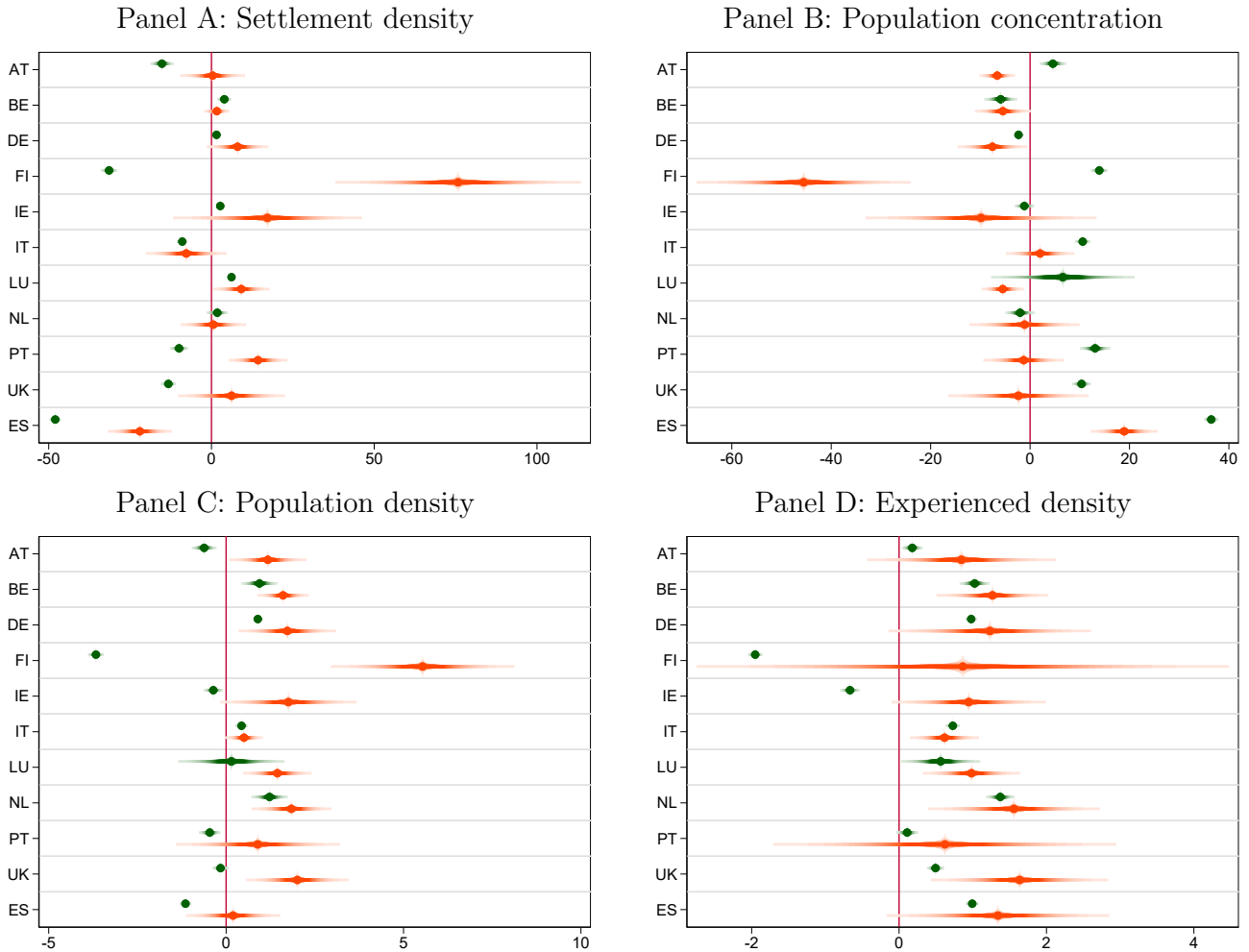
Figure 2 presents the estimated country fixed effects without controls (green dots) and with geographic and climatic controls (red dots) for the four indicators of settlement and population patterns.⁷ Panel A (settlement density) shows that Spain is the country with the lowest share of inhabited 10km² areas within each 250-km² grid cell. This is true when looking at the unconditional country averages as well as after controlling for geo-climatic controls. More specifically, the average 250-km² grid cell in Spain presents a settlement density 50 percentage points (pp.) lower than the corresponding average cell in France, while this figure is around 30 pp. in Finland, as shown by the green dots in Panel A. Interestingly enough, this difference becomes positive and statistically significant in the case of Finland after accounting for geographic and climatic factors, while it remains negative and significant in Spain, albeit smaller in magnitude (red dots). Thus, Panel A in Figure 2 clearly illustrates that geography and climate cannot fully explain the Spanish low settlement density, which is the lowest within Europe.

This pattern remains unaltered if we use population concentration instead of settlement density as the dependent variable in our regressions as shown in Panel B. Population concentration is thus the highest in Spain even after accounting for geography and climate. Indeed, both variables present a very high and negative correlation of -0.75. Turning to population density in Panel C, the difference between the average Spanish and French 250-km² grid cell is not statistically significant once we account for geo-climatic conditions. Finally, Panel D shows that in terms of experienced density (i.e., the ratio population to inhabited surface area), Spain is placed among the countries with the

⁶Table A3 in Appendix A in the Supplementary Material shows the estimated coefficients for the geographic and climatic variables included in the regressions.

⁷In order to focus on a set of countries with a similar level of economic development, we present the results for the sample of Eurozone founding countries and the UK. Figure A1 in Appendix A shows that these findings also hold in a larger set of European countries. Following Henderson et al. (2017), who show that some geographical characteristics are primarily important for agriculture, in this larger sample we interact the geo-climatic factors with the Eurozone dummy, and hence we account for the fact that Eastern European countries have a larger agriculture share.

Figure 2: The Spanish anomaly in settlement and population patterns.



highest values, reflecting that people tend to live concentrated in towns and cities, with a lower presence of dispersed settlement.

A potential concern is that we are comparing countries that are very heterogeneous in surface area. One could suspect that the singularity of Spain is perhaps conditional on the specific territorial division used in our analysis. In order to address this concern, we compare Spain to a set of virtual countries with similar surface areas. For that purpose, we randomly choose 1,000 grid cells and then create virtual countries by selecting up to 2,500 neighbors falling within 700 km from the cell's centroid (similar to Spanish land area). Then we run 1,000 regressions of settlement density on the full set of geo-climatic controls, the Spain dummy, and the virtual countries dummies (included one by one). Remarkably, in every case the coefficient on Spain for settlement density is lower than

on the virtual region (see Figure A2 in Appendix A). We find the opposite results for population concentration: Spain always carries the highest coefficient. This figure is about 20% in the case of population density, which somehow confirms the lack of statistical significance of the Spanish coefficient in Panel C of Figure 2, while it is 99.8% for experienced density (i.e., experienced density is almost always higher in Spain than in the virtual countries).

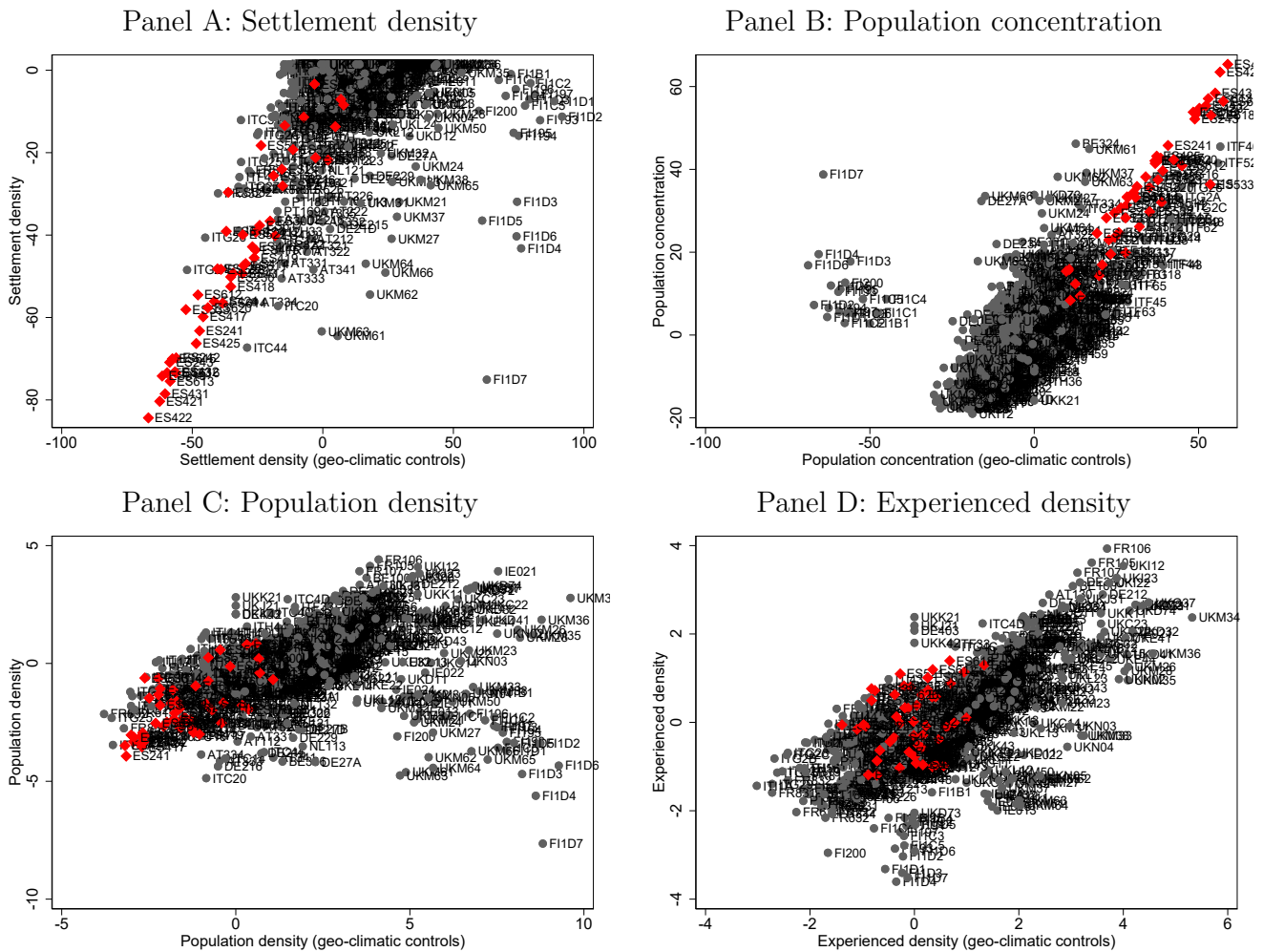
In order to further explore the Spanish anomaly, we now turn to the analysis at the regional level. In particular, we estimate models analogous to equation (1) but instead of including country dummies, we now include region dummies at the NUTS3 level corresponding to Spanish provinces. The omitted category is now the Paris area (FR101) so that all estimated dummies capture the difference with respect to Paris. Panels A to D in Figure 3 depict the estimated coefficients on the region dummies without geo-climatic controls (y-axis) and with those controls (x-axis). This allows to simultaneously visualize the distribution of coefficients for both specifications. The results show that a large number of Spanish regions (depicted in red) locate at the extremes of the distributions of settlement density and population concentration (Panels A and B, respectively). For instance, the 17 European regions with the lowest settlement density are located in Spain. In contrast, this anomaly is difficult to perceive in the case of population density (Panel C) or experienced density (Panel D).⁸

Finally, Figure 4 takes into account the existing heterogeneity in settlement patterns within Spain and splits the country into two parts, northern and southern Spain (*ES_N* and *ES_S*, respectively). Notably, the anomalous pattern of settlement density and population concentration is particularly accentuated in southern Spain. The latter has considerably less settlement density and more population concentration than northern Spain, which in turn stands out from the other countries. For instance, considering Panel B, in the average 250-km² grid cell in France, 25% of the population lives in the most populated 1% of the area, while in northern Spain this value increases to 53.5% and in southern Spain to 71.6%. Moreover, southern Spain also features the highest experienced density in terms of population, which reflects a remarkable lack of dispersed settlement in the territory.

One relevant thing to notice is that the GEOSTAT 2011 dataset represents census data for each European country in a common grid. While its methodology involves developing guidelines and methods to create a harmonized dataset, the spatial reference data as well as the production methods can vary by country (see EFGS (2011) and Eurostat and EFGS (2015)). In the case of Spain, Goerlich and Cantarino (2017) note that the Spanish grid is based on the aggregation of a large georeferenced census sample of dwellings, rather than the whole population of registered buildings. This can

⁸Figure A3 confirms the Spanish anomaly in population patterns using a larger set of European regions. In this case, the 7 regions with the lowest average settlement density are from Spain. Also, 9 out of the 10 regions with the lowest settlement density are located in Spain.

Figure 3: The Spanish anomaly by regions.

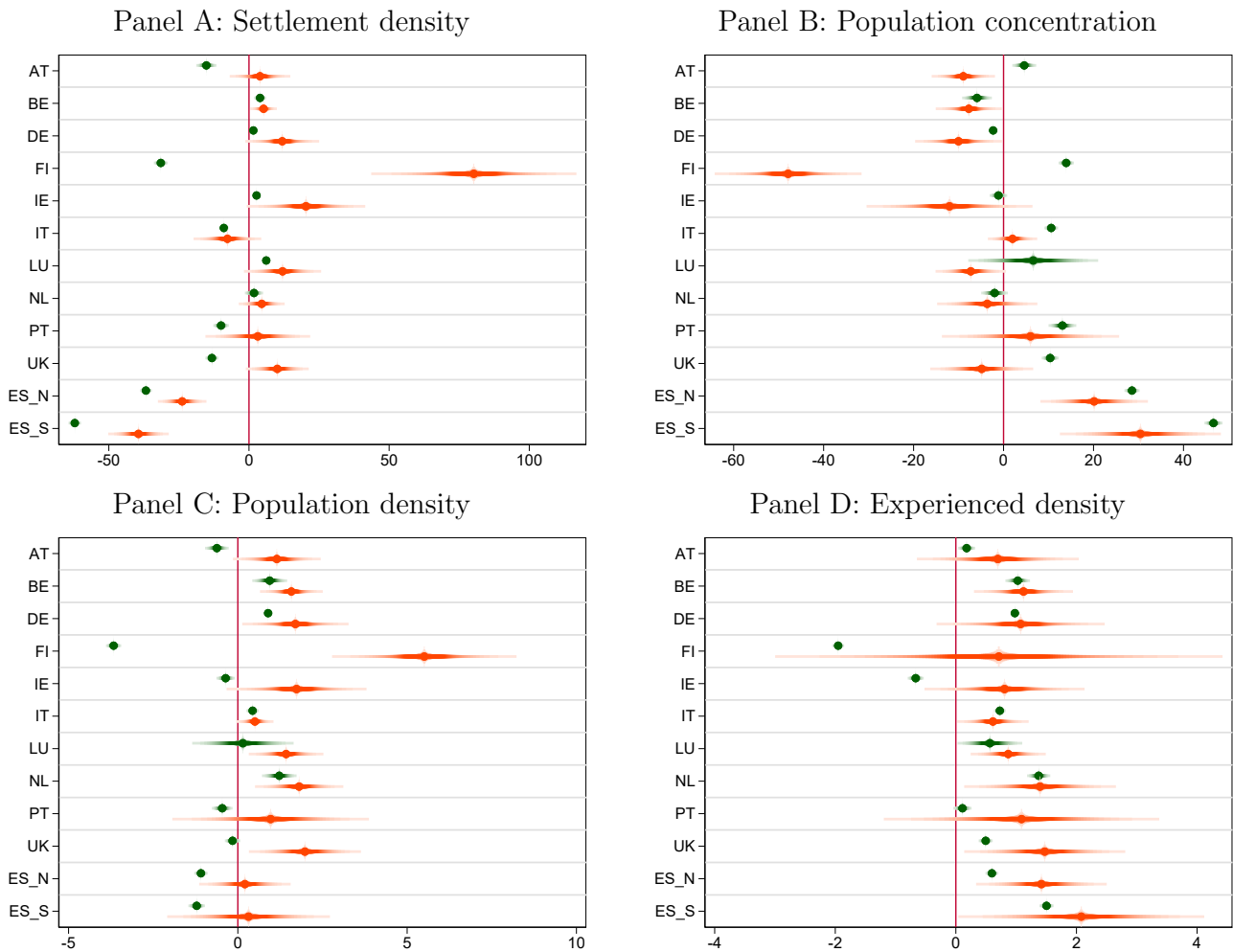


introduce a bias, insofar as dispersedly populated areas may appear uninhabited if buildings thereof are overlooked by the census sample.⁹ Goerlich and Cantarino (2017) estimate that the true amount of inhabited grid-cells in Spain could be around 16%, rather than the 13% reported by GEOSTAT 2011. While this figure would still keep Spain at the bottom of the percentage of inhabited grid-cells within the European countries (see Table 1), it is worth evaluating this methodological issue.

To address this concern, we repeat the baseline exercise with the GEOSTAT 2006 grid (Eurostat (2006)). In this dataset, the Spanish grid is constructed following a top-down approach, which consists of the disaggregation of census tract data on the basis of a land use dataset (Goerlich

⁹Note that this type of bias arises also if, for confidentiality reasons, inhabited grid-cells are converted to non-populated. One example of a country applying such policy in GEOSTAT 2011 is Germany.

Figure 4: The anomaly of northern and southern Spain in settlement and population patterns.



and Cantarino (2013)). As discussed in Goerlich and Cantarino (2017), this approach induces a bias (opposite to the previous one) since it overestimates the dispersion of the population compared to the more accurate bottom-up method followed in GEOSTAT 2011 (i.e. the aggregation of georeferenced points). This is so for mainly three reasons. First, errors in the land use dataset. Second, the existence of secondary residences, to which population are assigned, inducing a lot of false positives in the production process. And third, decisions on what building types can host population, which can create both false positives and false negatives.¹⁰ Hence, by performing the baseline analysis with the 2006 grid, we test the existence of a Spanish anomaly under a dataset that overestimates settlement

¹⁰An example put forward by the authors is religious buildings. Since most of the them are uninhabited, it is convenient to instruct the algorithm not to assign population to these buildings. This induces the creation of false negatives for the grid-cells that host a religious building that is actually inhabited.

density and underestimates population concentration, at least regarding the Spanish population patterns. For this reason, the following results can be interpreted as a lower bound on the Spanish anomaly.

The results can be found in Figure 5, which is the counterpart of Figures 2 and 4 using the 2006 grid dataset. We find that with the 2006 grid Spain still exhibits a very low settlement density and a very high concentration of the population, even after controlling for geo-climatic factors. However, it ceases to be the country at the extreme of the distribution, Italy showing a lower settlement density and Ireland a higher population concentration. Interestingly enough, when we split the country in its northern and southern parts, Southern Spain still exhibits the lowest settlement density and the highest population concentration and experienced density within Europe. Hence, the analysis with the 2006 grid confirms the idiosyncratic settlement patterns in Spain (especially in its lower half) though it dampens the anomaly when one looks at the whole country. Given that the 2006 grid tends to bias the distribution of the population towards more dispersion, the true results probably lie between those obtained with the 2006 and the 2011 grids.

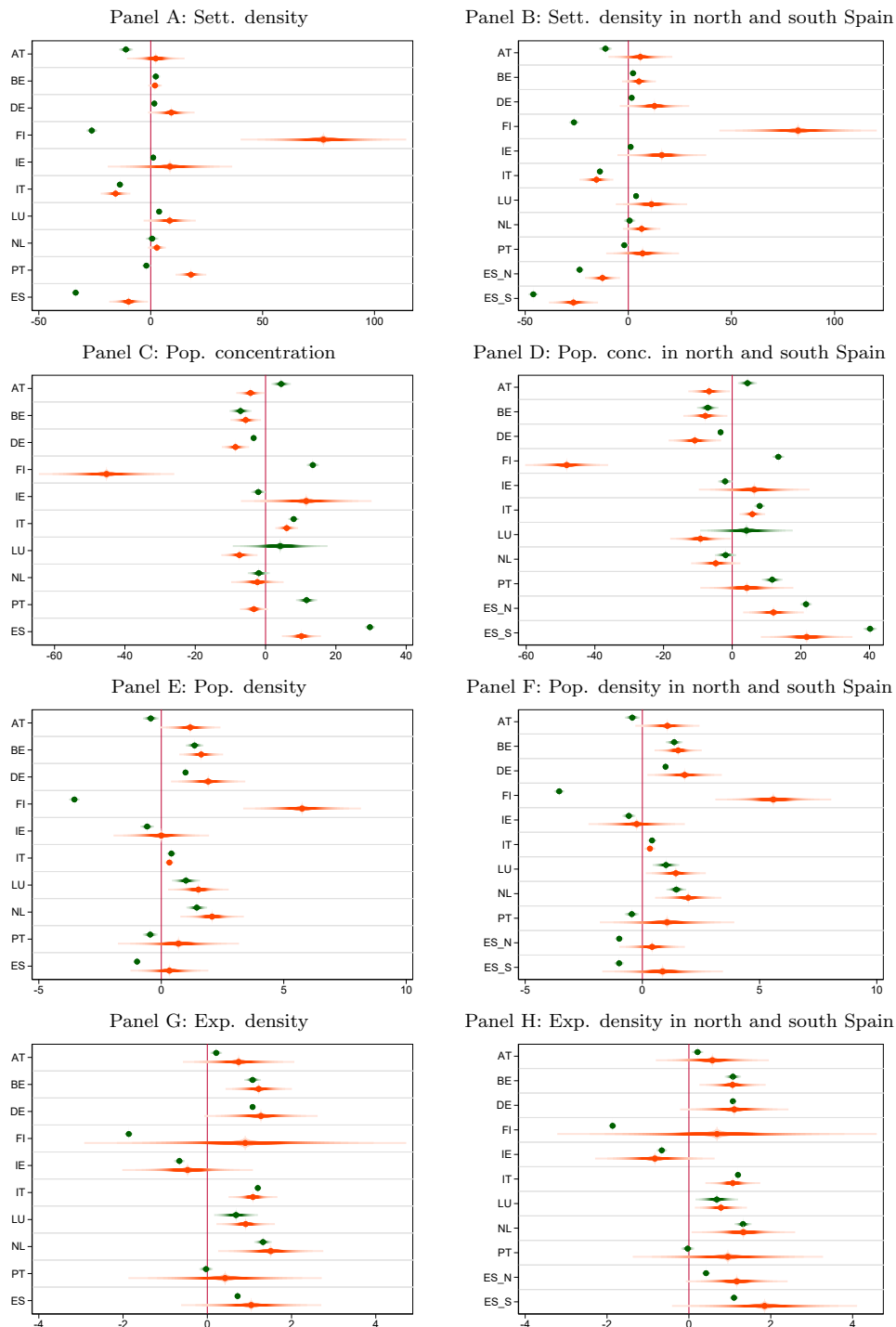
All in all, the evidence presented so far indicates that Spain has a remarkably low density of settlements and a high concentration of the population. Moreover, geographic and climatic factors fail to account for this anomaly. Concerning population levels, Spain is slightly underpopulated in relative terms when comparing its population to its total surface area, but is above the European average when looking at the actual density experienced by the population.

3.2 Discontinuities in settlement patterns at the borders

In this section, we shed further light on the Spanish settlement patterns by uncovering a discontinuity of these patterns at the Spanish borders. The purpose of this exercise is to further check that the anomalous pattern of settlement in Spain is not due to geographic or climatic factors. If there is a discontinuity in settlement density and population concentration across Spanish borders while other relevant factors only change gradually, then we will be more confident in ruling out (or at least, moderating) the explanatory power of these factors to account for the Spanish anomaly. Moreover, we investigate the existence of such discontinuities at the boundaries of the remainder countries in our dataset.

Figure 6 displays the discontinuity in our four indicators of settlement and population patterns at the borders of Spain with France and Portugal. For each observation, the x-axis represents distance to the border, while the y-axis represents the indicators' values. Grid-cells are grouped in 40 bins of 5-km length each, covering distances to the border between 0 and 200 km. Each observation represents the average value of the corresponding bin. Observations to the left of the vertical dotted

Figure 5: The Spanish anomaly in settlement and population patterns with the GEOSTAT 2006 grid.



line pertain to Spain, while those to the right belong to France and Portugal. The solid lines display the regression of a second-order polynomial at each side of the border. One can see that the discontinuity at both borders is evident in the two first rows of the figure: grid-cells corresponding to Spain showing a much lower degree of settlement density and a much higher level of population concentration. On the contrary, the last two rows of Figure 6 show that this non-linearity disappears when one analyzes population density or experienced density -except for the bottom graph of Panel B, which displays a jump in experienced density at the Spanish-Portuguese border.¹¹

In order to compare the discontinuities at the Spanish borders with respect to Europe, as well as to account for the effect of geo-climatic factors, we perform, for each border of our database, a set of regressions. Each regression is run on the grid cells lying within 200 km of the corresponding border and they take the following form:

$$y_i = \eta_c + x_i' \beta_0 + z_i' \beta_1 + \epsilon_i \quad (2)$$

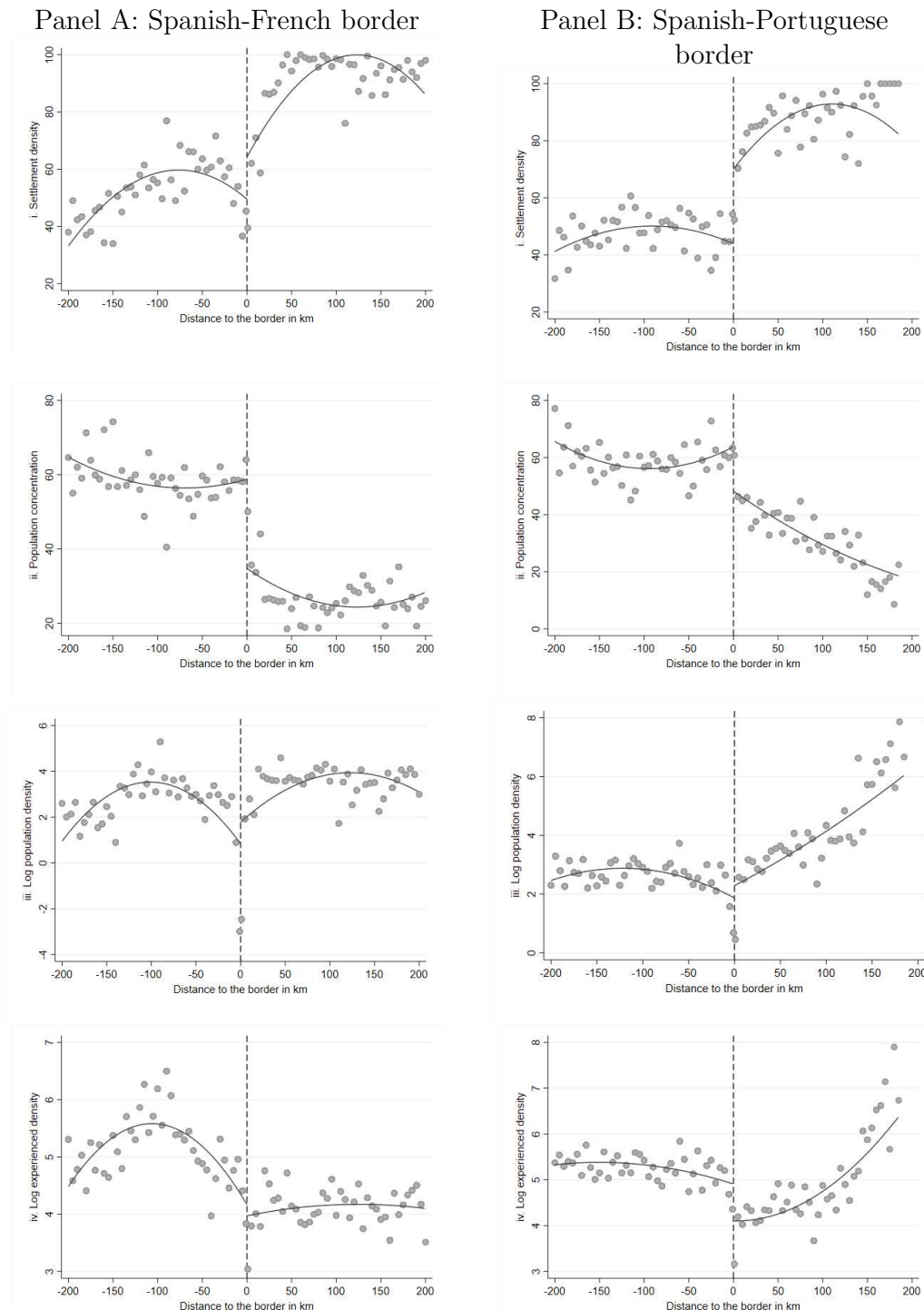
where y_i refers to one of the four indicators of settlement and population patterns in 250-km² grid cell i ; η_c is a dummy variable of one country forming the border; x_i is a vector of geographic and climatic factors, including temperature and its square, rainfall and its square, distance to the coast, surface area, and a set of border segment dummies, stretching 100 km each;¹² z_i are variable vectors describing the cell geographical position; namely, distance to the border and its square, the interaction of distance with η_c , or a quadratic polynomial in latitude and longitude. Note that there are 61 frontiers in the database. The sample is restricted to those that have at least 100 grid cells at each side of the border.

Figure 7 summarizes the results when adding the quadratic polynomial in latitude and longitude as geographical position variables. Each panel plots the coefficients of the country dummies, sorted in ascending order, along with 95% confidence intervals. The border between Spain and Portugal is depicted in green, while that between Spain and France is colored red. Panels A and B show discontinuities in settlement density and population concentration for both Spanish borders, which are consistent with Figure 6. Panels C and D show a less clear-cut pattern, with a significant discontinuity in population density only in the French border, and a large discontinuity in experienced density in the Portuguese border but lack of it in the French one. Overall, Figure 7 reveals most of the times absence of discontinuity across borders, particularly for population and experienced density. Among all these borders, the Spanish ones stand out in the extremes of the distribution of

¹¹Figures for the rest of countries of the database are available on request.

¹²If the last segment of the border is less than 50 km long, we joint it with the previous one. For this reason, some segments are longer than 100 km. Also, note that a few borders are shorter than 100 km long.

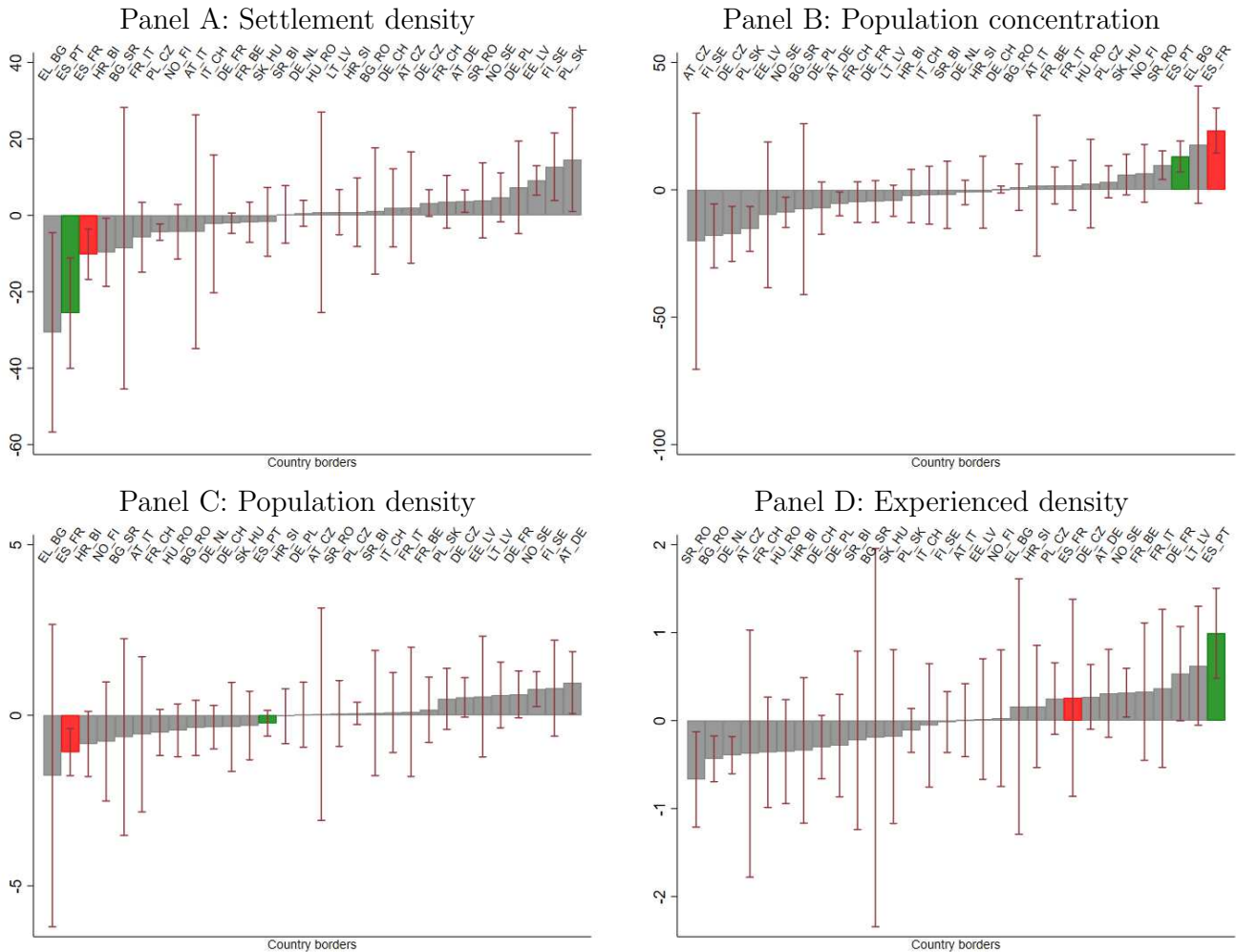
Figure 6: Spatial population patterns at the Spanish borders.



Notes: Observations to the left of the vertical dotted line pertain to Spain, while those to the right belong to France and Portugal. The solid lines display the regression of a second-order polynomial at each side of the border. Each observation is the local average of grid-cells falling within each 5 km distance-to-the-border bin.

coefficients.¹³

Figure 7: Settlement and population patterns at the European borders.



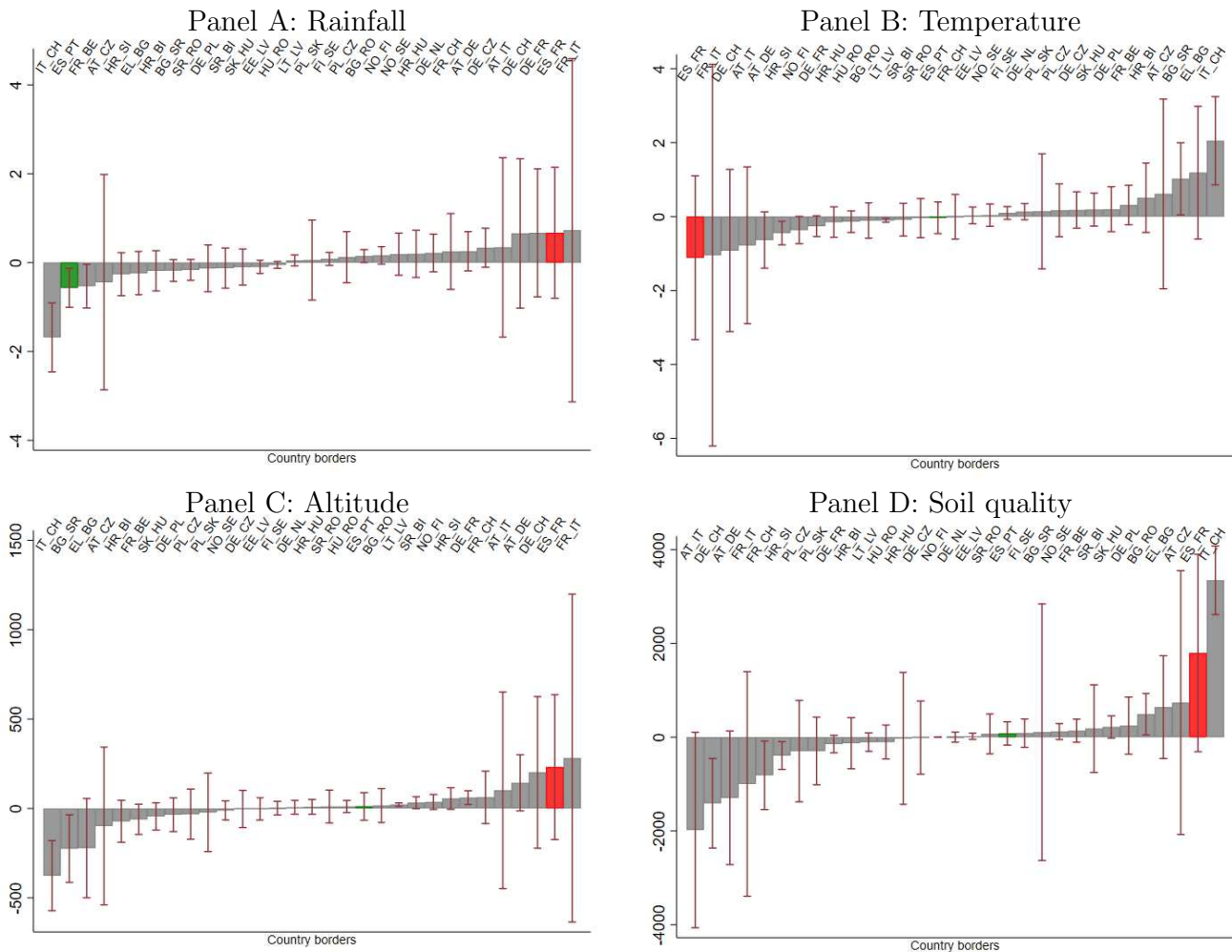
Notes: This figure uncovers settlement discontinuities at the European borders. It plots the country dummy coefficients of equation (2) for four different indicators of settlement and population patterns. In each border country-pair, the country dummy (η_c) refers to the first one. The border between Spain and Portugal is depicted in green, where that between Spain and France is colored red.

We also check whether the previous results can be driven by large discontinuities in geo-climatic factors between the borders above mentioned. To do so, we regress rainfall, temperature, altitude, and soil quality on the country dummy, segment dummies, and the geographical position variables. Figure 8 plots the country dummy coefficients in ascending order from regressions in which the geographical positions variables are again the quadratic polynomial in latitude and longitude. The

¹³Figures A4, A5 and A6 present analogous figures for the other specifications of the geographical position variables. We also corroborate that similar results are obtained when using the GEOSTAT 2006 population grid (Figures A7 and A8).

results suggest that the geo-climatic factors are very unlikely to be the drivers of the results above, since the differences at each side of the Spanish boundaries are comparable to those in other European countries and mostly statistically insignificant. There is only a small significant coefficient for rainfall, indicating that on the Spanish side of the Portuguese border it rains 56 milliliters less.¹⁴

Figure 8: Geo-climatic discontinuities at the European borders.



Notes: This figure uncovers geo-climatic discontinuities at the European borders. It plots country dummy coefficients after regressing geo-climatic factors on the country dummy, segment dummies, and geographical position variables. In each border country-pair, the country dummy refers to the first one. The border between Spain and Portugal is depicted in green, where that between Spain and France is colored red.

¹⁴We find similar results with the other specifications of the geographical position variables (Figures A9, A10 and A11).

4 History, warfare and the origins of the Spanish anomaly

4.1 Empirical evidence

Our results so far indicate that geographic and climatic factors do not satisfactorily explain Spain's population and settlement patterns, particularly in terms of settlement density and population concentration, which suggests that the peculiarities of Spanish history may be behind them, together with agglomeration forces. The Spanish anomaly is not a recent phenomenon. Already in the 17th century, European travelers were impressed by the scarcity of settlements: "One can travel for days on end without passing a house or village, and the country is abandoned and uncultivated"; "Spain gives the impression of being a desert of Libya, so unpopulated it is".¹⁵ We provide three pieces of evidence to show that the anomalous settlement pattern of Spain is historically rooted and predates the profound demographic changes (including rural exodus) that accompanied the structural transformation of the economy.

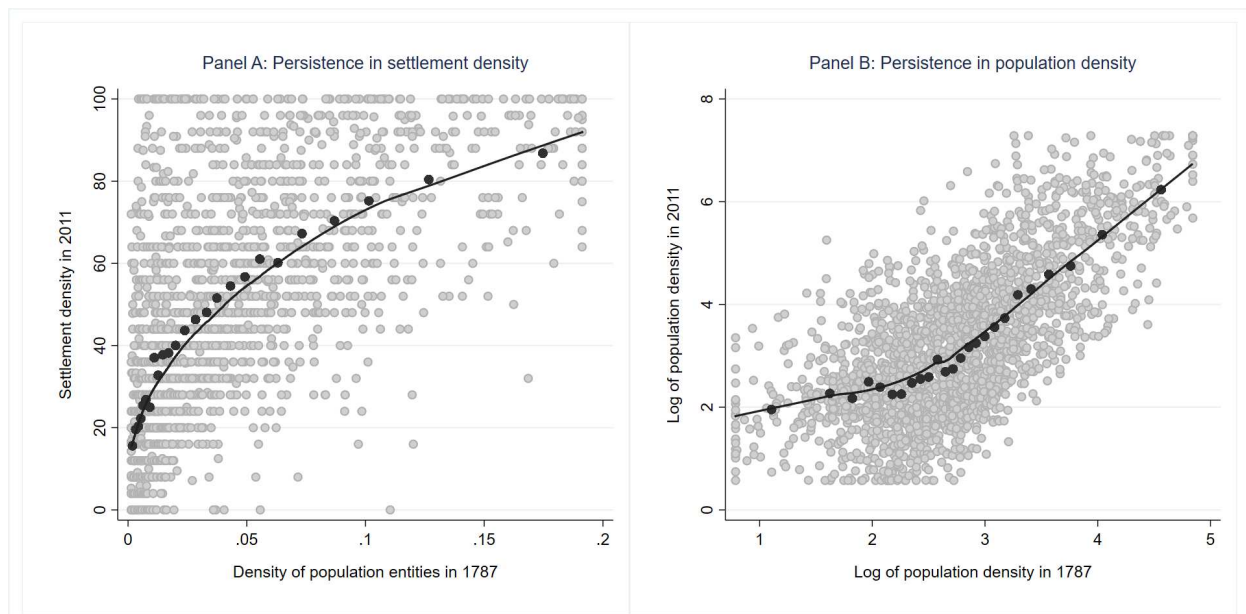
First, we focus on Spain to document a high degree of persistence in the spatial distribution of the population. The left panel in Figure 9 shows that there is a high correlation between density of localities in 1787 and settlement density in 2011 ($p = 0.65$), while the right panel depicts an also strong correlation between (log) population density in 1787 and (log) population density in 2011 ($p = 0.69$).¹⁶ This points to the fact that the main features of Spain's current spatial distribution of the population were already present at the end of the 18th century, well before Spain began its (late) industrialization process. Therefore, historical events taking place before this period are probably the responsible factors.

Second, we take a broader approach to compare Spain with other European countries in the second half of the 19th century (around 1871), when census data became available. We construct a dataset on settlements (i.e., localities) and population at the district level for France, Italy, Portugal, Prussia, and Spain. The sources used include ISTAT (1876), Estatística de Portugal (1881), Instituto Nacional de Estadística (1964), Salvati (2014), INSEE (2020), Becker and Cinnirella (2020), Alves et al. (2020), and Beltrán-Tapia et al. (2020). It is worth mentioning that the interpretation of this settlement data must be done with caution. According to the definition employed in this section, a settlement consists of a group of houses, ranging from villages to cities (we exclude isolated farmhouses as this

¹⁵From the Venetian ambassadors Federico Cornaro (1678-81) and Giovanni Cornaro (1681-82), quoted in Brenan (1950).

¹⁶To assess the magnitude of these correlations, it is important to bear in mind that the variable measuring density of population entities in 1787 is not directly comparable to settlement density in 2011. The former captures the number of population entities in 1787 divided by surface area while the latter is the percentage of populated 10-km² cells. Similarly, population density in 1787 is not exactly comparable to the indicator in 2011 as the quality of the data sources is very different. The implication is that correlations would be actually stronger if the methodology of the indicators were more similar. See Oto-Peralías (2020) for more details.

Figure 9: Persistence in settlement and population density (Spain).



Notes: Settlement density in 2011 refers to the percentage of 10-km² cells that are inhabited within each 250-km² grid cell and density of population entities in 1787 is the number of localities in 1787 per km².

information is seldom available), but the way historical censuses count settlements is likely to vary to some extent.¹⁷ Panel A of Figure 10 represents an indicator of settlement density calculated as the number of localities (i.e., villages, towns, and cities) per km². Remarkably, most of the Spanish territory, particularly its southern half, appears almost empty. Panel B also shows that population density in Spain is very low, although the north-south contrast within the country disappears.

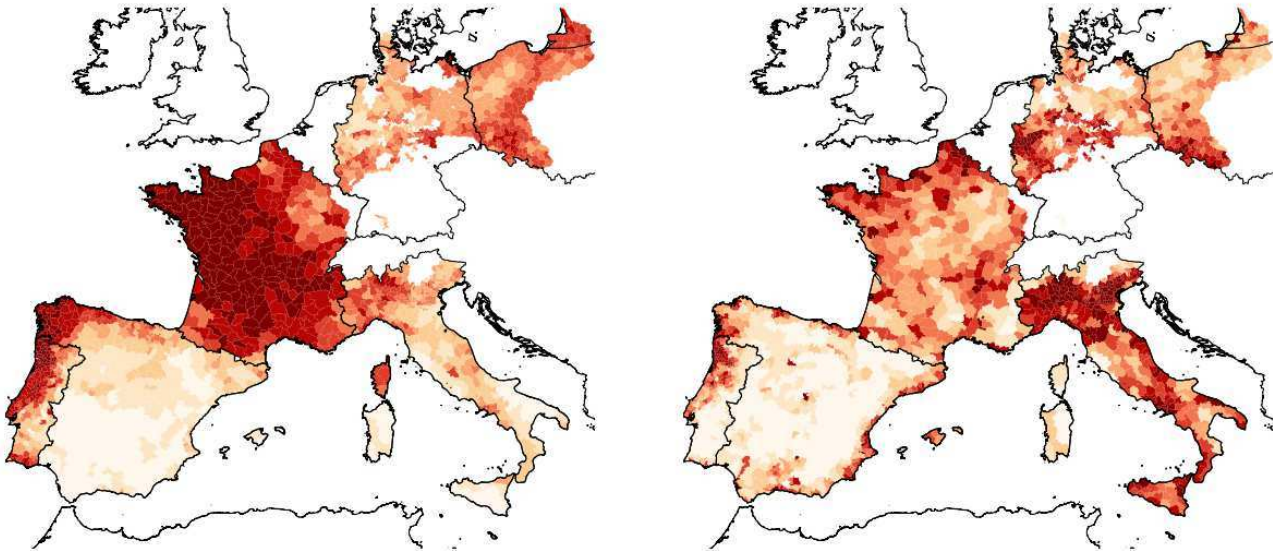
Figure 11 analyzes country differences in both indicators more systematically and controlling for geo-climatic factors. More specifically, this figure is created in an analogous way to Figure 2 but using the sample of European districts. Given the heterogeneous size of observations, we use districts' surface area as weights in the regressions. Panel A uses our baseline indicator of settlement density in 2011 for comparison purposes, while Panel B splits Spain into its northern and southern parts. We plot the country coefficients both without any control (green) and with geo-climatic controls (red). As expected, the figure shows that current settlement density is very low in Spain, particularly in the south. The next two panels (C and D) employ the indicator of settlement density during the second half of the 19th century (localities per km²). Panel C shows that all countries have lower settlement density than France and that the coefficient on Spain is the largest in absolute value. Moreover, when Spain is split into two parts in Panel D, the coefficient on southern Spain is remarkably large. Therefore, this clearly indicates that already in the 19th century Spain in general, and southern

¹⁷For instance, the distinction between an isolated farmhouse and a village/hamlet is not always clear. Appendix B in the Supplementary Material provides a detailed description of the dataset of 19th century European districts.

Figure 10: Settlement and population patterns in circa 1871.

Panel A: Settlement density (localities per km²)

Panel B: Population density



Notes: Settlement density in c. 1871 refers to the natural logarithm of the number of localities divided by surface area.

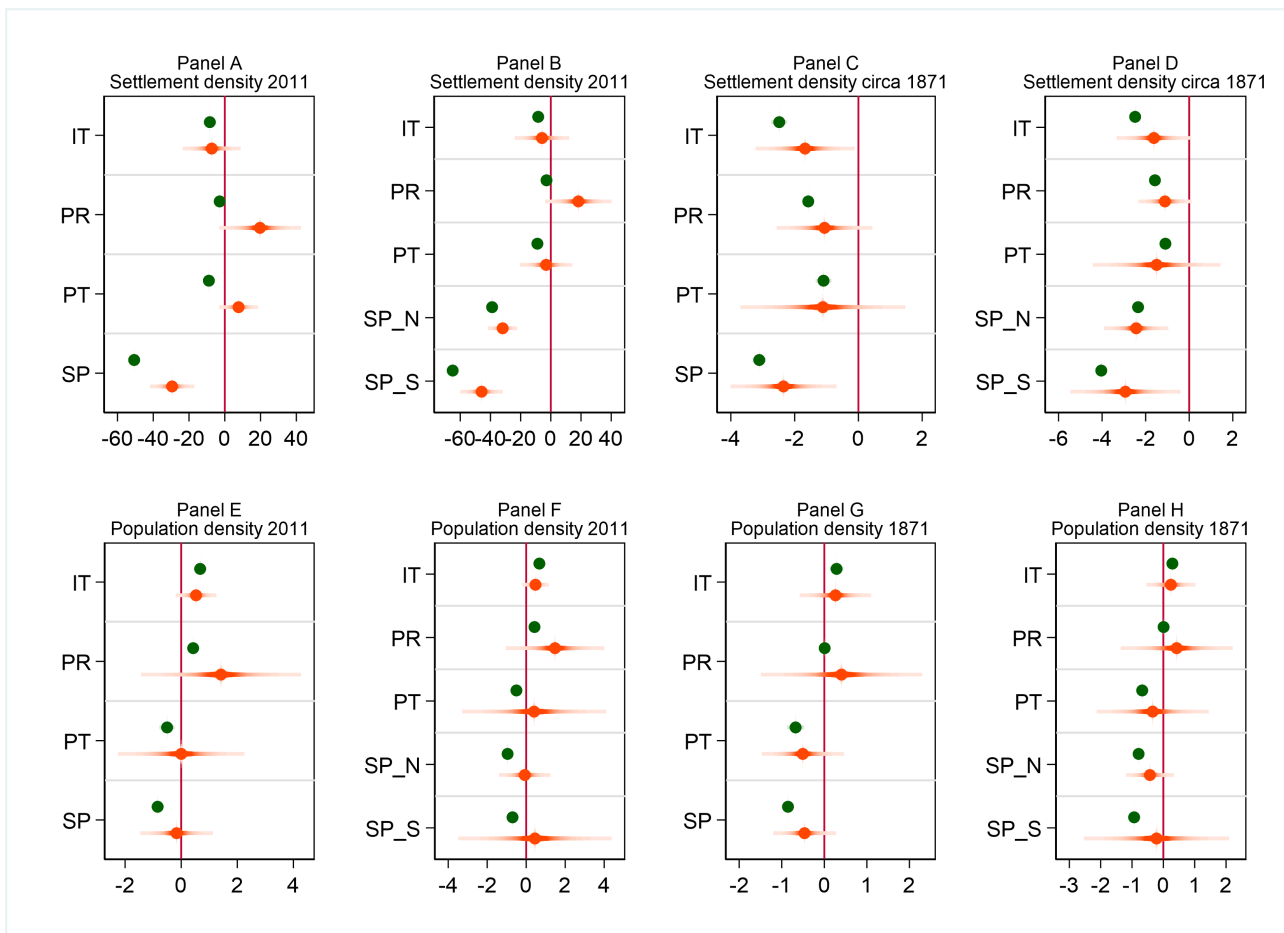
Spain in particular, had a remarkably low settlement density.¹⁸

Concerning population density, Panels E and F use the indicator measured in 2011 for comparison purposes with our baseline results in Figure 2, while Panels G and H use the indicator referred to 1871. The latter panels show that, after partialling out geographic and climatic features, the Spanish average value of population density is similar to the French one in the late 19th century, which is particularly true in southern Spain, where the difference is not statistically significant. Overall, the picture depicted in the second half of the 19th century is consistent with the outcomes observed today, namely, a very low level of settlement density in Spain, particularly in the south, while for population density the difference with respect to other countries is less pronounced.

Yet, one may wonder whether Spain has always had such a low density of settlements from remote times. If this were the case, it would suggest that some intrinsic characteristics of the country (its geography) determine the settlement structure in the very long-run, or alternatively, that its very early settlement patterns have been extraordinary persistent. In order to address this question, the third piece of evidence comes from data on ancient and pre-medieval settlements, sourced from the

¹⁸We cannot rule out that the very high settlement density of France is partially due to measurement issues. Although in Appendix B we discuss that our measure of French settlements refers to localities and that isolated farmhouses are not included, the strong contrast with its neighbors is a bit suspicious (for instance, the comparison of the island of Corsica, recently incorporated into France, with Sardinia). Therefore, differences with respect to France should be interpreted cautiously. That being said, it is beyond doubts that settlement density in Spain was well below France. In addition, the comparison among other countries is not affected by a supposedly inflated French value.

Figure 11: Settlement and population patterns in European countries in circa 1871.



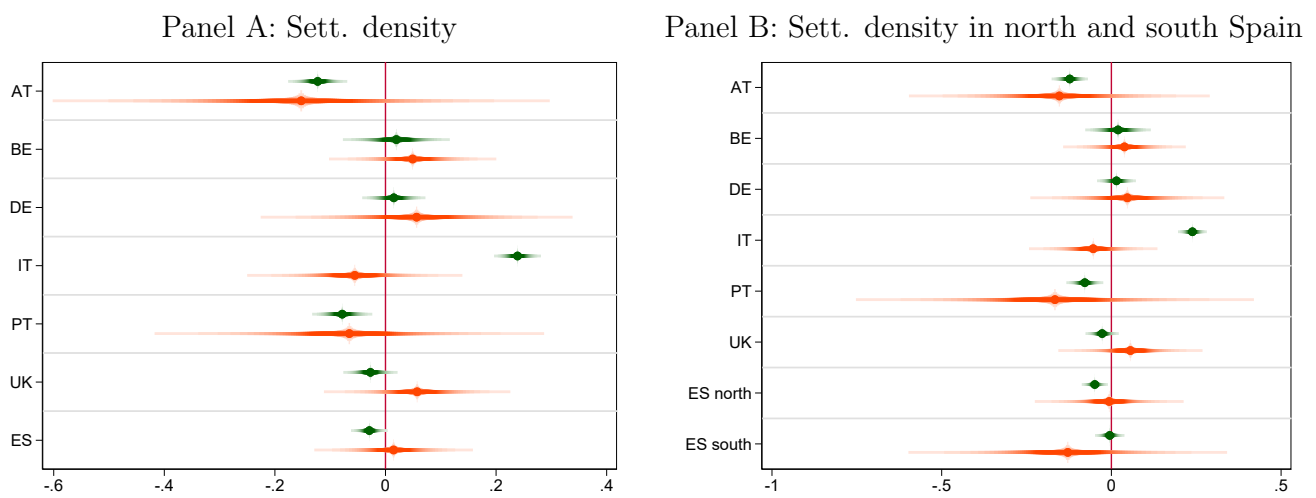
Notes: Settlement density in 2011 refers to the percentage of 10-km² cells that are inhabited within each 250-km² grid cell and settlement density in c. 1871 refers to the natural logarithm of the number of localities divided by surface area.

gazetteer of ancient places Pleiades (2021). Although a rough approximation to measure settlement patterns, given the difficulty to quantify them far back in time, this dataset provides a large collection of pre-modern attested places. It contains around 38,000 places and has good coverage of the Greek and Roman world. We construct a dummy variable capturing whether the grid cell contains at least one ancient place classified as a city, urban, town, village or settlement and 0 otherwise, as in Chronopoulos et al. (2020). The sample used comprises the area of the countries included in Figure 2 that spatially overlaps with the extent of the Roman empire c. 117 CE. Excluding areas outside the Roman Empire is convenient as a significant proportion of Pleiades' places refer to the Roman civilization. The percentage of grid cells containing ancient settlements is 27.7% for this sample.

Figure 12 shows that the percentage of cells containing settlements in Spain is similar to most of the countries. Therefore, Spain did not feature a similarly low level of settlement density in ancient and pre-medieval times as it does today. This suggests that it has not always been an outlier in

terms of scarcity of human settlements according to this data and, hence, its current anomalous settlement pattern does not seem to be a matter of geographic determinism, nor the consequence of an extraordinary long-term persistence from very remote times. Based on the results presented in this section, we should search in the historical processes happening during the middle ages and the early modern period for explanations for the Spain’s anomalous settlement pattern.

Figure 12: Ancient settlement patterns.



Notes: Settlement density refers to a dummy variable measuring the existence of at least one ancient place classified as settlement, city, urban, town or village, and attested before 700 CE.

4.2 Discussion

We turn now to discuss the origins of the Spanish anomaly. On the theoretical side, Allen and Donaldson (2018) show that path dependence is important in determining the level of economic activity across the space and their findings open up the potential for historical accidents, such as warfare, to play an outsized role in governing where economic activity occurs. In the case of Spain, previous research stresses the Reconquest as a key historical event in this regard (Oto-Peralías and Romero-Ávila (2016), Oto-Peralías (2020)). Two factors related to the Reconquest seem to be important here, namely, the pace or rate of the frontier expansion and military instability due to frontier warfare. Oto-Peralías and Romero-Ávila (2016) argue that a fast frontier expansion by the Christian kingdoms led to an imperfect colonization of the territory because the material and human resources available were insufficient relative to the magnitude of the colonization effort. The consequence was an occupation of the space characterized by a few settlements with large jurisdictional areas. Consistent with this, the authors find a positive relationship between the rate

of Reconquest and the size of municipalities' surface areas.

Oto-Peralías (2020) emphasizes frontier warfare as a key determinant of the way the territory was colonized by the Christian kingdoms. Military insecurity favored a colonization characterized by the concentration of the population in a few well defended settlements, scarcity of population and a livestock-oriented economy. Exploiting a discontinuity in military insecurity across the River Tagus during the 11th to 13th centuries, Oto-Peralías (2020) shows that the area south of the river, more exposed to frontier warfare, has today much lower levels of settlement and population density and higher population concentration. Interestingly, the difference in settlement density across the Tagus is close to the whole difference between northern and southern Spain, suggesting that the explanatory power of historical frontier warfare is important. The evidence supports the conjecture by some historians linking the prolonged exposure to frontier warfare to population concentration and scarcity of dispersed village communities (Bishko (1975)).

There are two additional explanations that could partially account for the population distribution in Spain, namely the sectoral composition of economic activity and the land planning regulation. However, both are unlikely to significantly affect the anomaly described. First, geographical conditions are more important for certain sectors, such as agriculture (see Henderson et al. (2017)), hence a high degree of industry specialization could shape the spatial distribution of population. Yet, it is worth noting that Spain's sector composition is very similar to that of other Western European countries. For example, in 2011 the agricultural share was 4.1%, while the average of the Eurozone sample was 3.7%. Second, rigidities in land-use planning can hinder the expansion of settlements and thereby increase the concentration of the population. In this regard, Riera et al. (1991) argue that the Spanish planning system bears a lot in common with other Central European countries, such as France. Moreover, according to the Doing Business Index, Spain ranks 79th in the variable assessing the easiness of carrying out construction activities ¹⁹, which is not far from the average of the Eurozone sample (51), and it is higher than the position of the Netherlands (88th) or Italy (97th). In fact, according to OECD (2017), in the context of a housing boom, Spain experienced the highest growth rate in developed land between 2000 and 2012. This evidence casts doubt on the significance of land regulation in shaping the spatial distribution of population observed in Spain.

¹⁹The variable is named Dealing with construction permits, and it gauges the procedures, time and cost to build a warehouse, including obtaining the necessary licenses and permits, submitting all required notifications, requesting and receiving all necessary inspections and obtaining utility connections.

5 Concluding remarks

Using the GEOSTAT 2011 population grid with information at the 1-km² level, this article uncovers an anomaly in the Spanish distribution of population across the space. Compared to other European countries, Spain presents the lowest density of settlements (almost 90% of its territory is uninhabited) together with the highest population concentration in certain areas. Interestingly enough, these patterns remain true even after accounting for geographical and climatic conditions.

We also find that these idiosyncratic features of the spatial distribution of the Spanish population exhibit a high degree of persistence. First, the country already presented a very low density of settlements in the second half of the 19th century. Second, there is a high correlation of both population density and settlement density between 1787 and 2011. Moreover, leveraging data from ancient sites we find that pre-medieval settlements are not abnormally low in Spain, thereby suggesting that the current anomalous settlement pattern has not always existed. Rather, the possible explanations should be found in the historical processes taking place during the Middle Ages and the Early Modern Period, as pointed out by Oto-Peralías (2020) and González-Jiménez (1992).

Economies of agglomeration rely on increasing returns to scale and lower transportation costs, and emphasize linkages between firms and suppliers as well as between firms and consumers. The emptiness of the Spanish territory may thus influence the distribution of economic activity across firms, sectors and space. The dynamic dimension related to the interaction between internal migration flows and the process of structural change of the Spanish economy over the last decades may also shape current economic outcomes.²⁰ Understanding the economic consequences of the Spanish anomaly in settlement patterns represents an exciting line of open research.

²⁰In related projects, Gutiérrez et al. (2020) analyze in detail the evolution over the last decades of the urbanization share in Spain and Budí (2020) explores the macroeconomic consequences of rural-urban migration flows.

References

- Allen, Treb and Costas Arkolakis, “Trade and the Topography of the Spatial Economy,” *The Quarterly Journal of Economics*, May 2014, 129 (3), 1085–1140.
- and Dave Donaldson, “The geography of path dependence,” July 2018. Mimeo.
- Alves, Daniel, Nuno Lima, and Pedro Machuqueiro, “Carta de Freguesias de Portugal Continental de 1878,” 2020. Luís Espinha da (ed.), Atlas, Cartografia Histórica, available at <http://atlas.fcsh.unl.pt/>.
- Becker, Sascha O and Francesco Cinnirella, “Prussia disaggregated: the demography of its universe of localities in 1871,” *Journal of Demographic Economics*, 2020, 86 (3), 259–290.
- Beltrán-Tapia, Francisco J., Alfonso Díez-Minguela, Julio Martínez-Galarraga, Alicia Gómez-Tello, and Daniel Tirado, “Spain, from the Ancien Régime to the Liberal State (ESPAREL),” 2020. Research project: <https://www.ehvalencia.es/en/espana-del-antiguo-regimen-al-estado-liberal-esparyl-en/>.
- Bishko, Charles Julian, “The Spanish and Portuguese Reconquest 1095-1492,” in Kenneth M. Setton, ed., *A History of the Crusades*, Vol. 3, Clarendon Press, Oxford, 1975, pp. 396–456.
- Brenan, Gerald, *The Spanish Labyrinth: An account of the Social and Political Background of the Spanish Civil War*, Cambridge University Press, Cambridge, 1950.
- Budí, Tomás, “Rural Depopulation, Urbanization, and Structural Change,” 2020. CEMFI Master Thesis.
- Chronopoulos, Dimitris K, Sotiris Kampanelis, Daniel Oto-Peralías, and John O S Wilson, “Ancient colonialism and the economic geography of the Mediterranean,” *Journal of Economic Geography*, 12 2020.
- Duranton, Gilles and Diego Puga, “The Economics of Urban Density,” *Journal of Economic Perspectives*, August 2020, 34 (3), 3–26.
- and Matthew A. Turner, “Urban form and driving: Evidence from US cities,” *Journal of Urban Economics*, 2018, 108, 170–191.
- EFGS, “GEOSTAT 1A. Representing Census data in a European population grid,” 2011. <http://www.efgs.info/wp-content/uploads/geostat/1a/GEOSTAT1A-final-report.pdf>.

- Estatística de Portugal**, “II Recenseamento Geral da População, 1878 - 1 de Janeiro,” 1881. Available at <https://censos.ine.pt/>.
- Eurostat**, “GEOSTAT 2006 population grid,” 2006. <http://ec.europa.eu/eurostat/>.
- , “GEOSTAT 2011 population grid,” 2016. <http://ec.europa.eu/eurostat/>.
- and **EFGS**, “GEOSTAT V2. Representing Census data in a European population grid,” 2015.
- Goerlich, Francisco and Isidro Cantarino**, “A population density grid for Spain,” *International Journal of Geographical Information Science*, 2013, 27 (12), 2247–2263.
- and —, “Grid Poblacional 2011 para España,” *Estudios Geográficos*, 2017, 78 (282), 135–163.
- González-Jiménez, Manuel**, “Frontier and Settlement in the Kingdom of Castile (1085-1350),” in Robert Bartlett and Angus MacKay, eds., *Medieval Frontier Societies*, Clarendon Press, Oxford, 1992, pp. 49–74.
- Gutiérrez, E., E. Moral-Benito, and R. Ramos**, “Tendencias recientes de la población en áreas rurales y urbanas en España,” 2020. Banco de España Documento Ocasional 2027.
- Harari, Mariaflavia**, “Cities in Bad Shape: Urban Geometry in India,” *American Economic Review*, August 2020, 110 (8), 2377–2421.
- Henderson, J Vernon, Tim Squires, Adam Storeygard, and David Weil**, “The Global Distribution of Economic Activity: Nature, History, and the Role of Trade,” *The Quarterly Journal of Economics*, 09 2017, 133 (1), 357–406.
- INSEE**, “Données historiques de la Statistique générale de France SGF 1800-1925,” 2020. Available at www.insee.fr.
- Instituto Nacional de Estatística**, “X Recenseamento Geral de População no continente e ilhas adjacentes às 0 horas de 15 de Dezembro de 1960 são editados em VI Tomos e um Anexo.” 1964. Available at <https://censos.ine.pt/>.
- ISTAT**, “Popolazione Presente ed Assente per Comuni, Centri e Frazioni di Comune. Censimento 31 dicembre 1871,” 1876.
- Krugman, Paul**, “Increasing returns and economic geography,” *Journal of Political Economy*, 1991, 99 (3), 483–499.
- OECD**, *The Governance of Land Use in OECD Countries* 2017.

- Oto-Peralías, Daniel**, “Frontiers, warfare and economic geography: The case of Spain,” *Journal of Development Economics*, 2020, *146*, 102511.
- **and Diego Romero-Ávila**, “The economic consequences of the Spanish Reconquest: the long-term effects of Medieval conquest and colonization,” *Journal of Economic Growth*, 2016, *21*, 409–464.
- Pleiades**, “Pleiades: a gazetteer of past places,” 2021. Available online at: <http://pleiades.stoa.org/> (accessed 30 March 2021).
- Rae, Alasdair**, “Think your country is crowded? These maps reveal the truth about population density across Europe,” 2018. The Conversation.
- Riera, Pere, Ian Munt, and John Keyes**, “The practice of land use planning in Spain,” *Planning Practice & Research*, 1991, *6* (2), 11–18.
- Roca, Jorge De La and Diego Puga**, “Learning by working in big cities,” *The Review of Economic Studies*, 2017, *84* (1), 106–142.
- Salvati, Luca**, “Population distribution and urban growth in Southern Italy, 1871–2011: emergent polycentrism or path-dependent monocentricity?,” *Urban geography*, 2014, *35* (3), 440–453.

Supplementary Material for

The spatial distribution of population in Spain: An anomaly in European perspective

E. Gutiérrez, E. Moral-Benito, D. Oto-Peralías, R. Ramos

Index

Appendix A: Additional tables and figures

Appendix B: Construction of the dataset on settlement and
population patterns in Europe in the second half of the
19th century

Appendix A Additional figures and tables

Table A1: Description of variables.

Variable	Description	Sources
Altitude	Average altitude of the surface area of the grid cell.	Authors' elaboration using data from GTOPO30 (Data available from the U.S. Geological Survey).
Border segment dummies	Dummy variables capturing the closest border segment to the grid cell centroid. Each border segment stretches 100 kms.	Authors' elaboration.
Distance to the coast	Geodesic distance between the centroid of the grid cell and the nearest point of the coast (in km).	Authors' elaboration.
Latitude/ Longitude	The geographic coordinates of the grid cell centroids, in decimal degrees.	Authors' elaboration.
Log of experienced density	Natural logarithm of experienced density, defined as the number of inhabitants divided by the occupied surface area in squared kilometers. Years of measurement: 2006 and 2011.	Authors' elaboration using data from GEOSTAT 2006 and GEOSTAT 2011.
Log of population density	Natural logarithm of 0.001 plus population density, where the latter is the number of inhabitants divided by the surface area in squared kilometers. Years of measurement: 2006 and 2011.	Authors' elaboration using data from GEOSTAT 2006 and GEOSTAT 2011.
Log of population density in 1787 (Spain)	Natural logarithm of population density in 1787.	Authors' elaboration using the 1787 population census (Instituto Nacional de Estadística (1987)).
Log of population density in c. 1871	Natural logarithm of population density in 1871 in Italy and Prussia, in 1878 in Portugal, in 1887 in Spain and in 1891 in France.	Authors' elaboration using several sources (see Appendix B).
Log of settlement density in c. 1871	Natural logarithm of the number of localities in c. 1871 divided by surface area.	Authors' elaboration using several sources (see Appendix B).

Table A1: Description of variables (continued).

Variable	Description	Sources
Population concentration	Percentage of the population living in the most populated one percent of the territory in 2006 and 2011.	Authors' elaboration using data from GEOSTAT 2006 and GEOSTAT 2011.
Precipitation	Annual precipitation, in hundred of milliliters. It corresponds to the average value of the surface area of the grid cell.	Authors' elaboration using data from WorldClim (Hijmans et al. (2005)).
Pre-medieval settlements	Dummy variable measuring the existence of at least one ancient place classified as settlement, city, urban, town or village, and attested before 700 CE.	Authors' elaboration using data from Pleiades (2021).
Ruggedness	Standard deviation of the altitude of the territory corresponding to the grid cell.	Authors' elaboration using data from GTOPO30 (Data available from the U.S. Geological Survey.).
Settlement density	Percentage of 10-km ² cells that are inhabited within each 250-km ² grid cell. Years of measurement: 2006 and 2011.	Authors' elaboration using data from GEOSTAT 2006 and GEOSTAT 2011.
Settlement density in 1787 (Spain)	Number of settlements (i.e., localities) in 1787 per km ² .	Authors' elaboration using the 1787 population census (Instituto Nacional de Estadística (1987)).
Soil quality	The indicator is the Caloric Suitability Index, which captures the potential agricultural output (maximum potential caloric yield attainable) based on crops that were available for cultivation before 1500 CE.	Authors' elaboration using data from Galor and Özak (2016).
Temperature	Annual average temperature. It corresponds to the average value of the surface area of the grid cell.	Authors' elaboration using data from WorldClim (Hijmans et al. (2005)).

Table A2: Descriptive statistics at the grid-cell level.

	(1) N	(2) mean	(3) sd	(4) min	(5) max
Altitude	23,965	370.7	403.7	-3.746	3,552
Distance to the coast	24,276	134.3	138.2	0.00126	660.1
Latitude	24,276	50.84	8.831	34.62	71.17
Longitude	24,276	11.62	11.45	-24.47	34.51
Log of experienced density	22,414	4.344	1.517	0	9.697
Log of experienced density with the GEOSTAT 2006 grid	21,470	4.378	1.487	0	9.933
Log of population density	24,276	2.471	3.313	-6.908	18.92
Log of population density with the GEOSTAT 2006 grid	24,276	2.115	3.870	-6.908	16.99
Log of population density in 1787 (Spain)	2,141	2.75	0.77	0.79	4.84
Log of population density in c. 1871	1,718	4.05	0.85	1.59	9.50
Log of settlement density c. 1871	1,722	-1.99	1.46	-5.24	1.19
Population concentration	22,638	37.99	23.51	2.688	100
Population concentration with the GEOSTAT 2006 grid	19,829	37.66	23.11	2.997	100
Precipitation	23,987	7.734	2.928	2.263	27.81
Pre-medieval settlements	7,703	0.277	0.447	0	1
Ruggedness	23,965	87.46	102.1	0	768.9
Settlement density	23,498	72.29	33.12	0	100
Settlement density with the 2006 GEOSTAT grid	20,875	77.37	31.53	0	100
Settlement density in 1787 (Spain)	2,142	0.04	0.04	0.001	0.19
Soil quality	22,037	6,759	3,408	0	10,941
Temperature	23,987	8.148	4.759	-9.589	19.58

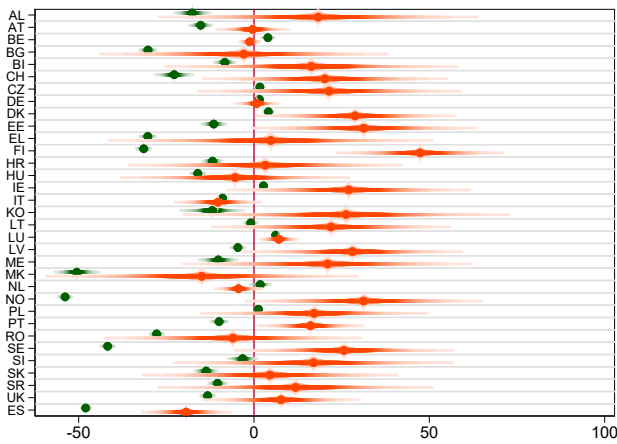
Table A3: The role of geography and climate in population patterns.

	(1) Sett. density	(2) Pop. concentration	(3) Pop. density	(4) Exp. density
Temperature_pt2	13.16** (5.123)	-8.898** (3.114)	0.892*** (0.248)	0.360*** (0.100)
Temperature_pt3	17.55** (5.654)	-11.81*** (2.514)	1.256*** (0.205)	0.565*** (0.148)
Temperature_pt4	16.49** (5.524)	-9.539*** (2.353)	1.602*** (0.283)	1.048*** (0.181)
Rainfall_pt2	3.468** (1.398)	-3.283 (2.490)	0.199 (0.269)	0.192 (0.138)
Rainfall_pt3	7.892*** (2.079)	-8.677*** (2.486)	0.341 (0.266)	0.0581 (0.190)
Rainfall_pt4	7.765** (3.373)	-11.15*** (2.575)	0.276 (0.266)	0.0552 (0.175)
Elevation_pt2	-0.999 (1.473)	2.437 (1.894)	-0.422** (0.154)	-0.343*** (0.0916)
Elevation_pt3	-8.616*** (2.182)	8.724** (2.849)	-1.334*** (0.190)	-0.701*** (0.103)
Elevation_pt4	-14.25*** (2.314)	12.21*** (2.627)	-1.825*** (0.207)	-0.948*** (0.0833)
Ruggedness_pt2	4.142** (1.369)	-3.421** (1.127)	0.343*** (0.0519)	0.148** (0.0532)
Ruggedness_pt3	5.448*** (0.904)	-4.325** (1.467)	0.452*** (0.140)	0.156 (0.117)
Ruggedness_pt4	4.137 (2.660)	-4.467** (1.649)	0.558*** (0.137)	0.219 (0.159)
Soil_quality_pt2	8.327** (2.965)	-3.371 (2.044)	0.467*** (0.122)	0.131* (0.0716)
Soil_quality_pt3	12.88*** (3.524)	-7.984*** (1.851)	0.501*** (0.159)	-0.0454 (0.103)
Soil_quality_pt4	12.14*** (3.438)	-6.832*** (1.880)	0.294 (0.175)	-0.161* (0.0863)
Dist_coast_pt2	-3.200 (2.895)	4.069* (1.970)	-0.356* (0.180)	-0.172** (0.0589)
Dist_coast_pt3	-10.01 (7.450)	6.681 (5.152)	-0.479 (0.361)	-0.0210 (0.149)
Dist_coast_pt4	-8.547 (6.523)	4.326 (4.219)	-0.0547 (0.326)	0.314 (0.183)
Island	-4.653 (4.109)	2.656 (1.828)	-0.794*** (0.224)	-0.637*** (0.0785)
Latitude	23.56*** (3.051)	-15.51*** (2.654)	1.369*** (0.175)	0.131 (0.149)
Longitude	2.261 (1.651)	0.867 (1.252)	0.149 (0.146)	0.0451 (0.0957)
Latitude ²	-0.250*** (0.0309)	0.165*** (0.0275)	-0.0153*** (0.00168)	-0.00186 (0.00120)
Longitude ²	-0.0128 (0.0201)	0.0126 (0.0143)	-0.00524*** (0.00124)	-0.00505*** (0.000644)
Latitude_longitude	-0.0316 (0.0408)	-0.0242 (0.0285)	-0.00111 (0.00303)	0.000827 (0.00168)
Observations	11,154	10,972	11,177	10,963
R-squared	0.682	0.552	0.570	0.544

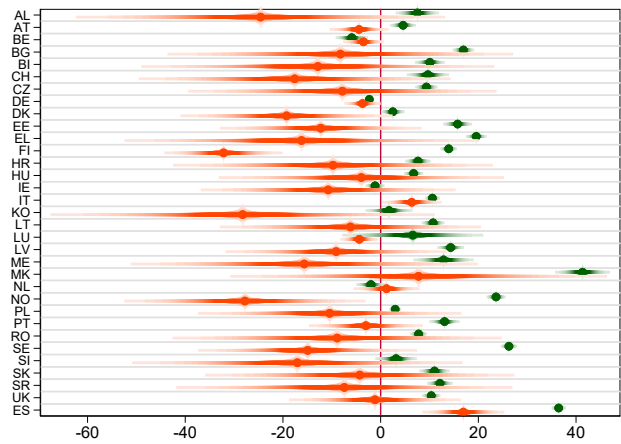
Notes: Standard errors clustered at the country level. *** p<0.01, ** p<0.05, * p<0.1

Figure A1: The Spanish anomaly in population patterns (Europe).

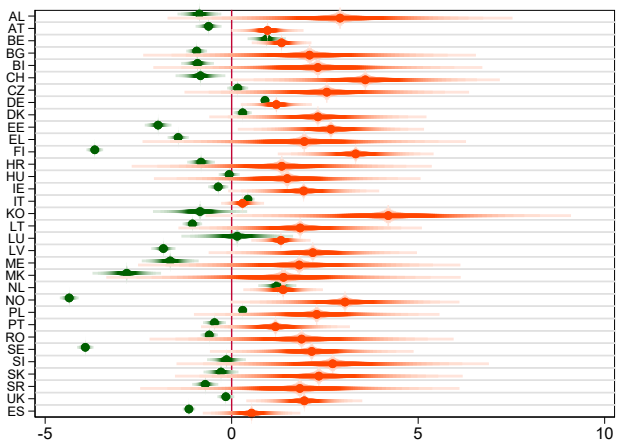
Panel A: Settlement density



Panel B: Population concentration



Panel C: Population density



Panel D: Experienced density

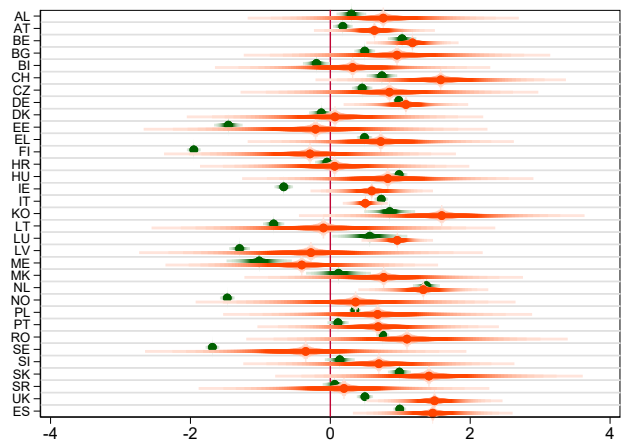


Figure A2: Cumulative distribution of the difference between the coefficient on Spain and on each virtual country from 1,000 regressions.

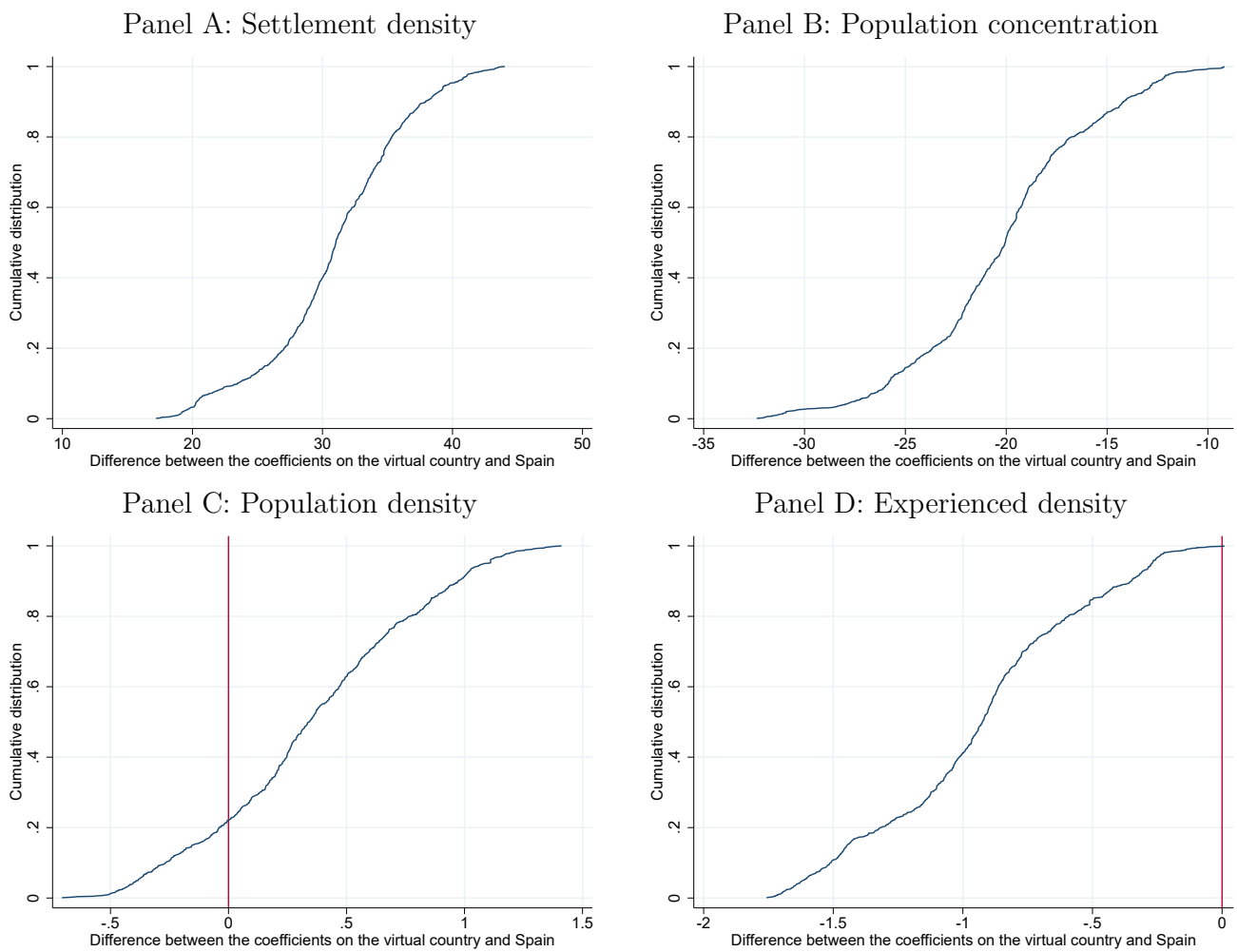


Figure A3: The Spanish anomaly by European regions.

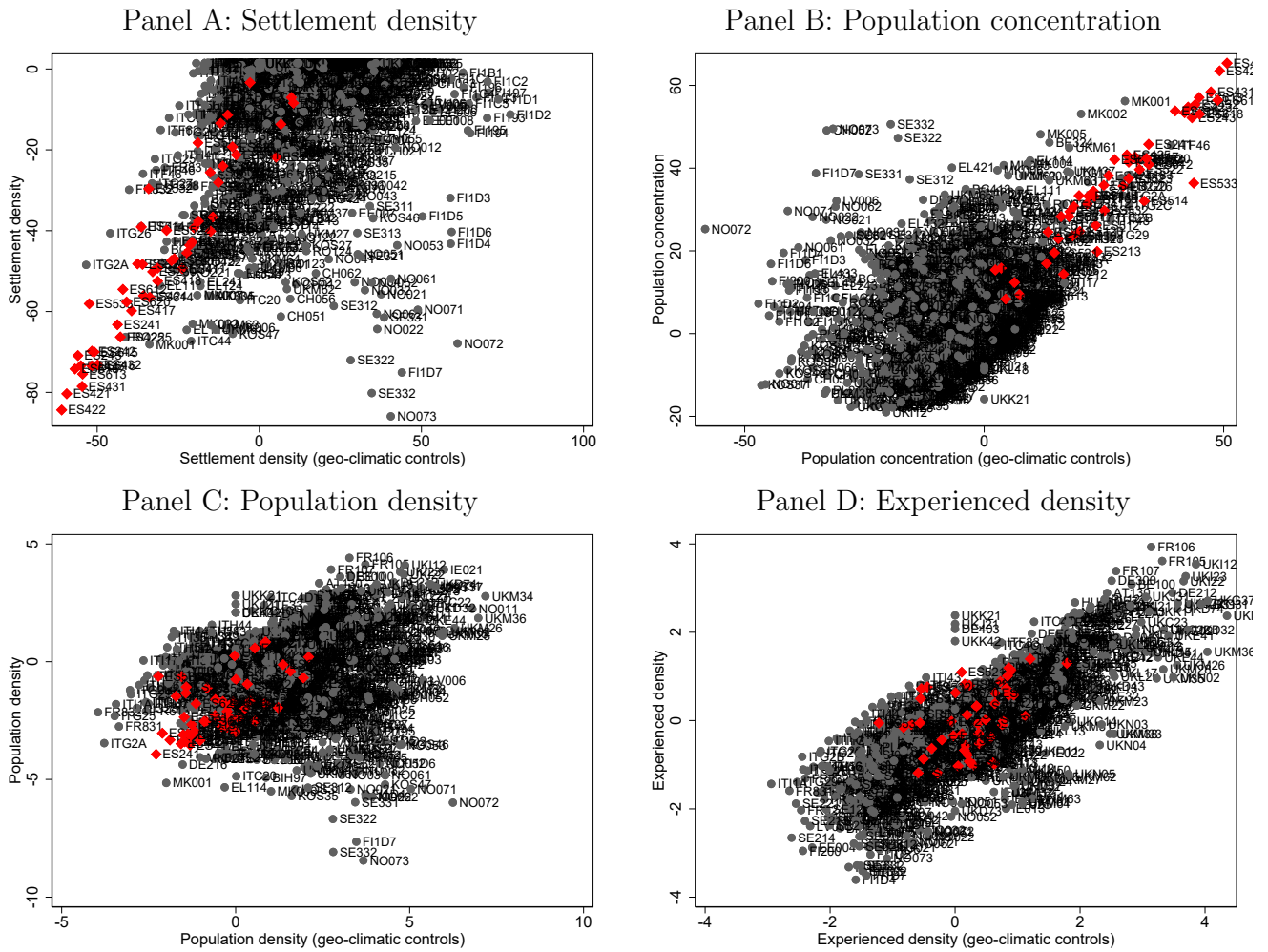
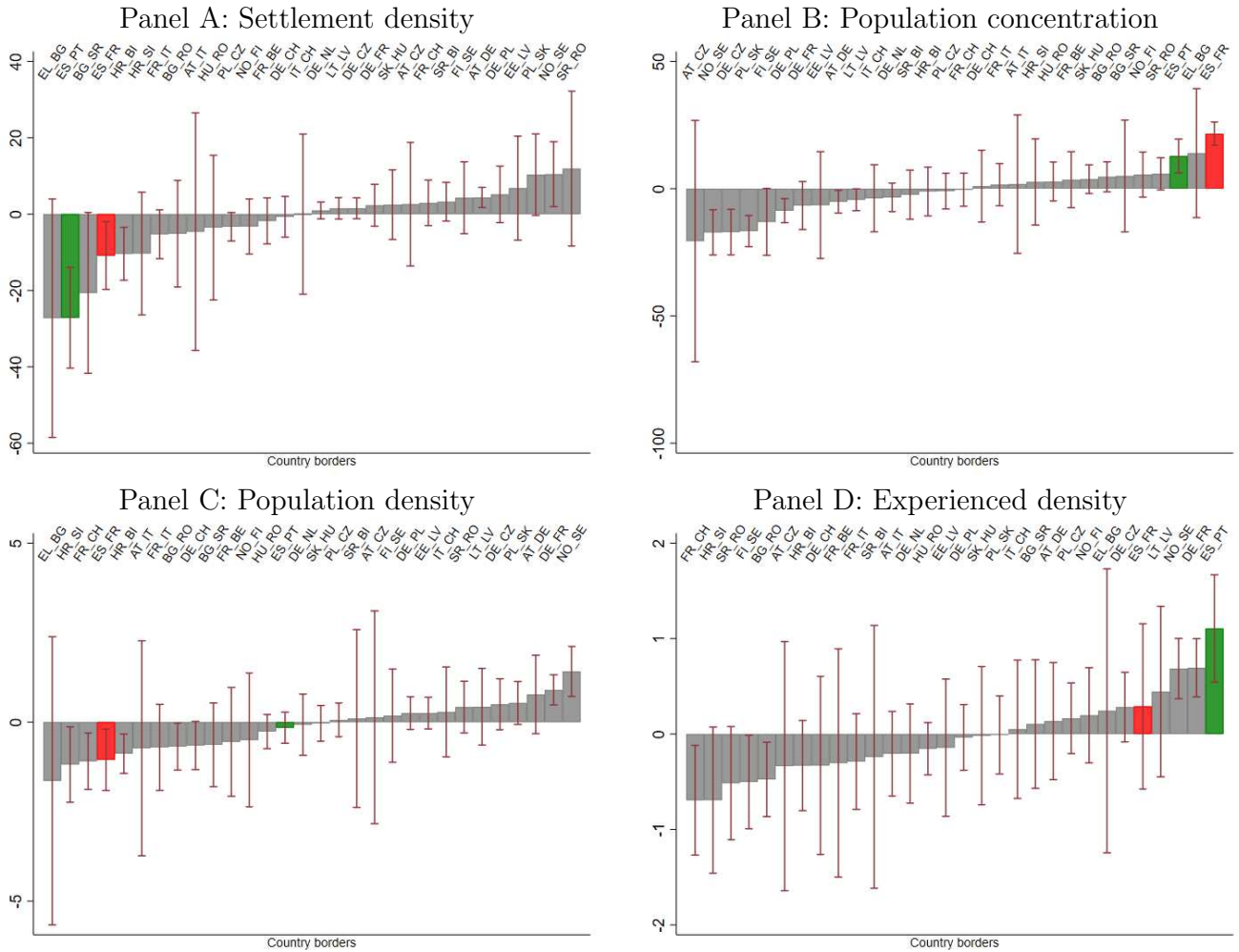
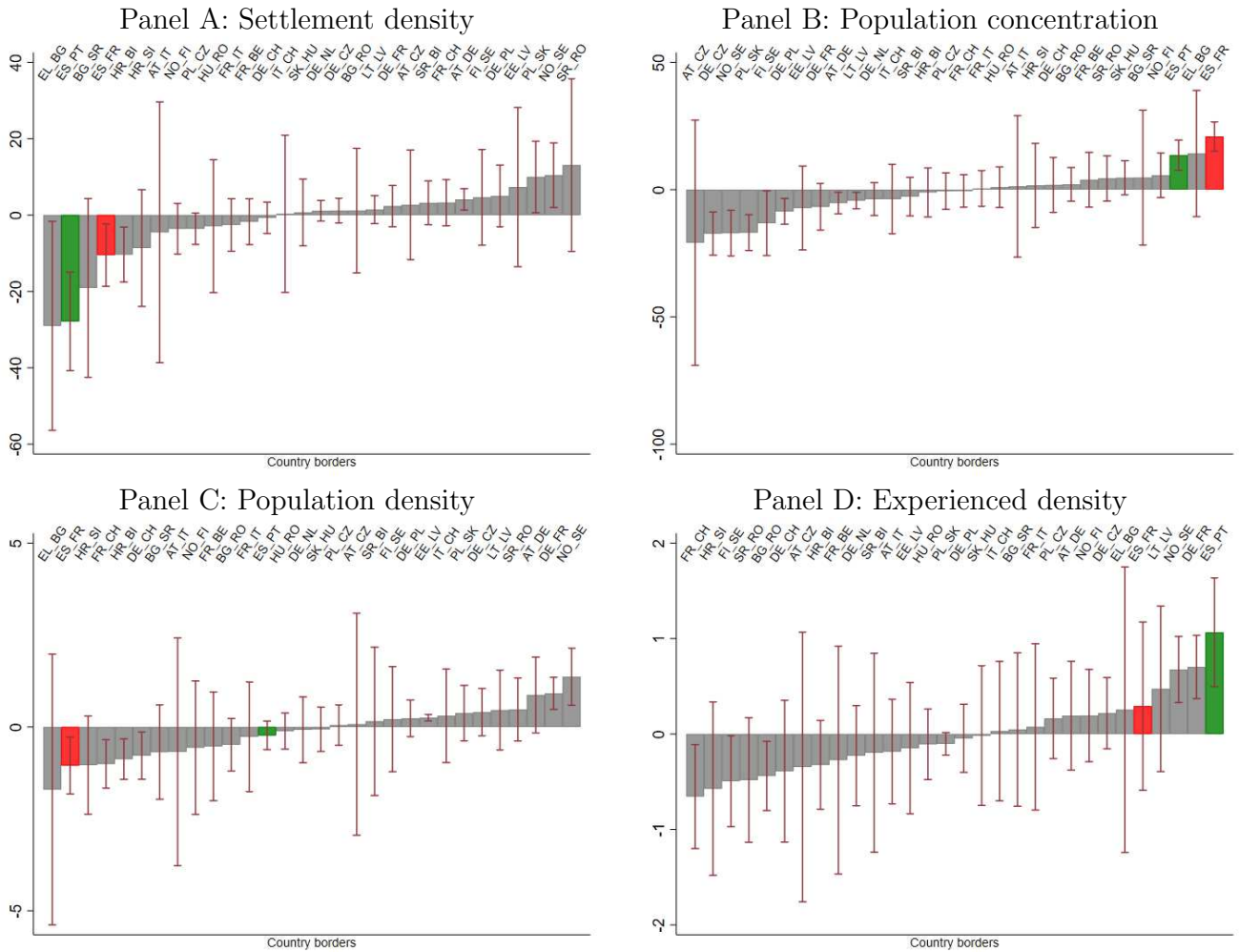


Figure A4: Settlement patterns at the European borders excluding geographical position variables.



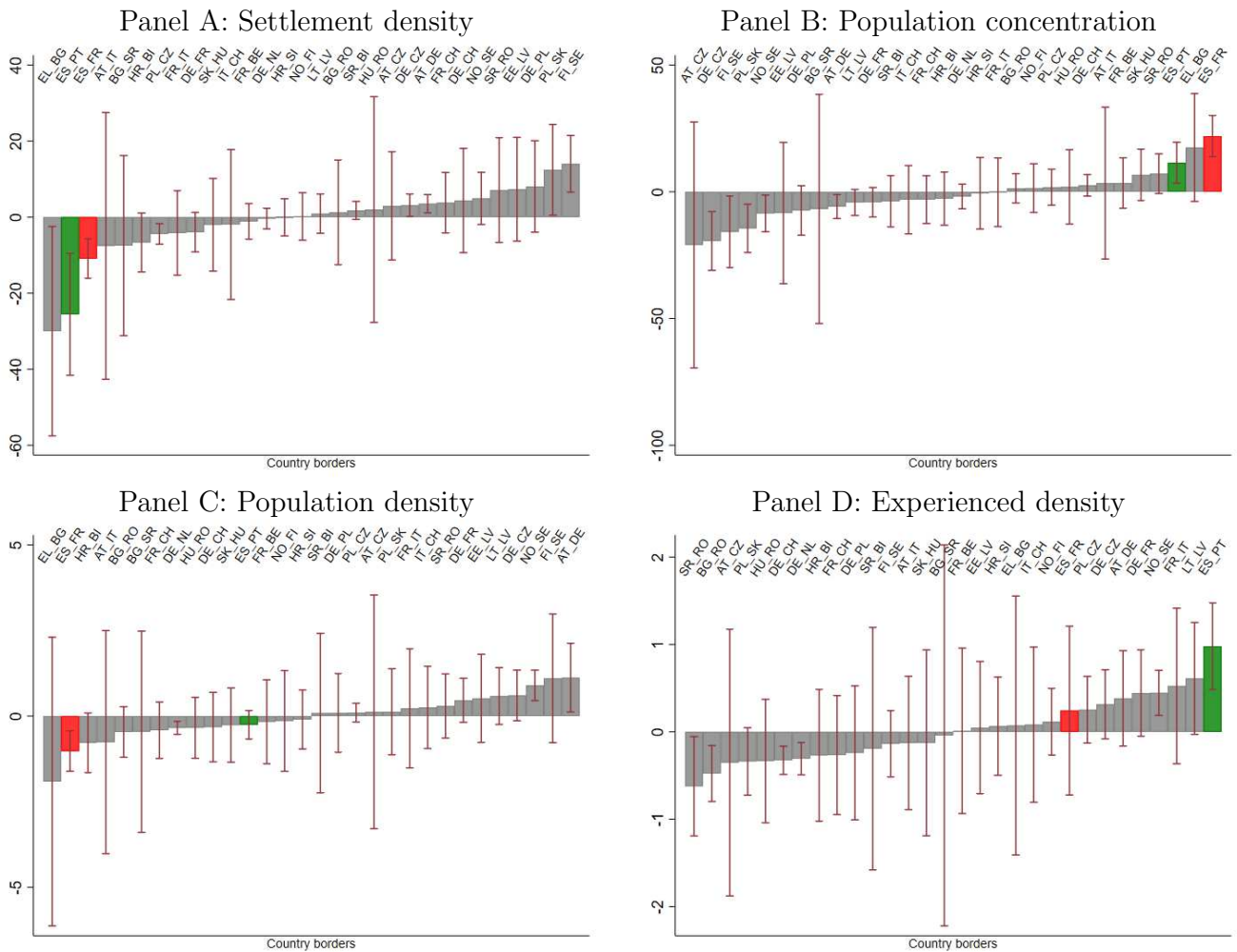
Notes: This figure uncovers settlement discontinuities at the European borders. It plots the country dummy coefficients of equation (2) in the main text for four different indicators of settlement and population patterns. In each border country-pair, the country dummy (η_c) refers to the first one. The border between Spain and Portugal is depicted in green, where that between Spain and France is colored red.

Figure A5: Settlement patterns at the European borders including distance and its square.



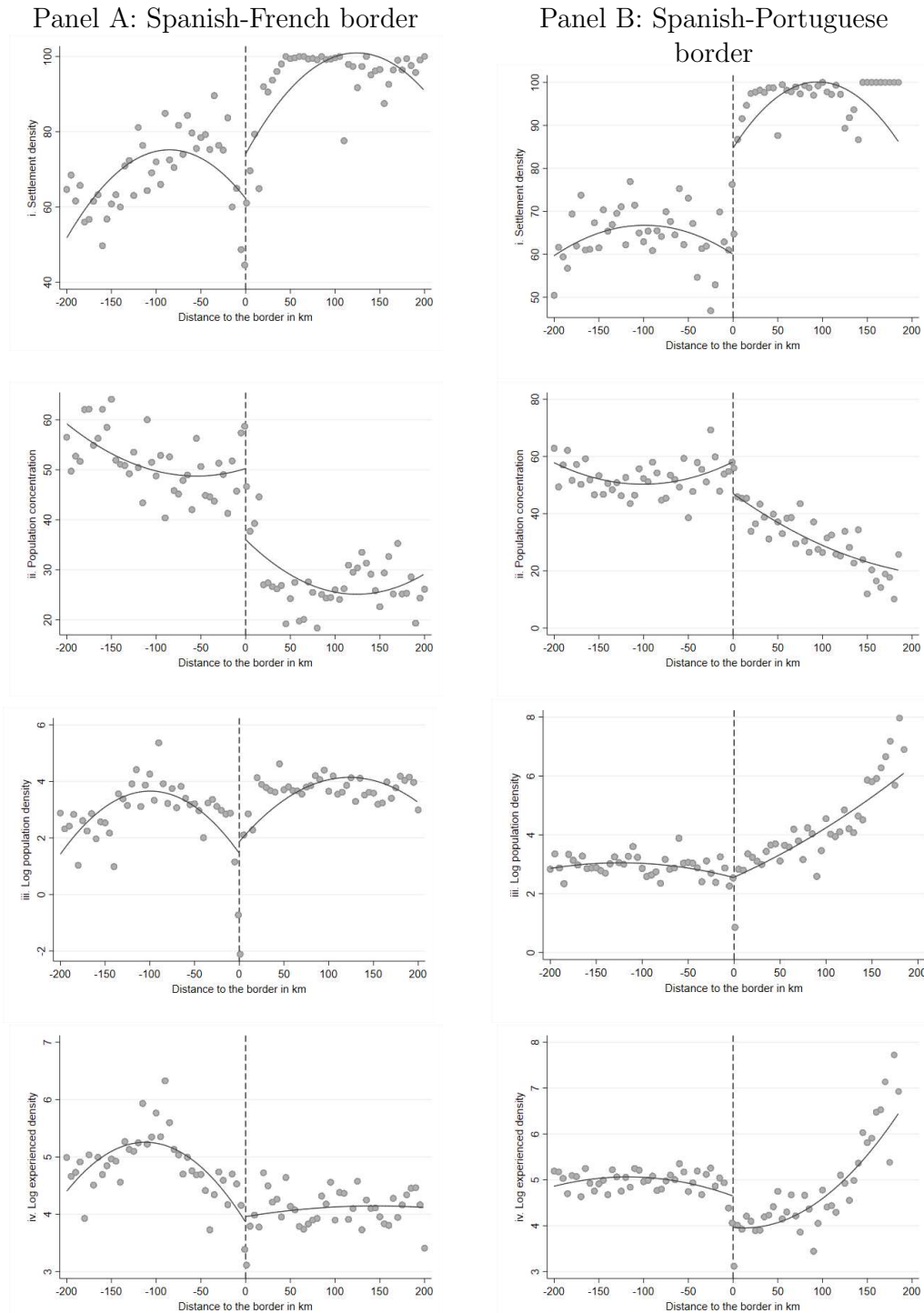
Notes: This figure uncovers settlement discontinuities at the European borders. It plots the country dummy coefficients of equation (2) for four different indicators of settlement and population patterns. In each border country-pair, the country dummy (η_c) refers to the first one. The border between Spain and Portugal is depicted in green, where that between Spain and France is colored red.

Figure A6: Settlement patterns at the European borders including distance and interaction with country dummy.



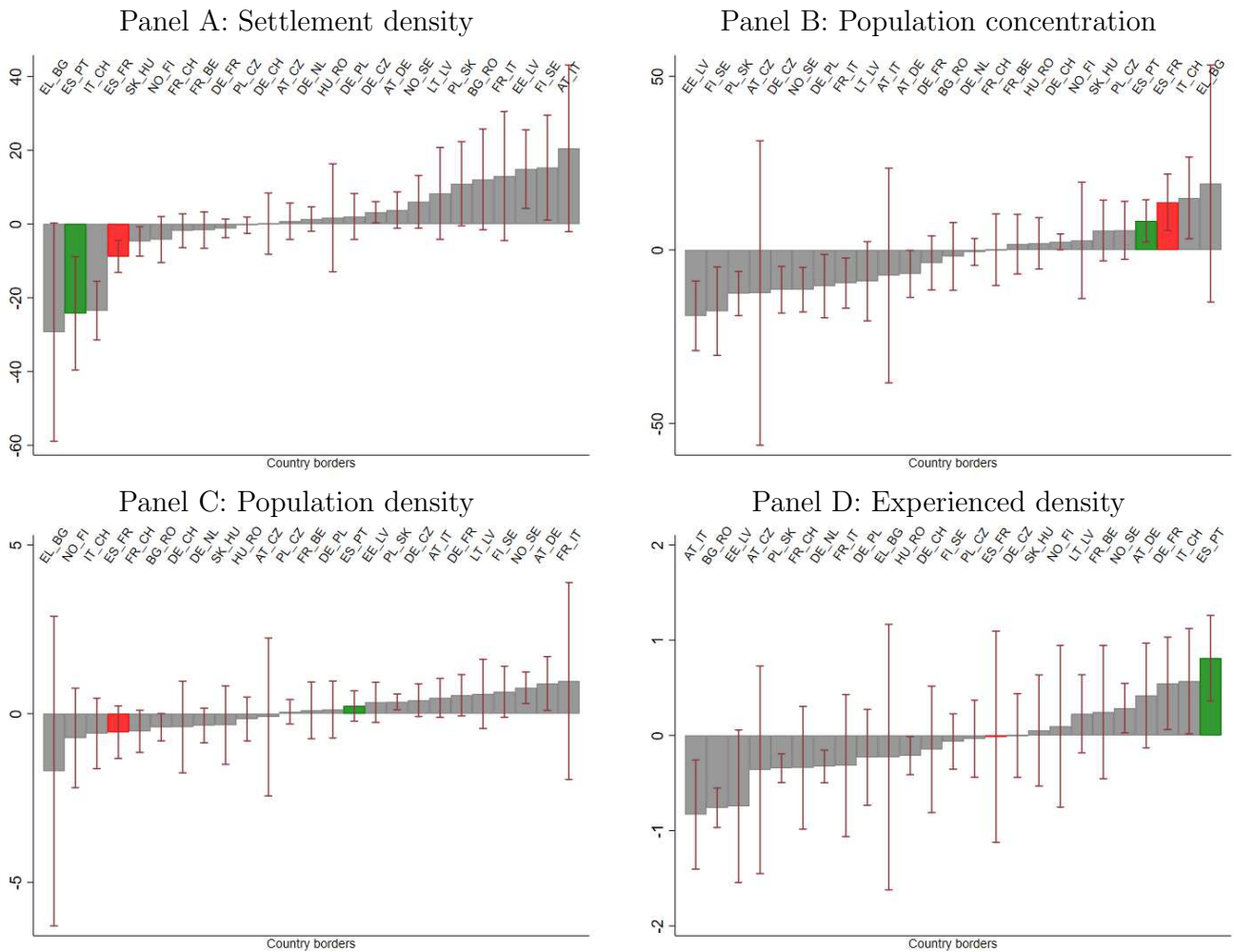
Notes: This figure uncovers settlement discontinuities at the European borders. It plots the country dummy coefficients of equation (2) for four different indicators of settlement and population patterns. In each border country-pair, the country dummy (η_c) refers to the first one. The border between Spain and Portugal is depicted in green, where that between Spain and France is colored red.

Figure A7: Spatial population patterns at the Spanish borders with the GEOSTAT 2006 grid.



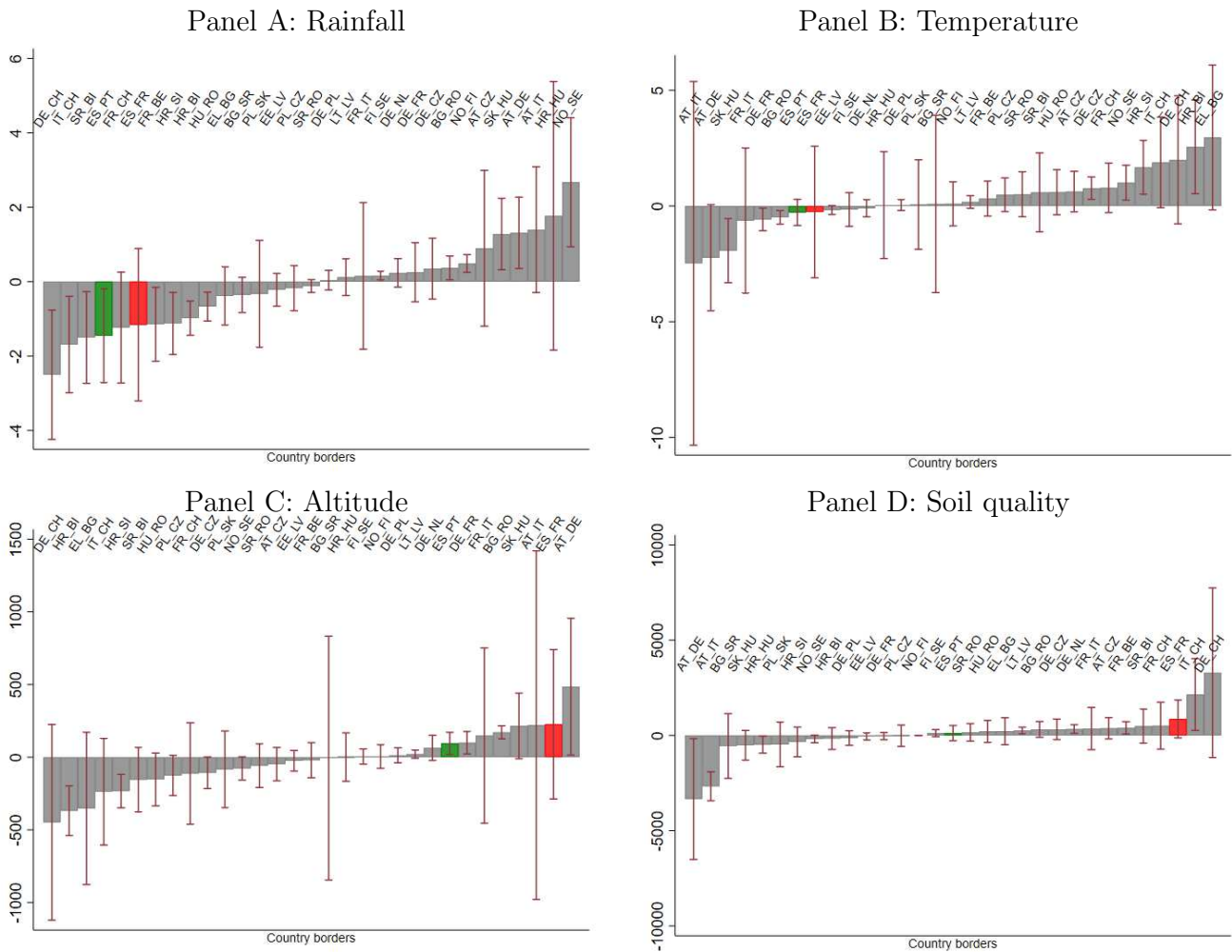
Notes: Observations to the left of the vertical dotted line pertain to Spain, while those to the right belong to France and Portugal. The solid lines display the regression of a second-order polynomial at each side of the border. Each observation is the local average of grid-cells falling within each 5 km distance-to-the-border bin.

Figure A8: Settlement and population patterns at the European borders with the GEOSTAT 2006 grid.



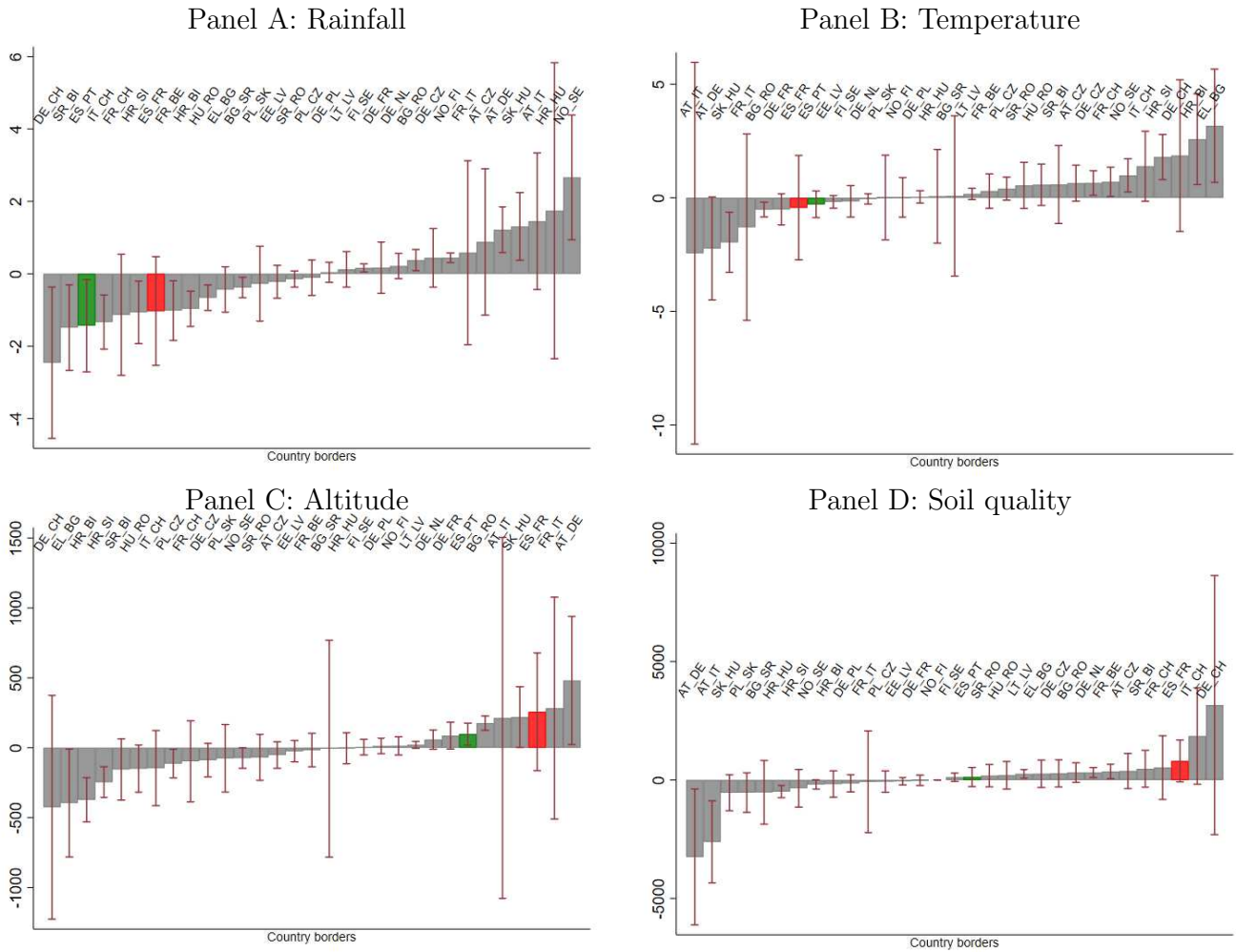
Notes: This figure uncovers settlement discontinuities at the European borders. It plots the country dummy coefficients of equation (2) for four different indicators of settlement and population patterns. In each border country-pair, the country dummy (η_c) refers to the first one. The border between Spain and Portugal is depicted in green, where that between Spain and France is colored red.

Figure A9: Geo-climatic discontinuities at the European borders excluding geographical position variables.



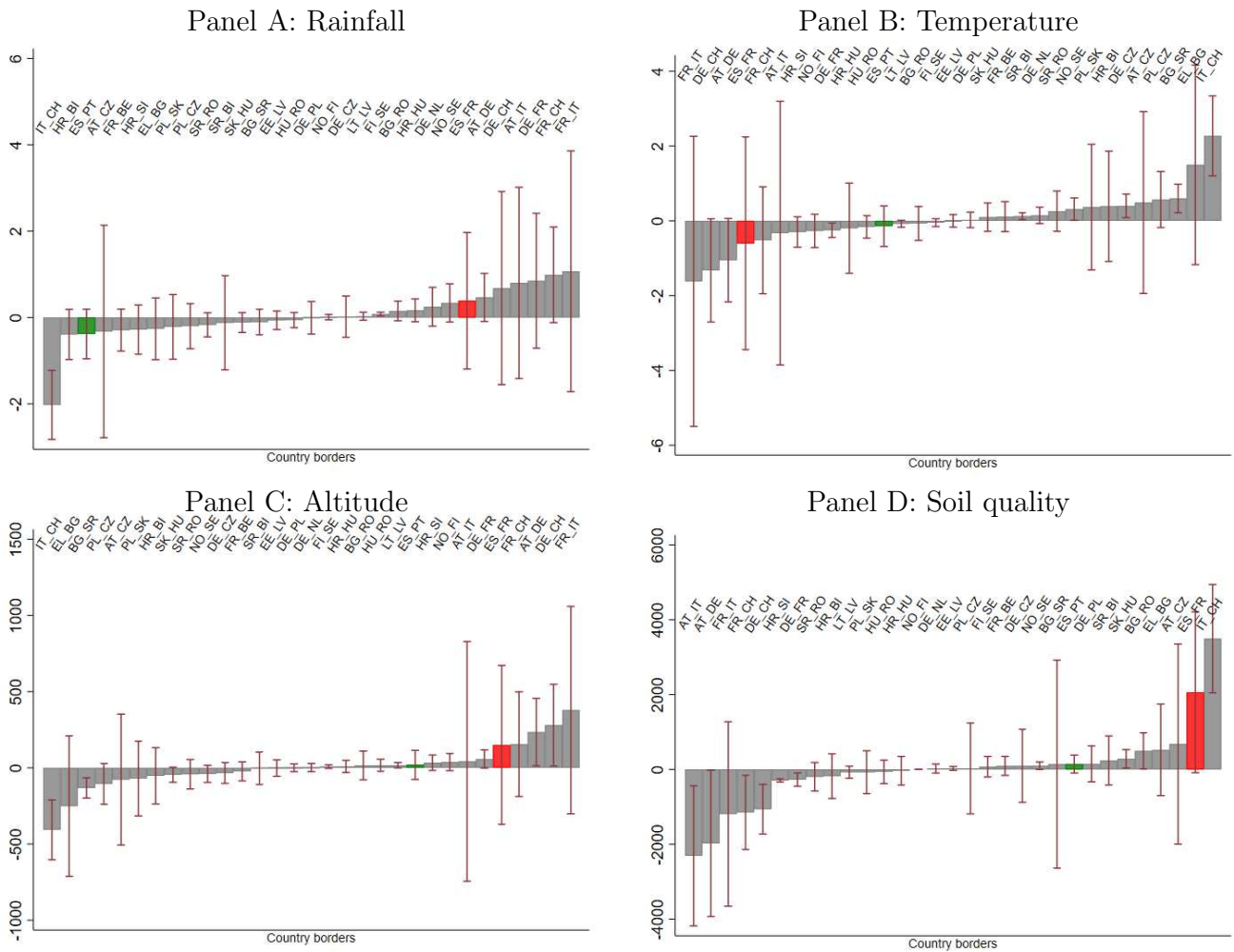
Notes: This figure uncovers geo-climatic discontinuities at the European borders. It plots country dummy coefficients after regressing geo-climatic factors on the country dummy and segment dummies. In each border country-pair, the country dummy refers to the first one. The border between Spain and Portugal is depicted in green, where that between Spain and France is colored red.

Figure A10: Geo-climatic discontinuities at the European borders including distance and its square.



Notes: This figure uncovers geo-climatic discontinuities at the European borders. It plots country dummy coefficients after regressing geo-climatic factors on the country dummy, segment dummies, and geographical position variables. In each border country-pair, the country dummy refers to the first one. The border between Spain and Portugal is depicted in green, where that between Spain and France is colored red.

Figure A11: Geo-climatic discontinuities at the European borders including distance and interaction with country dummy.



Notes: This figure uncovers geo-climatic discontinuities at the European borders. It plots country dummy coefficients after regressing geo-climatic factors on the country dummy, segment dummies, and geographical position variables. In each border country-pair, the country dummy refers to the first one. The border between Spain and Portugal is depicted in green, where that between Spain and France is colored red.

by regressing geo-climatic factors on the country dummy, segment dummies, and the geographical position variables.

Appendix B Construction of the dataset on settlement and population patterns in Europe in the second half of the 19th century

We gather data on localities and population c. 1871 at the district level for five European countries. The definition of settlements employed is a group of inhabited houses, ranging from villages to cities, and excluding isolated farmhouses and dispersed population. As mentioned in Section 4.1, 19th-century European censuses are not necessarily strictly homogeneous when defining and counting settlements. Therefore, results derived from this data must be interpreted with some caution. Regarding data on population density, the unit of measurement is clear (number of inhabitants) so it is more straightforward. We next describe the data sources and the steps to construct the dataset of European districts.

i. France

The 1891 census is the first available (to our knowledge) that provides data on localities (“hameaux de villages ou de sections de communes”) for each département (INSEE (2020)). France had about 36,000 communes in 1891 and more than 490,000 localities. Even though 490,000 seems a very large number, it plausibly reflects the very dispersed French population patterns. This global number renders an average of 13.6 localities per commune, which is plausible. Moreover, this number is very likely to exclude isolated farmhouses. According to INSEE (1951), there were 700,000 “lieux-dits” in France after WWII. “Lieux-dits” is a broad concept that includes any inhabited place, including farmhouses. Therefore, 490,000 “hameaux” (group of houses) in 1891, when the population was even more dispersed, is reasonable. Focusing on a specific département, Finistere, there were 7,880 “hameaux” in 1891, while INSEE (1951) mentions 26,777 “lieux dits” in the mid-20th century. Therefore, this points to an extremely dispersed settlement pattern in France.

We calculate the ratio hameaux/communes for each département, and then apply this ratio to the number of communes in each district (arrondissements) to obtain an estimation of the number of localities in each district. Regarding data on population, we gather district-level data from the 1872 population census (INSEE (2020)). There are a total of 284 French districts in our dataset.¹

¹We thank David Le Bris for providing guidance on French historical data.

ii. Italy

The 1871 census provides the number of “centri de popolazione” (localities) for each compartimenti (ISTAT (1876)). There were a total of 25,337 “centri” in Italy in 1871. First, we calculate the ratio centri/frazioni for each compartimenti. “Frazioni” are divisions of communes, which do not always correspond to “centri de popolazione”. Second, we estimate the number of “centri” at the provincial level by multiplying the number of “frazioni” by the previous ratio. We do so because the number of “centri” is not available at the provincial level. Third, we calculate the ratio centri/communes at the provincial level. Finally, we estimate the number of “centri” at the “circondari” (district) level by multiplying the number of communes by the ratio centri/communes calculated in the previous step. Concerning population data, we are very grateful to Luca Salvati for sharing data on population at the commune level for the year 1871 (Salvati (2014)). There are a total of 284 Italian districts in our dataset.

iii. Prussia

We obtain county level data on Prussian localities and population from Becker and Cinnirella (2020), who gathered data “for the universe of localities in 1871 Prussia provided by the official population census” (p. 264).² This dataset includes all Prussian towns, villages and manors, amounting to more than 50,000. The data corresponds to the 1871 census. We aggregate data on localities at the county (district) level. There are 428 Prussian districts in the sample.

iv. Portugal

We collect data on localities from the 1960 population census (Instituto Nacional de Estatística (1964)) and data on population from the 1878 census (Estatística de Portugal (1881)).³ Regarding settlement data, the 1960 census is the first one that provides data on localities (“lugares”). “Lugares” mean a group of houses, containing at least five homes (families), with a known toponym. This therefore refers to cities, towns or villages. The 1960 census contains the population of each one of the 38,950 “lugares” existing in 1960 and, importantly, their population in 1911. Based on this information, we first collect the number of “lugares” in each “concelho” (municipality) in 1960, which is provided in Table 2 of Tomo 1-vol 2. Second, we select a sample equal to the 10% of “lugares” in each district, amounting to 3,869 in total. Using this sample, we calculate for each district the ratio “lugares 1911/lugares 1960”, where the number of “lugares” in 1911 is extracted manually from the census on an individual basis. Finally, we estimate the number of localities (“lugares”) existing in 1911 at the municipality (concelho) level by multiplying the number of “lugares” in 1960 by the

²We are very thankful to Sascha Becker and Francesco Cinnirella for sharing the data.

³We thank Daniel Alves, Nuno Lima, and Pedro Machuqueiro for sharing the geographic layer of Freguesias in 1878 (Alves et al. (2020)).

previous ratio.

Concerning population, we gather data from the 1878 census. To be able to compare the population values of Portugal with the other countries (whose population is measured in 1871-72), we divide the municipality population values in 1878 by the rate of population growth between 1871 and 1878 at the national level.

There are a total of 257 municipalities in the dataset. Given the smaller surface area of Portugal, this means that the average size of observations for Portugal is smaller than for the rest of the countries analyzed.

v. Spain

Data on population and localities for Spanish districts (“partidos judiciales”) come from the 1887 population census and has been obtained from Beltrán-Tapia et al. (2020), who have kindly shared their data with us. Like in the previous cases, the types of localities considered include cities, towns, and villages, excluding farmhouses (“caseríos). There are approximately 43,000 localities in Spain in 1887. Regarding data on population, we proceed similarly to the case of Portugal and adjust the district level population by the growth rate at the national level between 1871 and 1887. There are a total of 469 Spanish districts in the dataset.

References

- Alves, Daniel, Nuno Lima, and Pedro Machuqueiro**, “Carta de Freguesias de Portugal Continental de 1878,” 2020. Luís Espinha da (ed.), *Atlas, Cartografia Histórica*, available at <http://atlas.fcsh.unl.pt/>.
- Becker, Sascha O and Francesco Cinnirella**, “Prussia disaggregated: the demography of its universe of localities in 1871,” *Journal of Demographic Economics*, 2020, 86 (3), 259–290.
- Beltrán-Tapia, Francisco J., Alfonso Díez-Minguela, Julio Martínez-Galarraga, Alicia Gómez-Tello, and Daniel Tirado**, “Spain, from the Ancien Régime to the Liberal State (ESPAREL),” 2020. Research project: <https://www.ehvalencia.es/en/espana-del-antiguo-regimen-al-estado-liberal-esparyl-en/>.
- Estatística de Portugal**, “II Recenseamento Geral da População, 1878 - 1 de Janeiro,” 1881. Available at <https://censos.ine.pt/>.
- Galor, Oded and Ömer Özak**, “The agricultural origins of time preference,” *American Economic Review*, 2016, 106 (10), 3064–3103.
- Hijmans, Robert J., Susan E. Cameron, Juan L. Parra, Peter G. Jones, and Andy Jarvis**, “Very high resolution interpolated climate surfaces for global land areas,” *International Journal of Climatology*, 2005, 25 (15), 1965–1978.
- INSEE**, “Nomenclature des hameaux, écarts, lieux-dits du Finistere. Première partie : Liste alphabétique par commune,” 1951.
- , “Données historiques de la Statistique générale de France SGF 1800-1925,” 2020. Available at www.insee.fr.
- Instituto Nacional de Estadística**, “CENSO de 1887 ”Floridablanca”,” 1887.
- Instituto Nacional de Estadística**, “X Recenseamento Geral de População no continente e ilhas adjacentes às 0 horas de 15 de Dezembro de 1960 são editados em VI Tomos e um Anexo.,” 1964. Available at <https://censos.ine.pt/>.
- ISTAT**, “Popolazione Presente ed Assente per Comuni, Centri e Frazioni di Comune. Censimento 31 dicembre 1871,” 1876.
- Pleiades**, “Pleiades: a gazetteer of past places,” 2021. Available online at: <http://pleiades.stoa.org/> (accessed 30 March 2021).

Salvati, Luca, “Population distribution and urban growth in Southern Italy, 1871–2011: emergent polycentrism or path-dependent monocentricity?,” *Urban geography*, 2014, 35 (3), 440–453.