The Medieval Frontier Origins of a Country's Economic Geography: The Case of Spain

Daniel Oto-Peralías
University of St Andrews

This version: September 2017

Abstract

This paper shows the potential of historical frontiers to shape the economic geography of countries. I focus on the case of Spain to explore how extreme frontier insecurity in the Middle Ages can condition the colonization of the territory in such a way to make it one of the most desert areas in Europe. First, I document that Spain stands out in Europe with an anomalous settlement pattern characterized by a very low density in its southern half, which is not explained by geographic and climatic factors. Second, I exploit a spatial discontinuity that was meaningful during the Christian colonization of central Spain in the Middle Ages to investigate the historical roots of this phenomenon. The findings suggest that medieval frontier insecurity heavily conditioned the colonization of the territory, resulting in a very sparse occupation of the affected region, a high degree of militarization, and a ranching orientation of the economy. These initial features of the colonization process led to a remarkably low level of settlement density that has persisted to this day.

Keywords: Medieval Frontiers, Settlement patterns, Urbanization, Spain, Europe, Spatial discontinuity, Warfare.

JEL Classification: C14, N90, O1, R10.

Daniel Oto-Peralías. University of St Andrews, School of Management. The Gateway, North Haugh, St Andrews, Fife, KY16 9AJ, UK. E-mail: dop2@st-andrews.ac.uk.

1. Introduction

This article shows that historical frontiers can decisively shape the economic geography of countries. Using Medieval Spain as a case study, I explore how extreme military insecurity in frontier regions can condition the political and economic occupation of the territory to such an extent to convert it into one of the most desert areas in Europe. The first part of the analysis documents that Spain has an anomalous settlement pattern characterized by a very low density in its southern half. This is a particularity of the country not found in its European neighbors, while econometric evidence seems to rule out geographic and climatic factors as main explanations for this anomaly. The second part investigates the historical origins of southern Spain's low settlement density, with the findings pointing to the dynamics of a large and insecure frontier region, and the associated militarized and "ranching" style of colonization, as the determinants of this.

Figure 1 shows an indicator of settlement patterns in Europe at the NUTS 3 level. It measures the density of settlements through the percentage of 10-km^2 grid cells in each region that are inhabited. In many European regions values are close to 100%, which means that it is rare to find uninhabited 10-km^2 cells. The picture is naturally different in high-latitude regions such as Iceland and the northern part of Scandinavia, where settlement density is low, but Spain stands out as a separate case. A large part of its territory, particularly the southern half, has a remarkably low density of settlements. For instance, among the ten European regions with the lowest settlement density, six are from southern Spain (while the other four are from Iceland, Norway, Sweden, and Finland). Extreme geographic and climatic conditions do not seem to be the reason for this. I use the term 'Spanish anomaly' for this remarkably low density of settlements not explained by geography and climate.[1] The average settlement density of southern Spain is indeed one of the lowest in Europe, and the lowest when controlling for geographic and climatic factors. This result is robust to an exercise in which I compare southern Spain to more than 1,000 virtual regions and when using 250-km^2 grid cells as units of analysis.

[*Insert Figure 1 about here*]

The central part of the paper investigates the medieval frontier origins of southern Spain's low settlement density. 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" (F. Cornaro, 1678-81, G. Cornaro, 1681-82, quoted in Brenan 1950, p. 128). To shed light on this issue, I exploit a geographic

---

[1] It is worth noting that low settlement density does not necessarily imply low population density. Both variables are related but correlation is far from perfect (it is about 56% between the settlement patterns indicator and the log of population density). Figure A1 in the Supplementary Material shows a map of population density, and Spain does not look so different. Thus, southern Spain's low settlement density reflects a spatial distribution of the population characterized by its concentration on urban centers and a relatively empty countryside.

discontinuity that was meaningful during the Middle Ages in the context of the Spanish Reconquest. "Historical accidents" made the colonization of the area south of the Tagus River very different from the colonization north of it. The invasions of the Almoravids and Almohads converted the territory south of the Tagus into a battlefield for one-and-a-half centuries, this river being a natural defensive border. Continuous warfare and insecurity heavily conditioned the nature of the colonization process in this frontier region, which was characterized by the leading role of the military orders as colonizer agents, scarcity of population, and a livestock-oriented economy (González Jiménez, 1992). The implications were the prominence of great castles and the absence of villages and towns, and consequently, a spatial distribution of the population characterized by a very low density of settlements. Bishko (1963) referred to this style of colonization as a "medieval ranching frontier". As argued below, the interplay between initial militarization and ranching contributed to the persistence of low settlement density.

The empirical analysis supports this historical narrative by revealing a statistically significant jump in settlement patterns across the River Tagus, whereas there are no geographic and climatic discontinuities across it. The results are robust to many specification tests and several falsification exercises. In addition, I collect census data on population entities in the 16th and 18th centuries to show that the discontinuity in settlement density already existed in that period, and therefore is not the result of migration movements and urban developments taking place in the modern or contemporary eras. Preliminary evidence also indicates that the territory exposed to the medieval ranching frontier is relatively poorer today.

This paper contributes to the new empirical literature on frontier societies (García-Jimeno and Robinson, 2011; Oto-Peralías and Romero-Ávila, 2016, 2017; Droller, 2017).[2] Compared to previous work, this article focuses on the effect of frontiers on the spatial distribution of the population across the territory, which is the basic layer on which all economic activities are built. As the Spanish case shows, historical frontiers can shape a country's economic geography. The exposure to warfare and insecurity –typical in medieval frontiers– creates incentives for a militarized colonization of the territory based on a few fortified settlements and a livestock-oriented economy. This mechanism can also operate in similar historical contexts, although the remarkably low settlement density in southern Spain is the legacy of remarkably high frontier

---

[2] García-Jimeno and Robinson (2011) analyze the divergent frontier experiences in the New World and argue that the outcome of frontier expansions depends on the initial political equilibrium, with negative outcomes being associated with oligarchic political systems. Oto-Peralías and Romero-Ávila (2016) focus on the Spanish Reconquest and propose that the political equilibrium and the outcome of the colonization process may be endogenous to the pace of the frontier expansion, with a fast expansion favoring the concentration of land and political power in the hands of the elite. Using the frontier of Granada as a case study, Oto-Peralías and Romero-Ávila (2017) argue that the dynamics of being a militarily insecure frontier region –and the associated defense needs– can create the conditions for highly unequal societies. Droller (2017) exploits the frontier expansion in the Argentinean Pampas to analyze the impact of population composition on economic development. Related from a methodological point of view, there are also several studies that analyze former historical borders between countries or empires to exploit discontinuities in culture or institutions (e.g., Grosjean, 2011; Becker et al. 2016).

violence and insecurity.

This paper also contributes to a growing literature on the effect of military conflicts on urban growth, state capacity and economic development. In contrast to previous works finding long-term positive effects of conflicts (e.g., Voigtländer and Voth, 2013; Dincecco and Onorato, 2016), this analysis shows that historical warfare may have negative consequences in certain contexts. Thus, in consolidated states or kingdoms it may foster urban growth and fiscal capacity, but in frontier regions it may lead to negative outcomes. As shown in this paper, under the conditions of frontier expansion, intense militarization and a ranching-oriented economy, a large region can get trapped in a vicious circle of low settlement density, with negative implications for development.[3]

Moreover, this article's focus on settlement patterns represents a new approach to look at a central topic in economic geography, namely the spatial distribution of the population and economic activity (Ottaviano and Thisse, 2004). How villages, towns and cities are distributed across the space significantly affects economic interactions and may influence economic development in the short- and long-run. The results of the analysis reveal a strong persistence in the settlement structure of the territory, which adds to the extant literature that shows that cities and the urban network are highly persistent and to the debate regarding the role played by location fundamentals in shaping the location of cities (Bleakley and Lin, 2012; Michaels and Rauch, 2017; Bosker and Buringh, 2017). Finally, this paper is also related to a body of research on the interplay between history and geography, more specifically, on the contingent role of geographic factors in development (e.g., Dell 2012, Nunn and Puga 2012, Belloc *et al.* 2017). As argued below, the River Tagus created a discontinuity in settlement patterns and development due to a contingent factor such as high military insecurity, which in turn was the consequence of the Almoravid and Almohad invasions during the 11th and 12th centuries. It was the interaction between history and geography that left its lasting imprint on Spain's economic geography.

The rest of the paper is organized as follows. Section 2 explores the determinants of settlement density in Europe, documents a Spanish anomaly in settlement patterns, and investigates whether geography can explain this. Section 3 provides a brief historical discussion about the medieval frontier origins of southern Spain's low settlement density. Sections 4 and 5 conduct an econometric analysis to shed light on this issue. Finally, Section 6 puts forward some implications and concludes.

---

[3] There are also other recent papers arguing that the consequences of warfare on state development depend on the historical context. Gennaioli and Voth (2015) build a model in which military conflicts are positive for state capacity when fiscal resources become crucial for winning wars, which happened after the "military revolution" of the 16th-17th centuries. Dincecco *et al.* (2016) find that historical warfare in Sub-Saharan Africa is associated with "special-interest states", characterized by high fiscal capacity and *high* social conflict, while in the rest of the Old World it is associated with "common-interest states" (i.e., high fiscal capacity and *low* social conflict).

2. The Spanish anomaly in settlement patterns in Europe

This section aims to analyze settlement patterns in Europe and to provide evidence on the existence of a Spanish anomaly, which is not satisfactorily explained by geographic and climatic factors. I first describe the data, then conduct an analysis at the NUTS 3 level, and finally use a grid of 250-km^2 cell size.

2.1 Data

The main data source employed to create indicators of settlement patterns is the GEOSTAT 2011 population grid (Eurostat 2016a). The GEOSTAT population grid is a convenient tool to measure settlements because it provides population data at a very high 1-km^2 resolution. The presence of a settlement can be inferred wherever there is a populated grid cell. Note that I am not interested in how many people live in each cell but in the existence of a populated cell as an indication of a settlement. An important advantage of this way of identifying settlements is its homogeneity across countries. Other alternatives such as data on municipalities or on local administrative units cannot be used for comparative purposes since they are heterogeneous and depend on the legal and political system of each country. The sample consists of the territory covered by GEOSTAT 2011 after excluding the overseas regions (see Figure 1). The total surface area is slightly larger than 5 million km^2, and there are about 2.08 million populated 1-km^2 grid cells.

The indicator of settlement density (*SD*) of cell-size *s* in region *i* is constructed as follows:

$$SD_i^S = \frac{\sum_1^N PC_{n,i}^S}{\sum_1^N C_{n,i}^S} \times 100$$

where $\sum_1^N PC_{n,i}^S$ is the number of populated grid cells of size *s* in region *i*, and $\sum_1^N C_{n,i}^S$ is the total number of grid cells of size *s* in region *i*. Therefore, SD_i^S represents the percentage of populated cells of size *s* over the total number of cells of that size in region *i*. Grid cell *n* in region *i* is considered to be populated if it contains at least one 1-km^2 populated cell. Note that if *s* is equal to 100 (i.e., 100-km^2), then SD_i^{100} measures the percentage of populated 100-km^2 cells, or, in other words, the percentage of 100-km^2 cells with at least one 1-km^2 populated cell within them. It may now become apparent why the indicator of settlement density is not closely correlated with population density. Indeed, the higher the cell size *s*, the lower the correlation.[4] To illustrate the methodology, Figure 2 shows the construction of the indicator for the NUTS 3-region ES612 (Cádiz, Spain) and a cell size of 100-km^2. To create the indicator, a layer with the region boundaries is intersected by a 10-km x 10-km grid. The total number of grid cells whose centroids are within the region's boundaries (i.e., $\sum_1^N C_{n,i}^S$) is 78. From these 78 cells, all but 10 are populated (that is, contain at least one 1-km^2 populated cell). Therefore, settlement density for this NUTS is 68/78*100, which equals 87.18%.

---

[4] The coefficient of correlation between log population density and the indicator of settlement density is 0.56, 0.48, and 0.36 for cell sizes of 10, 25 and 100-km^2, respectively.

[*Insert Figure 2 about here*]

I construct many geographic and climatic variables, including temperature, rainfall, average altitude, ruggedness, distance from the coast, etc. To save space, the definitions and sources of all the variables are presented in Table A1 at the end of the main text, while the descriptive statistics are reported in Table A2 in the Supplementary Material.

2.2 Analysis at the NUTS 3 level

Is southern Spain's low settlement density –as depicted in Figure 1– explained by geographic and climatic factors? After all, geography plays a crucial role in development (Gallup *et al*. 1999), and many observers and scholars have traditionally pointed to the adverse geography of Spain (particularly, its central plateau) as a reason for its economic backwardness.[5] However, on careful consideration, this does not seem very convincing. The low levels of settlement density in southern Spain are only surpassed in exceptionally adverse geographies (e.g., areas close to the Arctic). Arguably, climatic and geographic conditions in southern Spain are not so extreme as to account for its remarkably low settlement density. Therefore, an econometric study of this question is conducted. The empirical strategy is simple: to analyze the differences in settlement patterns between southern Spain and the rest of Europe through a regression model that controls for geographic and climatic factors. If after controlling for these factors there is still a sizable difference, then it can be inferred that historical factors –rather than geographic ones– are behind the Spanish anomaly.

To conduct this analysis, it is first necessary to select an appropriate cell size for the settlement density indicator. I need to reconcile economic meaningfulness with sufficient data variation. A value of 10-km^2 seems to meet both criteria. On the one hand, 10-km^2 is a meaningful size from an economic point of view. In a balance occupation of the territory, we would expect every 10-km^2 to have at least one settlement. For instance, the average size of a commune in France is 15-km^2 and, typically, each commune has more than one settlement. On the other hand, the data have sufficient variation to conduct a regression analysis since only 27.8% of the NUTS 3-regions have values of SD_i^{10} equal to 100, while if –for example– the cell size were 100-km^2 (SD_i^{100}), then 86.8% of the regions would have a value of 100. Importantly, in the regression exercise at the NUTS 3 level, observations are weighted by their surface area since the relevant unit of analysis in this study is the territory itself rather than the administrative divisions.[6] Throughout the analysis, I correct for spatial dependence by clustering standard errors at the country level.

The first four columns in Table 1 explore the climatic and geographic determinants of settlement patterns. I start in column 1 by regressing settlement density on a quadratic polynomial

---

[5] An example of the traditional view of adverse geographic conditions is the book *Los Males de la Patria* (i.e, "the evils of the country") by Lucas Mallada (1890), where it is argued that only 10% of Spanish land "leads us to suppose that we have been born in a privileged country" (quoted in Simpson, 1995, p. 34).

[6] In addition, this mitigates the problem of the high heterogeneity in NUTS surface areas and the consequent undue influence of countries like Germany (with a highly fragmented NUTS division) on the overall results.

in temperature, which turns out to be a powerful predictor of settlement density. It alone explains 52% of the variation in the dependent variable.[7] There is a (nonlinear) inverted U-shaped relationship between temperature and settlement density, with its peak being at approximately 9 degrees Celsius. Next, I include rainfall, which also shows a nonlinear relationship, although with a much lower explanatory power (the joint significance test is 0.09). Column 3 adds average elevation, which enters with a large negative coefficient increasing R^2 by 10 points. Column 4 further includes additional geographic indicators, namely ruggedness, soil quality, distance to the coast (linear and squared) and an island dummy. All these variables add little explanatory power to the model and are mostly statistically insignificant. Therefore, my baseline set of geo-climatic controls is that of column 3.

The rest of the table analyzes the differences in settlement density across Europe. Column 5 regresses settlement density solely on a set of country dummies. The reported coefficients consequently reflect the mean differences with respect to the reference category, which is France.[8] Spain appears split into two parts, northern and southern Spain.[9] Remarkably, southern Spain has lower settlement density than the Scandinavian countries; it is only surpassed by Iceland. The difference compared to its neighbors is huge. The average settlement density in France is more than three times higher than in southern Spain, while in Portugal it is almost three times higher. The contrast with the rest of Spain is also remarkable: the northern half of Spain –along with the Balearic Islands– has almost twice as much density as the southern half.

To analyze whether this difference can be explained by geography and climate, column 6 includes the baseline climatic and geographic variables along with the country dummies. Southern Spain is now the region with the lowest settlement density. Climatic and geographic factors can account for the settlement patterns of Norway, Finland, Sweden, and Iceland. In all these cases, the coefficient experiences a dramatic increase (i.e., becomes less negative). In contrast, the negative coefficient on southern Spain is barely affected. This suggests that climate and geography are not the keys to explaining the Spanish anomaly. Finally, column 7 includes additional geographic variables. As expected by the previous result in column 4, the explanatory power of the model is unaffected and the Spanish anomaly remains clear.

[*Insert Table 1 about here*]

Although this evidence seems convincing, a potential concern is that I am comparing countries that are very heterogeneous in surface area. One could suspect that the singularity of southern Spain is perhaps conditional on the specific territorial division used in the previous analysis. To

---

[7] I follow Burke *et al.* (2015) in using quadratic polynomials in temperature and precipitation.

[8] For space considerations, only the five largest negative coefficients are reported. Table A3 in the Supplementary Material contains the full list of coefficients.

[9] Southern Spain refers to the 17 mainland provinces located south of Madrid (i.e., the eight Andalusian provinces: Badajoz, Cáceres, Murcia, Albacete, Ciudad Real, Toledo, Cuenca, Alicante and Valencia). For simplicity, I do not further delimit here the region that actually corresponds to the Spanish anomaly. This is done in the next section.

address this concern, I compare southern Spain to “virtual regions” created using the following algorithm. For each NUTS *i*: i) its 100 nearest neighbors are identified, ii) the pairwise distance with each of them is calculated, and iii) *virtual region i* is created, being composed of NUTS *i* and its neighbors falling within 300 km.[10] Southern Spain is excluded from the algorithm, so none of its NUTS can be part of any virtual region. I estimate 1,423 regressions (one for each virtual region created) where the independent variables are the baseline set of geo-climatic controls, the southern Spain dummy, and the virtual regions one by one. Remarkably, in more than 99% of cases, the coefficient on southern Spain is lower (i.e., *more* negative) than on the artificial region. Figure 3 depicts the cumulative distribution of the difference between the coefficients on the virtual regions and on southern Spain. A positive difference means that settlement density is lower in southern Spain. As is apparent, almost all the distribution lies to the right of zero, confirming the above evidence about the Spanish anomaly.[11]

[*Insert Figure 3 about here*]

2.3 Analysis at the grid cell level

This section uses grid cells as units of analysis. This is a useful robustness check because –besides increasing the number of observations and, consequently, the statistical power– it mitigates any potential problem created by the endogenous formation of NUTS boundaries. In addition, I can increase the precision of the analysis at this finer level by narrowing down the geographical delimitation of the territory in southern Spain corresponding to the Spanish anomaly. As argued in Section 3, this is the area delimited by the Tagus to the north, Portugal to the west, the former Aragonese border to the east, and the former frontier of Granada to the south (see Figure 4). Grid cells have a surface area of 250-km^2, and the total number of cells in the sample is 23,498. The variable settlement density is measured as above but now the subscript *i* refers to grid cells rather than NUTS.

Column 1 in Table 2 shows that the region corresponding to the Spanish anomaly has about 70 percentage points less settlement density than France, the omitted category. The contrast within Spain is also large. The average settlement density of the rest of Spain is more than twice than in the region of interest (see Table A4 in the Supplementary Material). Even Norway has a higher value, and only Iceland has lower settlement density. This is remarkable since these are unconditional differences. When controlling for climatic and geographic factors in columns 2 and 3, the area corresponding to the Spanish anomaly stands out as the territory with the lowest settlement density.[12]

---

[10] This limit of 300 km is chosen to create regions with a surface area similar to southern Spain.

[11] Results are virtually the same without including the geo-climatic controls or when using the full set of controls.

[12] There are other regions with relatively low levels of settlement density, such as Macedonia, the rest of Spain and Greece, but the region that really stands out is the area corresponding to the Spanish anomaly. Figure A2 in the Supplementary Material represents the distribution of the coefficients on the region

[*Insert Table 2 about here*]

To check the robustness of this result, I also systematically compare the region of interest with virtual regions (as in the NUTS level analysis). Given the large number of observations, I do not create one virtual region for each 250-km^2 grid cell. Rather, I randomly choose 1,000 grid cells and create virtual regions by selecting up to 1,000 neighbors falling within 300 km from the cell's centroid. Then I run 1,000 regressions of settlement density on the set of geo-climatic controls, the region of interest, and the virtual regions (included one by one). In 98% of cases, the value of settlement density for the area corresponding to the Spanish anomaly is lower than for the virtual region.[13]

To sum up, the evidence presented so far indicates that a region that approximately corresponds to southern Spain has a remarkably low density of settlements, even lower than Scandinavian countries. Geographic and climatic factors fail to account for this anomaly, which suggests that the peculiarities of Spanish history are behind this. The rest of the paper tries to shed light on the medieval frontier origins of southern Spain's low settlement density.[14]

3. On the medieval frontier origins of settlement patterns in Spain: Historical background

This section introduces the hypothesis that the low level of settlement density in southern Spain is linked to its character as an insecure military frontier region during the Middle Ages. First, I conduct a brief historical overview of the key period in which southern Spain was conquered and resettled by Castile and Leon in the Middle Ages. Second, I focus on the geographic discontinuity created by the River Tagus as a meaningful defensive barrier. Finally, I discuss the persistence of settlement patterns over time.

3.1. The Spanish Reconquest and the Christian colonization of southern Spain

The Reconquest is the formative process of modern Spain. During a period of almost 800 years, the northern Christian kingdoms were gradually conquering the territory under Muslim rule. The conquest was followed by a process of repopulation or resettlement of the territory. This process of territorial expansion made Spain a frontier society over a long period of time. The frontier conditions prevalent in each period and the subsequent way the territory was colonized had lasting consequences for the country's future development. This is a well-known hypothesis in Spanish historiography, first developed by Sánchez-Albornoz (1962) and recently tested by Oto-Peralías and Romero-Ávila (2016, 2017). Several factors affected the type of colonization conducted in

---

dummies from the regression reported in column 3 (Table 2). As can be seen, there is a jump in the size of the coefficient in the case of the Spanish anomaly.

[13] Figure A3 in the Supplementary Material shows the cumulative distribution of the difference between the coefficients (as in Figure 3).

[14] This paper's results are fully robust to the use of the earlier version of GEOSTAT (i.e., GEOSTAT 2006) as a data source for settlement density.

each region. One was the size of the area that had to be colonized. A larger area made it difficult for the Monarchs to repopulate and defend the territory, thereby requiring the active involvement of the nobility and military orders, which led to the concentration of land and political power in their hands. Another important factor was military insecurity, which also led to the prominence of the military elite in the colonization.

When considering whether this process of medieval colonization is important to account for what I have called the Spanish anomaly, it is very revealing that the low settlement density area of southern Spain approximately matches a historically well-defined territory: most of Extremadura and Castilla-la-Mancha, Murcia, and the Castilian part of Andalusia. This is the area conquered by Castile and Leon during the 12th and 13th centuries. More precisely, this territory is delimited by the Tagus to the north, the former Aragonese border to the east, Portugal to the west, and the former Granada frontier to the south. Figure 4 represents this area along with the indicator of settlement density. Remarkably, the borders of the highlighted territory delimit reasonably well the area of low settlement density.[15]

[*Insert Figure 4 about here*]

Two periods can be differentiated in the conquest of this large territory. The first one dates from the conquest of Toledo in 1085, which brought the frontier down to the River Tagus, to the battle of Las Navas de Tolosa in 1212. During this period, Castile and Leon struggled to conquer and defend key positions in the southern plateau. The military insecurity characteristic of frontier regions –and exacerbated by the invasions of Almoravids and Almohads– largely conditioned the way the territory was colonized. Historians have long identified militarization as the main feature of medieval frontiers (Berend, 1999), and militarization reached a peak in the southern plateau, with no better example than the rise of the military orders as the best alternative to defend the dangerous frontier positions (Forey, 1984).[16] The second period extends from 1212 until the 1260s, with the consolidation of a stable frontier between Castile and the Nasrid Kingdom of Granada. I focus on the first period, which marks the transition to the area of low settlement

---

[15] One manifestation of the anomalous way this territory was settled –and the resulting low settlement density– was the existence of "agro-towns", which were the consequence of the concentration of the rural population in a few urban centers. "Agro-towns" were distributed across the southern part of Spain and represented large concentrations of farmers and landless peasants in urban centers which lacked the usual features of medieval towns and had a low level of economic activity associated with the non-agricultural sector (Reher 1990). They arose because "[t]owns provided security and lower transaction costs in a frontier economy during the repopulation process that followed the Reconquest" and due to the accumulation of land in a few hands (Álvarez-Nogal and Prados-de-la-Escosura, 2013, p. 13).

[16] The 12th century witnessed the birth of the Castilian-Leonese religious military orders of Calatrava, Santiago and Alcantara, which –like Templars and Hospitallers– had the goal of fighting the "infidel" and protecting the frontier. "The military orders were invaluable in resisting the onslaught of the Almohads and in wresting and holdings lands on the frontier" (Mackay, 1977, p. 32). In the words of Vann (1999, p. 28), "[t]he knights of the orders provided professionally trained and well-supplied forces for a wide range of offensive or defensive undertakings". In addition to this military function, the orders also fulfilled a colonizing and political function (AyalaMartínez, 2006).

density.

3.2. Extreme military insecurity in the southern plateau and the barrier of the River Tagus

What makes the period after the conquest of Toledo special with respect to the previous one is the extreme military insecurity that affected all the southern plateau, modern-day Extremadura and Castilla-la-Mancha. The territory south of the Tagus was a battlefield in the conflict between Christian and Muslim armies for almost one-and-a-half centuries (c. 1085-1212) (Bishko, 1963). This period of intensive warfare was due to an arguably exogenous factor: the invasions of the Almoravid and Almohad armies from North Africa. The fact that the Tagus was a natural defensive barrier that delimited the area subject to insecurity provides a way to identify the effect of medieval frontier conditions on settlement patterns, and consequently to study whether this region's frontier history is responsible for southern Spain's low settlement density.[17] Put differently, the Tagus was a natural military barrier that created a discontinuity in the intensity of warfare suffered by the territory, which determined how it was colonized and settled.[18]

Historical references indicate that the Tagus was indeed a frontier landmark and a natural defensive border. For instance, González Jiménez (1992) mentions that "the reconquest of Toledo had […] created a new frontier line based on the Tagus, against which all the Muslim attacks foundered" (p. 60). Rodríguez-Picavea (1999) points out that the Tagus was a landmark in that period which separated the rearguard (to the north) from the vanguard (to the south): "Obvious danger, incipient territorial articulation and sparse settlement awaited those who dared to cross this natural 'frontier'" (p. 33, author's translation). Toledo itself was well protected by the Tagus, forming a great natural moat surrounding more than two-thirds of the city (Ladero Quesada, 1984). Vann (1999) also states that, in the defensive strategy of 12th-century Castile, "[t]he geography of the area played an important role in the positional fighting that took place along the Tagus River" (p. 25), and "Castilians took, settled, held (and periodically lost) strategic areas along the Tagus River, as if engaged in a giant game of chess" (p. 24).[19]

Military insecurity heavily conditioned the nature of the colonization process south of the Tagus. Castilian-Leonese conquests were extremely precarious and territory was often lost. For

---

[17] As argued below, a key point in this identification strategy is that, while the Tagus was important from a military perspective under medieval war technology, it is not a major geographic obstacle for social and economic interactions.

[18] South of the Tagus there was a large frontier region that extended towards the mountains "Sierra Morena" (on the border with modern-day Andalusia) without precise boundaries. The only precise line was the Tagus, and north of it was the rearguard. This is why I focus on the Tagus as the delimitation between a vanguard frontier region to the south of it (affected by high military insecurity) and a relatively safer territory to the north.

[19] According to García Fitz (2001), the Almoravids developed the notion of a "front-line" located south of the Tagus that separated the territory under their control from the area beyond it, in which they undertook devastating raids. During the period of permanent and extreme violence due to the numerous attempts to retake Toledo by the Almoravids, all the Castilian positions south of the Tagus were lost, while only one north of it had the same fate (Talavera).

instance, most of the lands occupied and populated by Alfonso VII were relinquished as a result of the Almohad invasion (Sánchez-Albornoz, 1962). The resettlement of the area was based on castles and fortresses as strategic centers (Vann, 1999).[20] González Jiménez (1992), after emphasizing the ferocious violence brought to this region by the North African armies, states that "it was precisely this permanent insecurity which helps to explain the main features of New Castile and other regions" (p. 60). According to this author, central features of this region were shortage of population, a ranching-oriented economy, and the prominence of the military orders as colonizer agents. Military orders' lordships were indeed overwhelmingly located south of the Tagus and exposed to an almost permanent frontier friction (Ayala Martínez, 1996).[21]

In contrast, north of the Tagus, the territory was colonized under the king's control. This area became the rearguard of the kingdom and was vital for its defense. Towns and urban centers predominated, mostly under royal jurisdiction, with charters that granted extensive freedom and rights to settlers (González Jiménez, 1992). The area was much safer and attracted more settlers. This territory was protected by the Tagus and the system of castles surrounding Toledo, which allowed a more developed settlement and economic activities (Rodríguez-Picavea, 1999). All these conditions led to a more balanced and widespread settlement of the territory.

After defeat in Las Navas de Tolosa in 1212, Muslim military power was severely reduced, and in a short period of fifty years, Castile conquered the rest of southern Spain with the exception of the Nasrid Kingdom of Granada. Thus, in the second half of the 13th century the southern plateau finally became a secure region. The frontier had moved southward to Andalusia. However, conditions had already been created south of the Tagus for the persistence of low settlement density.

3.3 From frontier insecurity to ranching and to persistence in settlement patterns

The hypothesis that the high intensity of warfare in the southern plateau conditioned the resettlement of this frontier region, favoring the prominence of the military orders as colonizer agents and the development of a ranching-oriented economy, was put forward by Charles J. Bishko (1963). He traced the links between frontier insecurity, colonization strategy and ranching in Medieval Spain. Leonese and Castilian rulers had to rely on the military orders for the conquest, defense and colonization of the southern plateau, leading to a militarized and sparse occupation of the space. Military orders had extensive powers, including the monopoly of government in

---

[20] In this regard, González (1976a) notes that "in the 12th century an attempt to found a new city south of the Tagus required heavy expenses in money and manpower, not always available, besides an uncertain success" (p. 22, author's translation). Similarly, he observes "the frontier with the Almoravids and Almohads led, in large fields from the Tagus onwards and for many years, to the triumph of the horse and the castle" (p. 26, author's translation). The exigencies and repercussions of war prevented the full repopulation of the territory, with large areas still to be occupied in the 13th and (even) the 14th centuries (González, 1976b).

[21] According to a map provided by Ayala Martínez (1996), it can be estimated that almost 90% of the area under military orders' control was located in the territory conquered by Castile and Leon in the 12th-13th centuries.

their territory. They also sought to attract settlers by granting charters (*fueros*) to the new towns and settlements; however, these charters only conceded limited rights compared to royal towns, which helps explain the lower urban development of the area under the orders' control.[22]

The main economic activity in this frontier region "was stock raising in that advanced form, more fruitfully developed in the Iberian Peninsula than anywhere else in the medieval world, which is properly called ranching" (Bishko, 1963, p. 54). The ranching activity, developed earlier in the northern plateau and incentivized by military conditions favoring mobile assets over crops, was widespread south of the Tagus. In fact, the frontier conditions made rural labor scarce and crop-farming hazardous, thereby favoring cattle and sheep, which were mobile and less demanding (Bishko, 1952). Thus, "by the second half of the [12th] century towns […] and the military orders […] were sending their herds and flocks into the Guadiana Basin in spite of the ever-present danger of Almohade attack", and after the defeat of the Muslims and "the opening up of the richest Manchegan and Extremaduran grasslands, there occurred […] an explosive expansion of the ranching industry of the plains" (Bishko 1963, p. 54). Importantly, the main colonizer agents, the military orders, were heavily engaged in this ranching activity, owning large herds of sheep and cattle. Livestock indeed constituted one of their main sources of wealth, and was managed by specific commanders (Ayala Martínez, 1996).[23, 24]

Hence, extreme frontier insecurity led to the prominence of the military orders, ranching, and the concentration of the sparse population in a limited number of fortified locations. Once the starting point was established, factors of persistence prevented the whole region from converging on settlement density to the rest of Spain and Europe. First, the prevalence of pastoralism created

---

[22] As mentioned above, royal towns played a secondary role south of the Tagus compared to the situation to the north. Settlers under royal jurisdiction had more freedom than in lordships. For instance, lords had the power to appoint the most important local authorities and to provide justice. Even in the absence of labor services, lordships had the monopoly over several activities such as public ovens, mills, and the forest (González, 1976b).

[23] According to MacKay (1977), "[i]t was only when the Christians won the grass-lands of the plains and steppes of La Mancha and Extremadura that an integrated ranching economy emerged […] The military orders owned large flocks. In 1243, for example, the Templars and Alcantarans quarrelled over the control of 42,000 sheep in the Tagus valley" (p. 74). But not only the lords, the townsmen "also derived their wealth from pastoralism rather than from their small arable holdings" (MacKay, 1977, p. 74). This region was also the pastureland of the large transhumant flocks of northern owners. Besides, livestock was also a convenient activity in the sense that it allowed the occupation and integration of large tracks of lands. It is not questionable that "'livestock colonization' became at the end of the 'great Reconquest' a major instrument of articulation of frontier areas" (Ayala Martínez, 2006, p. 109, author's translation).

[24] Bishko (1963) also concedes a role to geography when explaining the development of ranching. Influenced by the Turner (1920)'s view of the North American frontier, he partially sees the Castilian expansion as a fight with the environment, and compares the expansion of American colonists through the Great Plains with that of Castilians through the southern plateau. In his view, both military insecurity and the adaptation to adverse natural conditions are what gave rise to the ranching activity. Below, I attempt to isolate the effect of military insecurity by focusing on a territory that is geo-climatically very similar but was subject to different intensities of warfare. The discontinuity created by the River Tagus is what allows this analysis.

vested interests among ranchers to –for instance– maintain pasturelands with the same use.[25] Second, a large percentage of the territory was under the jurisdiction of military orders. For example, in the provinces of Ciudad Real and Badajoz, 80% and 50% of the land, respectively, was in the hands of the orders in the early 16th century (López-González *et al.*, 1989). As mentioned above, they granted fewer rights to settlers than royal towns and heavily invested in livestock, which was an economic activity easier to manage both in frontier times and in subsequent centuries. Given that jurisdictional rights lasted until the 19th century, this mechanism of persistence helps explain the long-term effect of the initial frontier conditions. The 19th century also witnessed the beginning of the industrialization process in Spain. From then onwards, large tracks of underexploited land were no longer a powerful attraction for people.[26]

4. Frontiers and settlement patterns: Empirical evidence

4.1 Initial results

I first compare the level of settlement density of the region conquered by Castile and Leon during the 12th and 13th centuries –which I call, for brevity, *treated region*– with that of the rest of Spain (see Figure 4). With that purpose, Table 3 reports regressions of settlement density on the set of climatic and geographic controls and the treated region dummy. I use 250-km^2 grid cells as units of analysis and report Conley (1999)'s standard errors robust to spatial correlation of unknown form.[27] A grid cell is considered treated if its centroid falls within the treated region. To avoid having grid cells that only partially overlap with the treated area, I delete observations with a percentage of surface area falling within the treated region in the range 20%-80%. Column 1 starts with the bivariate relationship between settlement density and the treated region dummy. The coefficient indicates that settlement density is on average more than 30 percentage points lower in the treated region. This dummy variable alone explains 28% of the variation in the dependent variable. Column 2 adds the baseline geo-climatic controls, i.e., the quadratic polynomials of rainfall and temperature along with average altitude. Interestingly, the coefficient on the treated region is unaffected. When including the large set of geographic controls in column 3, the coefficient remains largely stable and highly statistically significant.

[*Insert Table 3 about here*]

One might argue that I am comparing very heterogeneous regions within a single regression,

---

[25] Once ranching became the main economic activity, and landlords (i.e., military orders and nobles) and urban oligarchs invested in livestock, there were incentives in place to maintain the same economic structure.

[26] The literature on agglomeration economies, although mainly concerned with the location of cities, is consistent with the idea of persistence in settlement patterns and the important role of history. Indeed, regarding the spatial distribution of economic activity, it is considered that there is "path dependency in the structure of the equilibrium, with history being as important as current circumstances" (Henderson *et al.*, 2001, p. 84).

[27] I employ cutoffs of 1 decimal degree, beyond which spatial correlation is assumed to be zero.

and although many climatic and geographic indicators are included in the model, there may be omitted factors which bias the coefficient downward. To address this issue, I create 25 virtual regions by dividing the sample into five quantiles of longitude, and then each quantile of longitude into five quantiles of latitude. Each resulting virtual region has approximately the same number of observations (87 or 88). Next, I include this set of 25 virtual region dummies in the equation, which implies that the regression is now comparing grid cells within relatively small and homogenous regions. Columns 4 and 5 report the results of this exercise. Reassuringly, the coefficient on the variable of interest remains largely unchanged. To sum up, this section shows that settlement density is much lower in the well-defined historical region delimited by the Castilian-Leonese conquests of the 12th and 13th centuries than in the rest of Spain.[28]

4.2. Spatial regression discontinuity (SRD) design: Discontinuity across the River Tagus

This section goes a step further and explicitly exploits the spatial discontinuity in warfare and insecurity created by the River Tagus from the 11th to 13th centuries. As described above, the Tagus was a meaningful military barrier and a recognized frontier landmark during this period of military instability, but arguably it is not a major obstacle to social and economic interactions. Therefore, this border can be considered exogenous and with a relevance circumscribed to the specific historical period in which the territory was conquered and resettled by Castile and Leon.

In this section, I narrow down the sample to grid cells falling within a distance of 50 km from the Tagus. The resulting sample size includes about 220 observations. In choosing the bandwidth, I have tried to reconcile a close geographic proximity with a sufficiently large number of observations to ensure good statistical power. The identification strategy contains the following assumptions: i) the River Tagus was a meaningful military barrier and largely delimited the territory subject to military insecurity; ii) there are no significant differences in climatic, geographic and pre-medieval historical factors across the river; and iii) the Tagus has not been a major barrier to social and economic interactions. If these assumptions are valid and there exists a significant discontinuity in settlement patterns across the Tagus, it would suggest that this discontinuity is due to the way the colonization of an insecure frontier region was achieved.

The validity of assumption “i” is justified above in the historical discussion. Besides, column 1 in Table 4 provides preliminary evidence consistent with the hypothesis about the importance of the Tagus in delimiting the area subject to violence and –therefore– affecting the way the territory was settled. Cells located south of the Tagus have about 20 percentage points less settlement density. Regarding assumption “ii”, its validity can be explicitly tested. Columns 2 to 7 in Table 4 compare the mean values in several climatic and geographic variables across the border. There are no statistically significant differences in rainfall, temperature, elevation,

---

[28] I have checked that the results are very similar when using the alternative indicators of settlement density employed in Section 4.2, namely density of municipalities and density of population entities. These results are available in Tables A5 and A6 in the Supplementary Material.

ruggedness, soil quality, and distance to the coast.[29] In addition, the last two columns show that there are no differences in pre-medieval settlements and distance to Roman roads either, which reduces concerns about potential preexisting historical differences before Christian colonization.[30]

[*Insert Table 4 about here*]

With respect to assumption "iii", the main argument to support its validity is that the Tagus is not wide enough to create a significant barrier to social and economic interactions. In the Middle Ages there were several bridges and in the dry season it was easily fordable. Before the conquest of Toledo in 1085, there were at least five bridges (three Roman and two Muslim), and it was also possible to cross the river through fords at specific points and by boat (Torres Balbás, 1957; Malalana Ureña, 1990). The absence of differences in pre-medieval settlements across the river (Table 4, column 8) is also consistent with this point. It is worth emphasizing that while the Tagus was not a major obstacle to social and economic interactions, it was a crucial barrier from a military perspective, since bridges and fords were strongly defended. Put differently, the same natural obstacle was relatively minor for civil purposes but important from a military perspective.

Table 5 reports the baseline SRD results from equations taking the following form:

$$Y_{i,j} = \alpha_0 + \emptyset_j + \beta \cdot T_{i,j} + X'_{i,j} \cdot \delta + f(geo.loc.) + \varepsilon_{i,j}$$

where $Y_{i,j}$ is the dependent variable (i.e., settlement density) in cell i along segment j of the border, α_0 is a constant term, $\emptyset_j$ is a set of four equal-length segments of the border –representing the closest one to the cell centroid–, $T_{i,j}$ is a dummy variable indicating whether the cell is located south of the River Tagus, $X'_{i,j}$ represents a vector of control variables, $f(geo.loc.)$ stands for a polynomial of variables referred to the geographic location of cell i, and $\varepsilon_{i,j}$ is the error term. The equation is estimated with OLS, reporting standard errors corrected for spatial dependence.[31]

For the sake of robustness, I employ three different indicators of settlement density as dependent variables. Panel A uses the variable of settlement density employed so far, that is, SD_i^{10}(from GEOSTAT 2011 dataset). Panel B uses a variable measuring the density of population entities, which is created from a comprehensive dataset of all population entities (i.e., villages,

---

[29] Table A7 in Supplementary Material also show that there are no differences in aridity nor in important characteristics of the soil for agriculture, such as subsoil and topsoil available water capacity, depth to rock, soil erodibility, topsoil organic carbon content, and soil texture.

[30] Consistent with the latter, historical references also indicate that settlement density before the Christian conquest was not higher north of the Tagus. For instance, Sánchez-Albornoz (1962) maintained that the Duero Valley, particularly north of the river, was completely unpopulated before the Christian conquest and colonization. González (1976a) mentions that when Alfonso VI conquered Toledo, most of the area between the Duero and Tagus was unpopulated. Consequently, areas north of the Tagus did not start with an advantage in terms of settlement density.

[31] A more detailed discussion of the methodology and implementation of SRD design to historical settings can be found in Dell (2010), Becker *et al.* (2016), and Oto-Peralías and Romero-Ávila (2017).

towns, etc.) existing in the country (Instituto Geográfico Nacional, 2016). Finally, Panel C uses an indicator capturing the size of municipalities' jurisdictional areas, with a higher density of municipalities implying a lower size (Eurostat 2016b). This variable is related to the management of the territory, and reflects its division into local governmental areas. Given the persistence in jurisdictional boundaries over time, it is a good proxy for the historical management of the territory, which is linked to how it was colonized.

Column 1 shows the results from a border specification, that is, without including the term $f(geo.loc.)$. In this regression, the coefficient of interest –β– reports the average difference in settlement density on both sides of the border. This is an informative test given that the bandwidth is only 100 km and Table 4 rules out differences in geographic and climatic variables. The results are highly supportive of a statistically significant discontinuity in settlement patterns across the border. The territory south of the Tagus –and therefore affected by high military instability during the 11th-13th centuries– has a much lower settlement density, the difference being almost 26 percentage points in Panel A. This difference is similar in magnitude to that reported in Table 3, which is remarkable given the geographical proximity among observations. Panels B and C also reveal large differences in settlement patterns across the border.

[*Insert Table 5 about here*]

The rest of the columns in Table 5 estimate spatial discontinuity regressions aimed at identifying discontinuous jumps at the border in the dependent variable. Columns 2 and 3 use the quadratic polynomial in distance to the Tagus and to Madrid as forcing variables, while column 4 uses the quadratic polynomial in latitude and longitude. I prefer to avoid high-order polynomials because new evidence indicates that low-order polynomials perform better than their high-order counterparts, which often provide misleading confidence intervals (Gelman and Imbens, 2014). The results are remarkably robust. The effect of having been affected by high insecurity is always negative, statistically significant, and economically important.

I next provide some graphical evidence on the discontinuity at the border. Figure 5 plots the predicted value from a regression of settlement density on distance to the border along with the 90 percent confidence intervals. The existence of a discontinuous jump at the border is apparent in the three variables of settlement density. Figure 6 follows the two-dimensional RD style of plots proposed by Dell (2010), in which the color of the figure represents the predicted value of settlement density from a regression on a quadratic polynomial in latitude and longitude and the treatment dummy variable. This style of RD figure is a good complement to represent spatial discontinuities, particularly in the case of long borders with a relatively short bandwidth. Consistent with the results reported in column 4, the figure reflects a clear discontinuity at the border.

[*Insert Figures 5 and 6 about here*]

4.3. SRD design: Sensitivity analysis

Table 6 conducts specification tests by using alternative polynomials in the variables of geographic location: i) a linear polynomial in distance to the Tagus, ii) an interacted polynomial between distance to the Tagus and $T_{i,j}$, iii) a linear polynomial in distance to Madrid, iv) an interacted polynomial between distance to Madrid and $T_{i,j}$, and v) a linear polynomial in latitude and longitude. In all cases, the coefficients on the variable of interest are negative and highly statistically significant, thereby providing support to the previous findings.

[*Insert Table 6 about here*]

Table 7 provides the results from a falsification test consisting of moving the frontier 50 km northward and southward. The purpose of this exercise is to double-check that the treatment variable is not capturing a north-south gradient in settlement patterns. Reassuringly, the coefficients are mostly insignificant and their magnitude is much smaller. Table 8 reports the results from another placebo exercise testing differences across the Duero and Guadiana rivers, which have similar courses to the Tagus –north and south of it, respectively. According to this paper's argument, the Tagus created a discontinuity due to the coincidence of the progress of the Reconquest over its course with the Almoravid and Almohad invasions. Therefore, one should not find similar discontinuities in settlement patterns across the other rivers, and this is indeed what Table 8 shows.

[*Insert Table 7 and Table 8 about here*]

I also conduct a more systematic placebo test, which involves drawing 1,000 random placebo borders. More precisely, placebo borders follow latitude lines between 37 and 42 degrees north, trying to replicate the roughly horizontal orientation of the Tagus. Observations falling to the south of the random borders are considered "treated". As in the main analysis, the sample is restricted to cells whose centroids are located within 50 km of the border. Figure A4 (Supplementary Material) illustrates the cumulative distribution of coefficients of the placebo treatments, where the vertical line shows the "true" coefficients reported in Table 5. The results from this falsification exercise again provide support for the existence of a genuine discontinuity at the Tagus border. Thus, in less than 1% of cases the placebo treatment has a larger *negative* coefficient than their counterparts in Table 5; in other words, the placebo effect is very seldom greater than the "true" effect.

Finally, I also conduct the following robustness checks: i) to employ a larger set of geographic controls; ii) to exclude grid cells with centroids located in the province of Madrid; iii) to use alternative bandwidths of 50, 75, 125, 150, 175, and 200 km (note that the baseline bandwidth is 100 km), iv) to remove cells that are not completely squared, and v) to use an alternative larger cell size of 500 km^2. The results are very similar to the baseline findings, in that the coefficient on the variable 'south of the Tagus' is always negative and statistically significant.[32]

---

[32] These results are available in the Supplementary Material (Tables A8 to A17).

5. Further evidence and discussion

5.1. A confirmatory analysis: Portugal

The validity of this paper's hypothesis and the general implications of the analysis would be stronger if similar results were found in analogous historical contexts. Portugal is undoubtedly the country with the most similar historical experience to medieval Spain, and therefore the best candidate on which to conduct a "confirmatory analysis". The River Tagus also divides Portugal into two parts, and this kingdom also suffered the invasions of the Almoravid and Almohad armies. The weakening of the Almoravid power allowed Portuguese armies to retake the frontier stronghold of Santarém and conquer Lisbon in 1147, establishing their control over the Tagus (Lay, 2009). In the following decades, the Portuguese expanded southward as far as Silves in the Algarve, but their conquests were fragile since almost all the territory south of the Tagus was lost against the Almohads. As in the case of Castile and Leon, the control south of the river was not possible until the Christian victory at Las Navas de Tolosa in 1212.

Table 9 analyzes the existence of a discontinuity in settlement density across the Tagus in Portugal. I focus on the indicators SD^{10} and density of municipalities, which are the variables constructed using Eurostat data. Figure 7 represents the geographic area of study along with the value of the indicator SD^{10}. As in the main analysis, the sample is restricted to grid cells whose centroids are within 50 km of the Tagus. Both the results reported in the table and the figure itself show that the density of settlement is lower south of the Tagus. Besides the four baseline specifications, I include another with the quadratic polynomial in distance to Lisbon, which is much more relevant to the case of Portugal. The coefficients are always statistically significant and their size similar to the baseline results in Table 5.

[Insert Figure 7 and Table 9 about here]

5.2. Historical measures of settlement density

The evidence presented in Section 4 indicates the presence of a discontinuity in settlement patterns across the Tagus. My interpretation of the results, based on the historical account in Section 3, is that it was the consequence of the high military insecurity experienced during the Middle Ages, which affected both sides of the river differently. If this interpretation is correct, then the discontinuity across the Tagus should exist not only today, but also in the past, just after the territory was colonized. To shed some light on this, I collect census data on population entities from the *Censo de Pecheros de Carlos I* of 1528 (INE, 2008), and from the *Censo de Floridablanca* of 1787 (INE, 1987). These data sources include all the population entities of the country at the time.[33] As information is available about the modern-day municipality to which each 16th or 18th-century settlement belongs, data can be georeferenced. These indicators of the

---

[33] The 1528 census only covers 16th-century Castile, excluding the Kingdom of Granada, the Basque Country and Navarra, but given the restriction of the sample to 50 km from the Tagus, this does not affect the analysis.

16th and 18th centuries are convenient because they are not affected by the profound developments in urbanization that came about with the industrial revolution, thereby reducing potential confounding factors and providing credibility to this paper's hypothesis.[34] Table 10 shows that there was also a discontinuity in settlement density across the Tagus in these early periods, reflecting that it was the result of something happening during the Middle Ages and thereby supporting my reading of the results.

[*Insert Table 10 about here*]

5.3. Contemporary outcomes

Thus far the analysis has focused on the medieval frontier origins of Spain's settlement patterns. In this section, I briefly discuss whether settlement patterns matter for economic development. There are several reasons why this should be the case. From a historical perspective, the way in which the territory is occupied has important implications when land is the main factor of production. In pre-industrial times, when agriculture was the main source of wealth, a balanced occupation of the territory with many settlements scattered across the space was necessary for intensive use of land. Before the age of motor vehicles, geographical proximity made a difference in transportation costs and allowed more intensive forms of land exploitation (e.g., agriculture rather than livestock). Thus, areas with higher settlement density may have developed more intensive forms of land exploitation, thereby becoming wealthier. Given path dependency in prosperity (e.g., Guiso *et al.*, 2016, and Maloney and Valencia, 2016), early economic development may lead to better economic outcomes today.

Although it is beyond the scope of this paper to investigate this issue, Table 11 provides some preliminary evidence. Before reporting the results, it is worth noting that identifying the effect of settlement patterns is a very difficult task. They are the result of historical processes of colonization, and other factors can also be considered co-original. My aim here is to show that the reported discontinuity in settlement patterns overlaps with discontinuities in several current outcome variables. I use as dependent variables the following four proxies for economic development at the local level: light density at night, average socioeconomic condition, number of vehicles per household, and labor force activity rate. Interestingly, Table 11 shows that a similar discontinuity arises when using these outcome variables. Given the absence of geographic differences across the border and the fact that the River Tagus hardly implies a major geographic barrier to economic interactions, it is plausible to assume that these differences in outcomes are the consequence of the way the territory was colonized in the Middle Ages. Low settlement density along with other related factors –such as political and economic inequality and the prevalence of livestock– could have created the conditions for slower economic growth in the long-run.[35]

---

[34] Note that the indicator of settlement density from the 1528 census is not even affected by the transformations that took place during the Modern era.

[35] I have checked that the results also hold when using the approach employed in Section 4.1, that is, with

[*Insert Table 11 about here*]

5.4. Discussion

The findings of this section provide support for the frontier origins hypothesis of Spain's settlement patterns. According to this hypothesis, the high level of military insecurity suffered by the large frontier region of the southern plateau from the 11th to 13th centuries conditioned the colonization of the territory by Castile and Leon. The main features of the initial colonization strategy were intense militarization, the existence of a few well-defended settlements and castles, and a ranching-oriented economy. This extreme form of initial colonization sowed the seeds for the sparse settlement density of the territory to this day. There were a number of interrelated factors that contributed to the persistence in low settlement density, some of them the direct consequence of the colonization strategy itself.

First, the prominent role of the military orders as colonizer agents did not contribute to the attraction of settlers. When establishing new settlements, the military orders kept for themselves important seigniorial privileges, which made these places less attractive to settlers than other available territories (López Pita, 1994). Second, the ranching orientation of the economy could have perverse effects on population patterns. On the one hand, it is a system that works well in a context of low population density, thereby reducing the necessity of attracting farmers to exploit the land. On the other hand, ranching can create vested interests, which contribute to its perpetuation by –for example– maintaining the use of land for pasture. Third, the fact that the frontier experienced a large southward expansion just at the time when the Muslim threat disappeared slowed down the resettlement process even further. This is important because it meant that the ranching character of the economy and the initial pattern of colonization had more time to become rigid.

Fourth, geographic conditions may also matter. The relative aridity of the territory may help to explain why the region failed to converge on settlement density to the rest of Spain and Europe. In a context of a dry environment, the initial form of colonization may be more persistent. Yet it is worth emphasizing that the previous analysis shows that geography does not explain the discontinuity across the River Tagus. In addition, most of Spain is dry and the southern plateau is not the driest part of the country. It was the extreme medieval frontier insecurity that created a pattern which has persisted over time, and in this process of path dependence, geography may have played a role, but not the main one. Institutional factors such as the prominence of military orders and the early ranching orientation of the economy were more crucial. That geography played a secondary role is also supported by the fact that the difference in settlement density between the territory conquered by Castile and Leon in the 12th-13th centuries and the rest of Spain remains largely unchanged when including geo-climatic controls (as shown in Table 3). Moreover, the difference is similar in magnitude even when focusing on the territory within 50 km of the

---

the whole sample of Spain and the inclusion of virtual region dummies (see Table A19).

Tagus (Table 5).

6. Conclusions

This article shows that historical frontiers can shape the economic geography of countries. I focus on Medieval Spain to explore how extreme frontier insecurity can condition the occupation of the territory in such a way to make it one of the most desert areas in Europe. First, I provide evidence that Spain stands out with an anomalous settlement pattern, characterized by a very low density in its southern half, which is not explained by geographic and climatic factors. The result is robust at the NUTS 3 and the grid cell levels, and to the use of virtual regions. The second part of the article investigates the medieval frontier origins of the spatial distribution of settlements in Spain. The southern plateau was subject to extreme warfare and insecurity during the High Middle Ages, which led to a colonization of the territory characterized by a very small number of settlements, the prominence of military orders, and a ranching orientation of the economy. Standard regression analysis and spatial regression discontinuity techniques provide support for the medieval frontier origins hypothesis. I further show that differences in settlement patterns overlap with differences in several indicators of current economic development at the local level, which suggests that the Spanish anomaly in settlement patterns may be related to southern Spain's economic backwardness.

This article is linked to several strands of the economic literature. First, it contributes to the new empirical literature on historical frontier societies by investigating the impact of medieval frontiers on the social occupation of the space (García-Jimeno and Robinson, 2011; Oto-Peralías and Romero-Ávila, 2016, 2017; Droller, 2017). Also discussed are a number of institutional features associated with the initial colonization strategy that can persist over time, perpetuating a pattern of low settlement density. Second, this paper is also related to a growing literature on the legacy of military conflicts on urban growth and development. In contrast to previous works finding long-term positive effects of conflicts (e.g., Voigtländer and Voth, 2013; Dincecco and Onorato, 2016), the analysis shows that military insecurity may have negative implications. Arguably, the consequences of military conflicts depend on the context (Gennaioli and Voth, 2015; Dincecco *et al*. 2016). In consolidated states or kingdoms, such conflict may foster urban growth and fiscal capacity, but in frontier regions, it may lead to negative outcomes. Third, this work also adds to a body of research on the contingent role that geographic factors play in development (e.g., Dell, 2012; Nunn and Puga, 2012). The River Tagus created a discontinuity in settlement patterns due to a contingent factor, namely high military insecurity, which in turn was the consequence of "historical accidents" such as the Almoravid and Almohad invasions.

Finally, this article's results relating to the importance of historical events in the settlement structure of the territory –and the potentially high-level persistence of this– also contribute to the debates regarding the role played by location fundamentals in the location of cities and to the literature on the persistence and dynamics of the urban system (Bleakley and Lin, 2012; Michaels and Rauch, 2017; Bosker and Buringh, 2017). While there is a large tradition of research on urbanization processes, little work has been done on how regions and countries are actually settled, which can be viewed as the most basic layer of interaction between economic agents and the territory. The implications of settlement patterns are also worth studying. The importance of settlement density for agriculture before the age of mechanization has been noted. It may also stimulate communications and trade, with long-term positive effects. In addition, settlement patterns may have effects in the short-term, notably in the labor market. For a given population density, higher settlement density may facilitate interactions and increase labor mobility. These and related topics are interesting areas of research.

References

Álvarez-Nogal, C., and De La Escosura, L. P. 2013. "The rise and fall of Spain (1270 ‐ 1850)," *The Economic History Review* 66(1): 1-37.

Ayala Martínez, Carlos de. 1996. "Órdenes militares hispánicas: reglas y expansión geográfica" *Codex aquilarensis: Cuadernos de investigación del Monasterio de Santa María la Real,* 12: 57-86.

Ayala Martínez, Carlos de. 2006. "Frontera y órdenes militares en la Edad Media castellano-leonesa (siglos XII-XIII)." *Studia Historica, Historia Medieval* 24: 87-112.

Becker Sascha O., Boeckh Katrin, Hainz Christa, Woessmann Ludger. 2016. "The Empire Is Dead, Long Live the Empire! Long-Run Persistence of Trust and Corruption in the Bureaucracy." *Economic Journal,* 126(590), 40–74.

Belloc, Marianna, Francesco Drago, and Roberto Galbiati. 2017. "Earthquakes, Religion, and Transition to Self-Government in Italian Cities," *The Quarterly Journal of Economics*, forthcoming doi: 10.1093/qje/qjw020.

Berend, Nora. 1999. "Medievalists and the Notion of the Frontier," *The Medieval History Journal* 2 (1): 55-72.

Bishko, Charles Julian. 1952. "The peninsular background of Latin American cattle ranching," *The Hispanic American Historical Review* 32 (4): 491-515.

Bishko, Charles Julian. 1963. "The Castilian as plainsman: The medieval ranching frontier in La Mancha and Extremadura." In: *The New World looks at its history*, edited by Archibald R. Lewis and Thomas F. McGann, pages 47-69.

Bleakley, Hoyt, and Jeffrey Lin. 2012. "Portage and path dependence." *The Quarterly Journal of Economics* 127(2): 587-644.

Brenan, Geral. 1950. *The Spanish Labyrinth: An Account of the Social and Political Background of the Spanish Civil War.* Cambridge University Press, Cambridge.

Burke, Marshall, Solomon M. Hsiang, and Edward Miguel. 2015. "Global non-linear effect of temperature on economic production," *Nature* 527 (7577): 235-239.

Conley, Timothy. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics* 92: 1–45.

Dell, Melissa. 2010. "The Persistent Effects of Peru's Mining Mita." *Econometrica* 78, 1863–1903.

Dell, Melissa. 2012. "Path dependence in development: Evidence from the Mexican Revolution," Working Paper, University of Harvard.

Dincecco, Mark, James E. Fenske, and Massimiliano G. Onorato. 2016. "Is Africa different? Historical conflict and state development." Working paper.

Dincecco, Mark, and Massimiliano G. Onorato. 2016. "Military conflict and the rise of urban Europe," *Journal of Economic Growth* 21: 259.

Droller, Federico. 2017. "Migration, Population Composition, and Long Run Economic Development: Evidence from Settlements in the Pampas," *The Economic Journal*, forthcoming, doi: 10.1111/ecoj.12505.

Eurostat. 2016a. GEOSTAT 2011 population grid. Available at http://ec.europa.eu/eurostat/

Eurostat. 2016b. Communes 2013. Available at http://ec.europa.eu/eurostat/

Fischer, G., Nachtergaele, F., Prieler, S., van Velthuizen, H.T.,Verelst, L., & Wiberg, D. 2008. *Global agro-ecological zones assessment for agriculture (GAEZ 2008)*. IIASA, Laxenburg, Austria and FAO, Rome, Italy.

Forey, Alan J. 1984. "The Military Orders and the Spanish Reconquest in the Twelfth and Thirteenth Centuries." *Traditio* 40: 197–234.

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

García Fitz, Francisco. 2001. "Una frontera caliente. La guerra en las fronteras castellano-musulmanas (siglos XI-XIII)." In: *Identidad y representación de la frontera en la España medieval (siglos XI XIV)*: 159-180.

García-Jimeno Camilo, Robinson James A. 2011. "The Myth of the Frontier." In *Understanding Long-Run Economic Growth*, edited by Costa Dora L., Lamoreaux Naomi R.. University of Chicago Press, Chicago, pp. 49–88.

Gelman Andrew, and Imbens Guido. 2014. "Why High-order Polynomials Should Not Be Used in Regression Discontinuity Designs." NBER Working Paper 20405. National Bureau of Economic Research, Cambridge, Massachusetts.

Gennaioli, Nicola, and Hans-Joachim Voth. 2015. "State capacity and military conflict." *The Review of Economic Studies* 82, no. 4: 1409-1448.

González, Julio. 1976a. *Repoblación de Castilla la Nueva*. Vol. 2. Universidad Complutense, Facultad de Filosofía y Letras, 1976.

González, Julio. 1976b. *Repoblación de Castilla la Nueva*. Vol. 1. Universidad Complutense, Facultad de Filosofía y Letras, 1976.

González Jiménez, Manuel. 1992. "Frontier and Settlement in the Kingdom of Castile (1085–1350)." In *Medieval Frontier Societies,* edited by Bartlett Robert, MacKay Angus. Clarendon Press, Oxford, pp. 49–74.

Grosjean, Pauline. 2011. "The Institutional Legacy of the Ottoman Empire: Islamic Rule and Financial Development in South Eastern Europe." *Journal of Comparative Economics* 39, 1–16.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2016. "Long ‐ term persistence." *Journal of the European Economic Association* 14 (6): 1401-1436.

Henderson, J. Vernon, Zmarak Shalizi, and Anthony J. Venables. 2001. "Geography and development," *Journal of Economic Geography* 1(1): 81-105.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. 2005. "Very high resolution interpolated climate surfaces for global land areas." *International Journal of Climatology*, 25, 1965–1978.

Instituto Geográfico Nacional. 2016. *Nomenclátor Geográfico de Municipios y Entidades de Población*. Instituto Geográfico Nacional, Ministerio de Fomento, Gobierno de España, Madrid. Available at http://www.ign.es.

INE. 1987. *Censo de 1787 "Floridablanca"*. Madrid: Instituto Nacional de Estadística.

INE. 2008. *Censo de Pecheros. Carlos I 1528*. Madrid: Instituto Nacional de Estadística.

Ladero Quesada, Miguel Á. 1984. "Toledo en época de la frontera," *Anales de la Universidad de Alicante. Historia Medieval* 3: 71-98.

Lay, Stephen. 2009. *The Reconquest Kings of Portugal. Political and Cultural Reorientation on the Medieval Frontier*. Palgrave Macmillan.

Liedekerke Marc van, Jones Arwyn, Panagos Panos. 2006. "ESDBv2 Raster Library—A Set of Rasters Derived from the European Soil Database Distribution v2.0." European Commission and the European Soil Bureau Network, CD-ROM.

López-González, C., Castellano, E. P., and Rodríguez, J. I. R. 1989. "Las Ordenes Militares Castellanas en la Época Moderna: Una Aproximación Cartográfica," In M. E. Burgos & D. Ozanam (Eds.), *Las Ordenes Militares en el Mediterráneo Occidental*. Siglos XIII–XVIII (pp. 291–303). Toledo: Casa de Velázquez, Instituto de Estudios Manchegos. Imprenta Taravilla.

López Pita, Paulina. 1994. "La sociedad manchega en vísperas del Descubrimiento," *Espacio, Tiempo y Forma, Serie IV, Historia Moderna*, t. 7, 349-366.

MacKay, Angus. 1977. *Spain in the Middle Ages: From Frontier to Empire, 1000-1500.*

MacMillan Publishing Company.

Malalana Ureña, Antonio. 1990. “Puentes-fortaleza en el Tajo: El tramo Zorita de los Canes (Guadalajara)-Castros (Cáceres),” *Boletín de arqueología medieval* 4: 195-222.

Mallada, Lucas. 1969. *Los Males de la Patria*, (1st ed: 1890), Madrid.

Maloney, William F., and Felipe Valencia Caicedo. 2016. “The Persistence of (Subnational) Fortune,” *The Economic Journal* 126, 2363–2401.

McCormick, Michael, Guoping Huang, Giovanni Zambotti, and Jessica Lavash. 2013. “Roman Road Network (version 2008),” DARMC Scholarly Data Series, Data Contribution Series #2013-5. DARMC, Center for Geographic Analysis, Harvard University, Cambridge MA 02138.

Michaels, Guy, and Ferdinand Rauch. 2017. “Resetting the Urban Network: 117 ‐ 2012.” *The Economic Journal*, doi:10.1111/ecoj.12424.

Nunn, Nathan, and Diego Puga. 2012. “Ruggedness: The Blessing of Bad Geography in Africa” *Review of Economics and Statistics* 94: 20–36.

Oto-Peralías, Daniel, and Diego Romero-Ávila. 2016. “The Economic Consequences of the Spanish Reconquest: The Long-term Effects of Medieval Conquest and Colonization.” *Journal of Economic Growth* 21: 409–464.

Oto-Peralías, Daniel, and Diego Romero-Ávila. 2017. “Historical Frontiers and the Rise of Inequality: The Case of the Frontier of Granada,” *Journal of the European Economic Association* 15 (1): 54-98.

Ottaviano, Gianmarco, and Jacques-François Thisse. 2004. “Agglomeration and economic geography.” *Handbook of regional and urban economics* 4: 2563-2608.

Panagos Panos. 2006. “The European Soil Database.” *GEO: Connexion*, 5, 32–33.

Panagos Panos, van Liedekerke Marc, Jones Arwyn, Montanarella Luca. 2012. “European Soil Data Centre: Response to European Policy Support and Public Data Requirements.*” Land Use Policy*, 29, 329–338.

Pleiades. 2016. Ancient world mapping center and Institute for the Study of the Ancient World. Institute for the Study of the Ancient World, New York University. Accessed November 28, 2016, from http://pleiades.stoa.org/.

Reher, David-Sven. 1990. *Town and Country in Pre-industrial Spain*. Cambridge: Cambridge University Press.

Rodríguez-Picavea, Enrique. 1999. “Aproximación a la geografía de la frontera meridional del reino de castilla (1157-1212),” *Cuadernos de Historia Medieval* Secc. Miscelánea, 2: 29-42.

Sánchez Albornoz, Claudio. 1962. *España, un enigma histórico*. Sudamericana: Buenos Aires.

Simpson, James. 1995. *Spanish agriculture: the long Siesta, 1765-1965*. Cambridge University

Press.

Torres Balbás, Leopoldo. 1957. *Ciudades Yermas Hispanomusulmanas*. Madrid: Imprenta Y Editorial Maestre.

Trabucco Antonio, Zomer Robert J. 2009. "Global Aridity Index (Global-Aridity) and Global Potential Evapo-Transpiration (Global-PET) Geospatial Database." CGIAR Consortium for Spatial Information . Available from the CGIAR-CSI GeoPortal at http://www.csi.cgiar.org.

Turner, Frederick J. 1920. *The Frontier in American History*. H. Holt and Co., New York.

Vann, Theresa M. 1999. "Twelfth-Century Castile and its frontier strategies." *The Circle of War in the Middle Ages: Essays on Medieval Military and Naval History*, 6: 21.

Voigtländer, Nico and Hans-Joachim Voth. 2013. "Gifts of Mars: Warfare and Europe's Early Rise to Riches," *Journal of Economic Perspectives* 27: 165-86.

TABLES AND FIGURES

Table 1

Determinants of settlement patterns in Europe and the Spanish anomaly: NUTS level analysis

Dependent variable is settlement density (SD^{10})

	(1)	(2)	(3)	(4)		(5)		(6)		(7)
Panel A: Geographic and climatic variables										
Temperature	12.483***	11.923***	11.11***	10.616***				11.155***		11.306***
	(1.233)	(1.482)	(1.323)	(1.801)				(1.392)		(1.434)
Temperat. sq	-0.691***	-0.663***	-0.613***	-0.585***				-0.441***		-0.432***
	(0.099)	(0.096)	(0.084)	(0.104)				(0.076)		(0.067)
Precipitation		7.131	8.537*	9.497**				5.165		4.936
		(5.929)	(4.595)	(4.435)				(3.303)		(3.208)
Precipit. sq		-0.392	-0.4**	-0.425**				-0.259*		-0.253*
		(0.266)	(0.194)	(0.192)				(0.137)		(0.135)
Altitude			-0.029***	-0.017				-0.006		-0.001
			(0.008)	(0.014)				(0.006)		(0.006)
Ruggedness				-0.032						-0.007
				(0.021)						(0.008)
Soil quality				0.108						-0.654
				(1.662)						(1.383)
Dist to coast				-0.004						-0.029
				(0.027)						(0.026)
Dist to coast sq				0.000						0.000
				(0.00)						(0.00)
Island dummy				14.854*						-8.64
				(8.244)						(5.759)
Panel B: Regions/country dummies: The five largest negative coefficients										
					Iceland	-87.847***	South Sp.	-58.407***	South Sp.	-61.391***
						(0.907)		(8.258)		(8.746)
					South Sp.	-64.965***	Macedonia	-45.941***	Macedonia	-46.705***
						(2.99)		(4.404)		(4.174)
					Norway	-57.768***	North Sp.	-36.151***	North Sp.	-38.103***
						(4.626)		(4.37)		(4.631)
					Macedonia	-52.337***	Iceland	-32.67***	Greece	-31.956***
						(3.007)		(10.025)		(6.4)
					Sweden	-43.962***	Greece	-30.178***	Iceland	-30.931***
						(9.675)		(5.298)		(10.229)
Constant	30.52***	3.749	6.783	2.912		94.796***		4.855		12.016
	(5.797)	(26.256)	(18.618)	(24.999)		(0.868)		(21.092)		(22.959)
R-sq	0.52	0.55	0.65	0.66		0.67		0.87		0.87
Obs	1,440	1,440	1,440	1,418		1,440		1,440		1,418

Notes: Variables descriptions are provided in Table A1. Columns 5 to 7 include a set of country dummies, with France being the omitted category. For space considerations, only the five largest negative cofficients are reported. Regressions are weighted by NUTS surface area. Standard errors clustered at the country level are in parentheses (in column 5 robust standard errors are reported). *, ** and *** denote significance at the 10, 5 and 1% level, respectively.

Table 2

Determinants of settlement patterns in Europe and the Spanish anomaly: Pixel level analysis

Dependent variable is settlement density (SD^{10})

	(1)	(2)	(3)
Panel A: Geographic and climatic variables			
Temperature		10.001***	10.135***
		(1.359)	(1.423)
Temperat. sq		-0.37***	-0.367***
		(0.059)	(0.062)
Precipitation		3.658	3.852
		(2.264)	(2.294)
Precipit. sq		-0.164*	-0.174**
		(0.083)	(0.083)
Altitude		-0.011*	-0.009
		(0.006)	(0.007)
Ruggedness			-0.005
			(0.008)
Soil quality			-0.034
			(0.698)
Dist to coast			-0.008
			(0.026)
Dist to coast sq			0.000
			(0.00)
Island dummy			-7.195
			(4.402)

Panel B: Regions/country dummies: The five largest negative coefficients

(1)		(2)		(3)	
Iceland	-84.868***	Sp.-Anom.	-65.689***	Sp.-Anom.	-68.127***
	(0.86)		(7.13)		(7.232)
Sp.-Anom.	-69.178***	Macedonia	-40.572***	Macedonia	-41.553***
	(0.771)		(4.155)		(3.734)
Norway	-53.872***	Cyprus	-34.979***	Sp.-Rest	-34.847***
	(0.898)		(8.637)		(4.674)
Macedonia	-50.524***	Iceland	-34.852***	Iceland	-32.943***
	(2.585)		(10.234)		(10.393)
Sweden	-41.731***	Sp.-Rest	-33.085***	Greece	-29.695***
	(0.908)		(4.65)		(4.738)

	(1)	(2)	(3)
Constant	93.836***	16.522	14.773
	(0.322)	(14.535)	(18.02)
R-sq	0.43	0.67	0.72
Obs	23,498	23,441	21,330

Notes: Variables descriptions are provided in Table A1. Regressions include a set of country dummies, with France being the omitted category. 'Sp.-Anom.' refers to the territory corresponding to the Spanish anomaly, while 'Sp.-Rest' to the rest of Spain. For space considerations, only the five largest negative cofficients are reported. Standard errors clustered at the country level are in parentheses (in column 1 robust standard errors are reported). *, ** and *** denote significance at the 10, 5 and 1% level, respectively.

Table 3

Frontiers and settlement patterns: Initial results

Dependent variable is settlement density (SD^{10})

	(1)	(2)	(3)	(4)	(5)
Treated region (conquered by	-32.322***	-33.283***	-25.96***	-26.815***	-21.658***
Leon and Castile 12th-13th cent.)	(3.01)	(3.188)	(3.353)	(3.879)	(3.869)
Temperature		7.411***	7.563***	6.215**	6.15**
		(2.842)	(2.304)	(2.864)	(2.662)
Temperat. sq		-0.232***	-0.299***	0.024	-0.04
		(0.088)	(0.072)	(0.092)	(0.082)
Precipitation		2.854	3.859	11.537***	11.705***
		(3.265)	(2.983)	(3.523)	(3.099)
Precipit. sq		0.06	-0.061	-0.418**	-0.511***
		(0.178)	(0.153)	(0.18)	(0.155)
Altitude		-0.017*	-0.008	0.000	0.006
		(0.009)	(0.008)	(0.013)	(0.011)
Ruggedness			-0.048***		-0.034***
			(0.013)		(0.011)
Soil quality			-0.779		0.322
			(0.659)		(0.507)
Dist to coast			-0.335***		-0.309***
			(0.052)		(0.05)
Dist to coast sq			0.001***		0.001***
			(0.00)		(0.00)
Island dummy			-4.879		-8.489
			(5.141)		(5.431)
Virtual regions fixed-effects				Yes	Yes
R-sq	0.28	0.47	0.54	0.56	0.61
Obs	2,091	2,091	2,000	2,091	2,000

Notes: Variables descriptions are provided in Table A1. Regressions include a constant term which is omitted for space considerations. Standard errors corrected for spatial dependence are in parentheses. *, ** and *** denote significance at the 10, 5 and 1% level, respectively.

Table 4

Differences across the Tagus River: Settlement density and geo-climatic variables

	Settlement density (SD^{10})	*Precipitation*	*Temperature*	*Altitude*	*Ruggedness*	*Soil quality*	*Distance to the coast*	*Pre-medieval settlements*	*Distance to Roman roads*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
South of the Tagus (high military insecurity)	-19.82***	0.301	0.967	-65.251	-18.818	-0.075	-13.67	-0.007	4.241
	(5.207)	(0.255)	(0.675)	(93.1)	(17.915)	(0.291)	(19.495)	(0.006)	(5.569)
R-sq	0.24	0.03	0.05	0.01	0.02	0.00	0.01	0.00	0.01
Obs	219	219	219	219	219	217	219	218	219

Notes: Variables descriptions are provided in Table A1. South of the Tagus (high military insecurity) is a dummy variable indicating whether the grid cell is located south of the Tagus. Sample restricted to grid cells within 50 km of the border. Regressions include a constant term which is omitted for space considerations. Standard errors corrected for spatial dependence are in parentheses. *, ** and *** denote significance at the 10, 5 and 1% level, respectively.

Table 5

Baseline results: Border specification and spatial RD regressions

	OLS (1)	Quadratic polynomial in distance to the Tagus (2)	Quadratic polynomial in distance to Madrid (3)	Quadratic polynomial in latitude and longitude (4)
Panel A: Settlement density (SD^{10})				
South of the Tagus (high military insecurity)	-25.738*** (6.001)	-25.32*** (5.938)	-16.069*** (3.671)	-20.32*** (3.632)
Standardized coefficient	-0.638	-0.627	-0.398	-0.503
R-squared	0.42	0.43	0.54	0.52
Observations	219	219	219	219
Panel B: Density of 'population entities'				
South of the Tagus (high military insecurity)	-0.308*** (0.063)	-0.31*** (0.064)	-0.191*** (0.045)	-0.209*** (0.044)
Standardized coefficient	-0.575	-0.577	-0.356	-0.389
R-squared	0.50	0.50	0.57	0.58
Observations	218	218	218	218
Panel C: Density of municipalities				
South of the Tagus (high military insecurity)	-0.894*** (0.166)	-0.916*** (0.171)	-0.601*** (0.173)	-0.66*** (0.151)
Standardized coefficient	-0.441	-0.451	-0.296	-0.325
R-squared	0.40	0.41	0.44	0.46
Observations	219	219	219	219
Boundary fixed effects	Yes	Yes	Yes	Yes
Geo-climatic controls	Yes	Yes	Yes	Yes

Notes: Variables descriptions are provided in Table A1. South of the Tagus (high military insecurity) is a dummy variable indicating whether the grid cell is located south of the Tagus. Regressions include a constant term which is omitted for space considerations. Sample restricted to grid cells within 50 km of the border. The set of geographic–climatic controls includes rainfall and temperature (both in linear and quadratic terms) and altitude. Standard errors corrected for spatial dependence are in parentheses. *, ** and *** denote significance at the 10, 5 and 1% level, respectively.

Table 6

Specification tests

	Linear polynomial in distance to the Tagus (1)	Interacted polynomial in distance to the Tagus (2)	Linear polynomial in distance to Madrid (3)	Interacted polynomial in distance to the Madrid (4)	Linear polynomial in latitude and longitude (5)
Panel A: Settlement density (SD^{10})					
South of the Tagus (high military insecurity)	-25.582***	-19.152***	-18.768***	-41.463***	-19.382***
	(5.959)	(4.748)	(4.092)	(6.696)	(4.753)
Standardized coefficient	-0.634	-0.474	-0.465	-1.027	-0.48
R-squared	0.42	0.43	0.47	0.54	0.45
Observations	219	219	219	219	219
Panel B: Density of 'population entities'					
South of the Tagus (high military insecurity)	-0.311***	-0.236***	-0.205***	-0.42***	-0.194***
	(0.064)	(0.069)	(0.048)	(0.07)	(0.054)
Standardized coefficient	-0.578	-0.44	-0.382	-0.782	-0.362
R-squared	0.50	0.51	0.56	0.59	0.55
Observations	218	218	218	218	218
Panel C: Density of municipalities					
South of the Tagus (high military insecurity)	-0.912***	-0.7***	-0.665***	-1.323***	-0.671***
	(0.17)	(0.227)	(0.168)	(0.22)	(0.194)
Standardized coefficient	-0.45	-0.345	-0.328	-0.652	-0.331
R-squared	0.41	0.42	0.43	0.45	0.42
Observations	219	219	219	219	219
Boundary fixed effects	Yes	Yes	Yes	Yes	Yes
Geo-climatic controls	Yes	Yes	Yes	Yes	Yes

Notes: Variables descriptions are provided in Table A1. South of the Tagus (high military insecurity) is a dummy variable indicating whether the grid cell is located south of the Tagus. Regressions include a constant term which is omitted for space considerations. Sample restricted to grid cells within 50 km of the border. The set of geographic–climatic controls includes rainfall and temperature (both in linear and quadratic terms) and altitude. Standard errors corrected for spatial dependence are in parentheses. *, ** and *** denote significance at the 10, 5 and 1% level, respectively.

Table 7

Placebo tests (I): Moving the frontier southward and northward

	Moving the frontier 50 km northward				Moving the frontier 50 km southward			
	OLS	Quadratic polynomial in distance to the Tagus	Quadratic polynomial in distance to Madrid	Quadratic polynomial in latitude and longitude	OLS	Quadratic polynomial in distance to the Tagus	Quadratic polynomial in distance to Madrid	Quadratic polynomial in latitude and longitude
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Settlement density (SD ^{10})								
South of the placebo border	-5.84*	-5.74*	-3.511	-0.143	3.565	3.348	5.313**	2.486
	(3.175)	(3.228)	(2.723)	(2.762)	(2.677)	(2.669)	(2.469)	(3.215)
Standardized coefficient	-0.145	-0.142	-0.087	-0.004	0.124	0.116	0.185	0.087
R-squared	0.41	0.41	0.43	0.47	0.25	0.26	0.30	0.34
Observations	224	224	224	224	242	242	242	242
Panel B: Density of 'population entities'								
South of the placebo border	-0.05	-0.049	-0.031	0.024	0.052	0.05	0.082**	0.04
	(0.037)	(0.037)	(0.034)	(0.041)	(0.047)	(0.049)	(0.042)	(0.044)
Standardized coefficient	-0.107	-0.106	-0.066	0.052	0.115	0.112	0.182	0.089
R-squared	0.37	0.37	0.41	0.45	0.32	0.33	0.37	0.43
Observations	223	223	223	223	241	241	241	241
Panel C: Density of municipalities								
South of the placebo border	-0.213	-0.197	-0.298	-0.287	-0.021	-0.032	0.035	-0.01
	(0.237)	(0.234)	(0.197)	(0.315)	(0.134)	(0.138)	(0.14)	(0.171)
Standardized coefficient	-0.094	-0.087	-0.132	-0.126	-0.014	-0.022	0.024	-0.007
R-squared	0.32	0.33	0.33	0.33	0.36	0.37	0.37	0.40
Observations	224	224	224	224	242	242	242	242
Boundary fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo-climatic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Variables descriptions are provided in Table A1. South of the placebo border is a dummy variable indicating whether the grid cell is located south of the placebo border. Regressions include a constant term which is omitted for space considerations. Sample restricted to grid cells within 50 km of the border. The set of geographic–climatic controls includes rainfall and temperature (both in linear and quadratic terms) and altitude. Standard errors corrected for spatial dependence are in parentheses. *, ** and *** denote significance at the 10, 5 and 1% level, respectively.

Table 8

Placebo tests (II): Duero and Guadiana rivers

	Testing differences across the Duero river				Testing differences across the Guadiana river			
	OLS	Quadratic polynomial in distance to the Tagus	Quadratic polynomial in distance to Madrid	Quadratic polynomial in latitude and longitude	OLS	Quadratic polynomial in distance to the Tagus	Quadratic polynomial in distance to Madrid	Quadratic polynomial in latitude and longitude
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Settlement density (SD ^{10})								
South of the placebo border	3.193	3.717	3.103	2.67	-1.18	-1.726	-1.633	-6.238*
	(3.28)	(3.061)	(3.212)	(3.81)	(2.67)	(2.709)	(2.99)	(3.682)
Standardized coefficient	0.109	0.127	0.106	0.091	-0.049	-0.072	-0.068	-0.259
R-squared	0.14	0.16	0.15	0.16	0.23	0.23	0.24	0.29
Observations	141	141	141	141	205	205	205	205
Panel B: Density of 'population entities'								
South of the placebo border	0.012	0.019	0.007	-0.04	-0.007	-0.008	0	-0.044
	(0.036)	(0.034)	(0.034)	(0.046)	(0.031)	(0.032)	(0.031)	(0.036)
Standardized coefficient	0.036	0.056	0.022	-0.119	-0.019	-0.022	0.001	-0.115
R-squared	0.19	0.22	0.19	0.24	0.28	0.29	0.30	0.35
Observations	141	141	141	141	205	205	205	205
Panel C: Density of municipalities								
South of the placebo border	-0.021	0.054	-0.113	-0.169	0.091	0.099	0.154	-0.047
	(0.323)	(0.277)	(0.289)	(0.296)	(0.092)	(0.097)	(0.11)	(0.095)
Standardized coefficient	-0.009	0.023	-0.049	-0.073	0.081	0.089	0.137	-0.042
R-squared	0.19	0.27	0.20	0.25	0.22	0.23	0.26	0.32
Observations	141	141	141	141	205	205	205	205
Boundary fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geo-climatic controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Variables descriptions are provided in Table A1. South of the placebo border is a dummy variable indicating whether the grid cell is located south of the placebo border. Regressions include a constant term which is omitted for space considerations. Sample restricted to grid cells within 50 km of the border. The set of geographic–climatic controls includes rainfall and temperature (both in linear and quadratic terms) and altitude. Standard errors corrected for spatial dependence are in parentheses. *, ** and *** denote significance at the 10, 5 and 1% level, respectively.

Table 9

Discontinuity across the Tagus River in Portugal

	OLS	Quadratic polynomial in distance to the Tagus	Quadratic polynomial in distance to Madrid	Quadratic polynomial in distance to Lisbon	Quadratic polynomial in latitude and longitude
	(1)	(2)	(3)	(4)	(5)
Panel A: Settlement density (SD^{10})					
South of the Tagus (high military insecurity)	-19.109***	-19.902***	-19.915***	-20.014***	-18.333***
	(5.774)	(5.057)	(5.92)	(5.758)	(6.4)
Standardized coefficient	-0.542	-0.565	-0.565	-0.568	-0.52
R-squared	0.68	0.71	0.70	0.69	0.73
Observations	67	67	67	67	67
Panel B: Density of 'population entities'					
South of the Tagus (high military insecurity)	-2.313**	-2.328**	-2.098**	-2.154**	-2.218**
	(0.954)	(0.965)	(0.915)	(0.974)	(0.902)
Standardized coefficient	-0.539	-0.542	-0.489	-0.502	-0.517
R-squared	0.59	0.59	0.64	0.61	0.66
Observations	67	67	67	67	67
Boundary fixed effects	Yes	Yes	Yes	Yes	Yes
Geo-climatic controls	Yes	Yes	Yes	Yes	Yes

Notes: Variables descriptions are provided in Table A1. South of the Tagus (high military insecurity) is a dummy variable indicating whether the grid cell is located south of the Tagus. Regressions include a constant term which is omitted for space considerations. Sample restricted to grid cells within 50 km of the border. The set of geographic–climatic controls includes rainfall and temperature (both in linear and quadratic terms) and altitude. Standard errors corrected for spatial dependence are in parentheses. *, ** and *** denote significance at the 10, 5 and 1% level, respectively.

Table 10

Historical variables of settlement density

	OLS	Quadratic polynomial in distance to the Tagus	Quadratic polynomial in distance to Madrid	Quadratic polynomial in latitude and longitude
	(1)	(2)	(3)	(4)
Panel A: Census of "Pecheros" of Carlos I (1528)				
South of the Tagus (high military insecurity)	-0.016***	-0.016***	-0.013***	-0.007*
	(0.004)	(0.004)	(0.004)	(0.004)
Standardized coefficient	-0.453	-0.455	-0.355	-0.20
R-squared	0.45	0.45	0.46	0.49
Observations	211	211	211	211
Panel B: Census of "Floridablanca" (1787)				
South of the Tagus (high military insecurity)	-0.014***	-0.014***	-0.009***	-0.006**
	(0.003)	(0.003)	(0.003)	(0.003)
Standardized coefficient	-0.425	-0.428	-0.27	-0.178
R-squared	0.47	0.47	0.51	0.55
Observations	219	219	219	219
Boundary fixed effects	Yes	Yes	Yes	Yes
Geo-climatic controls	Yes	Yes	Yes	Yes

Notes: Variables descriptions are provided in Table A1. South of the Tagus (high military insecurity) is a dummy variable indicating whether the grid cell is located south of the Tagus. Regressions include a constant term which is omitted for space considerations. Sample restricted to grid cells within 50 km of the border. The set of geographic–climatic controls includes rainfall and temperature (both in linear and quadratic terms) and altitude. Standard errors corrected for spatial dependence are in parentheses. *, ** and *** denote significance at the 10, 5 and 1% level, respectively.

Table 11

Discontinuity in current economic outcomes

	OLS	Quadratic polynomial in distance to the Tagus	Quadratic polynomial in distance to Madrid	Quadratic polynomial in latitude and longitude
	(1)	(2)	(3)	(4)
Panel A: Log light density at night				
South of the Tagus (high military insecurity)	-1.392***	-1.346***	-0.81**	-1.08***
	(0.247)	(0.245)	(0.339)	(0.329)
Standardized coefficient	-0.417	-0.403	-0.243	-0.323
R-squared	0.45	0.46	0.50	0.52
Observations	219	219	219	219
Panel B: Average socioeconomic condition				
South of the Tagus (high military insecurity)	-0.087***	-0.085***	-0.062***	-0.067***
	(0.02)	(0.02)	(0.016)	(0.019)
Standardized coefficient	-0.546	-0.53	-0.386	-0.419
R-squared	0.52	0.53	0.59	0.58
Observations	219	219	219	219
Panel C: Average number of vehicles per household				
South of the Tagus (high military insecurity)	-0.15***	-0.142***	-0.109***	-0.155***
	(0.043)	(0.041)	(0.03)	(0.032)
Standardized coefficient	-0.461	-0.438	-0.335	-0.477
R-squared	0.45	0.48	0.52	0.47
Observations	219	219	219	219
Panel D: Labor force activity rate				
South of the Tagus (high military insecurity)	-3.294***	-3.253***	-2.725***	-2.892***
	(0.936)	(0.916)	(0.995)	(0.989)
Standardized coefficient	-0.469	-0.463	-0.388	-0.412
R-squared	0.30	0.30	0.31	0.32
Observations	219	219	219	219
Boundary fixed effects	Yes	Yes	Yes	Yes
Geo-climatic controls	Yes	Yes	Yes	Yes

Notes: Variables descriptions are provided in Table A1. South of the Tagus (high military insecurity) is a dummy variable indicating whether the grid cell is located south of the Tagus. Regressions include a constant term which is omitted for space considerations. Sample restricted to grid cells within 50 km of the border. The set of geographic–climatic controls includes rainfall and temperature (both in linear and quadratic terms) and altitude. Standard errors corrected for spatial dependence are in parentheses. *, ** and *** denote significance at the 10, 5 and 1% level, respectively.

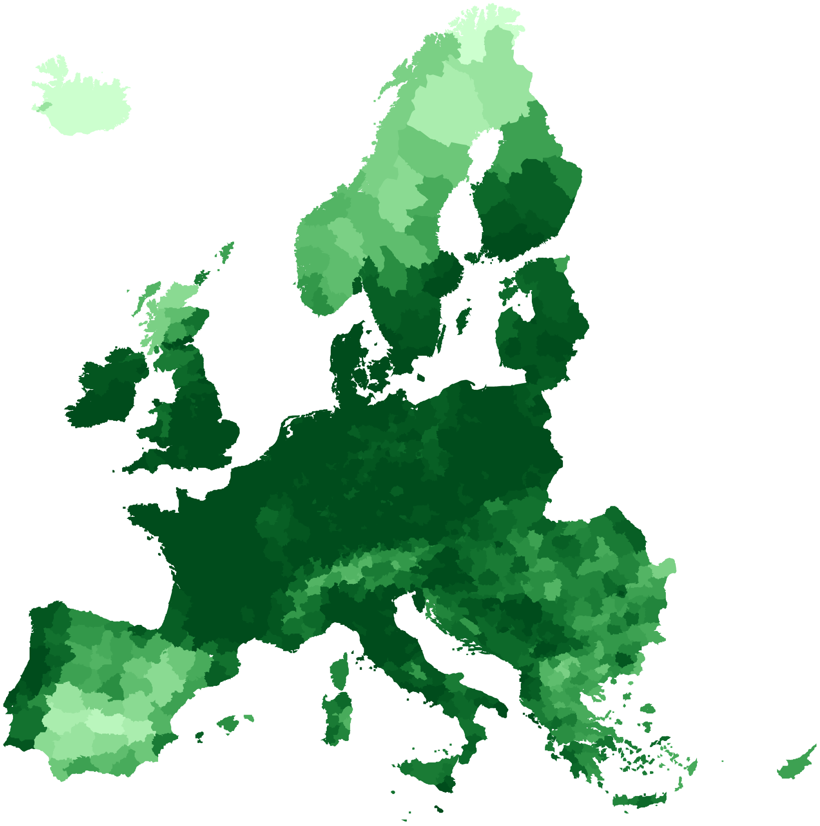


Figure 1. Settlement density in Europe in 2011

Notes: The figure represents the indicator *SD*^{10} that measures the percentage of 10km^2 grid cells that are populated in each region.

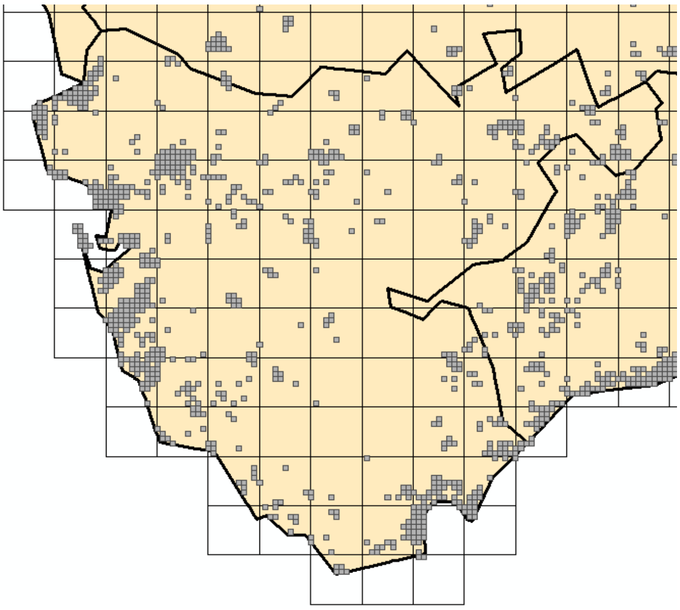


Figure 2. Example of the construction of the indicator of settlement density

Notes: The figure illustrates the construction of an indicator of settlement density of cell size 100 for the NUTS ES612 (Cadiz, Spain). There are three layers of data: i) the NUTS 3 administrative boundaries, ii) a 100-km^2 grid, and iii) the GEOSTAT 1-km^2 population grid.

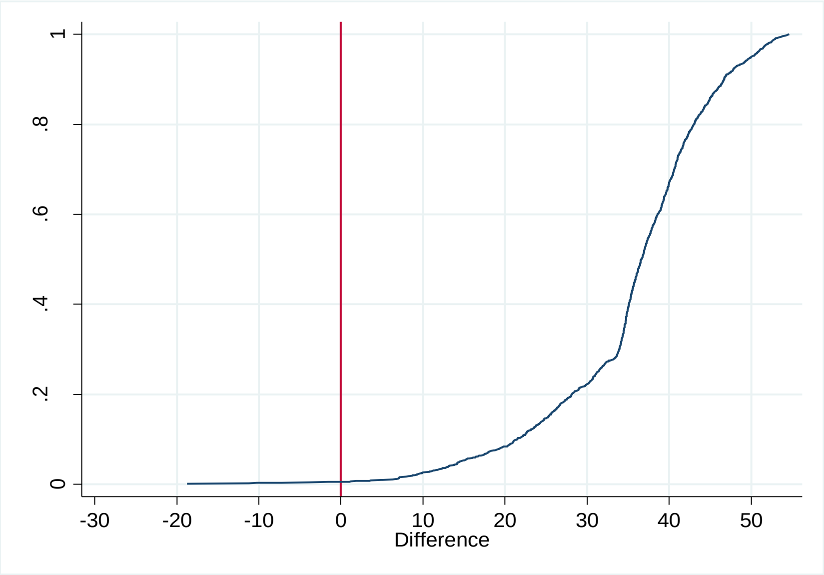


Figure 3. Cumulative distribution of the difference between the coefficient on the virtual region and on southern Spain from 1,423 regressions

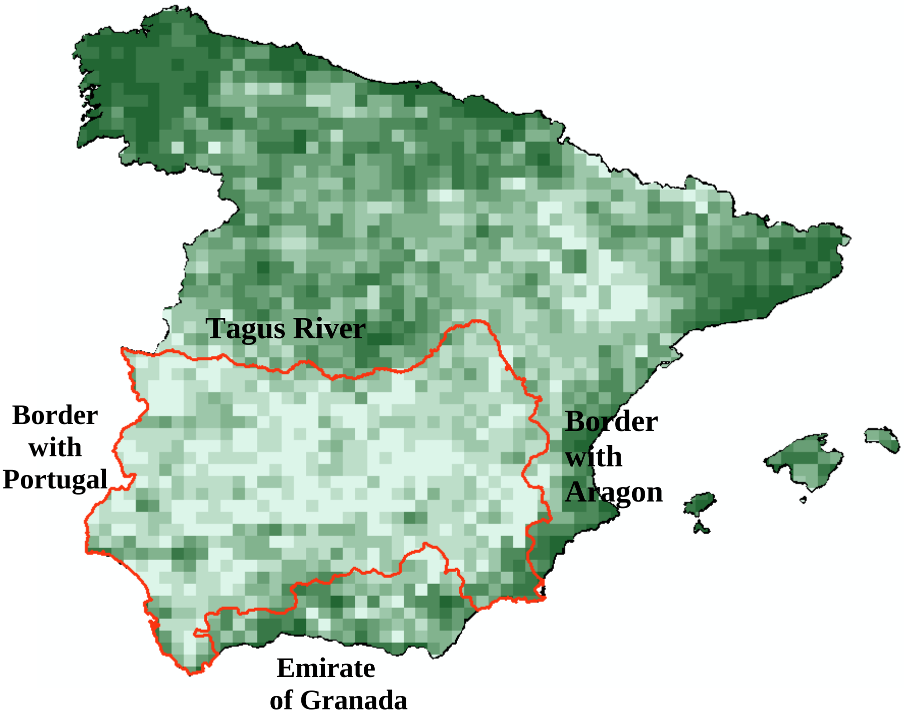
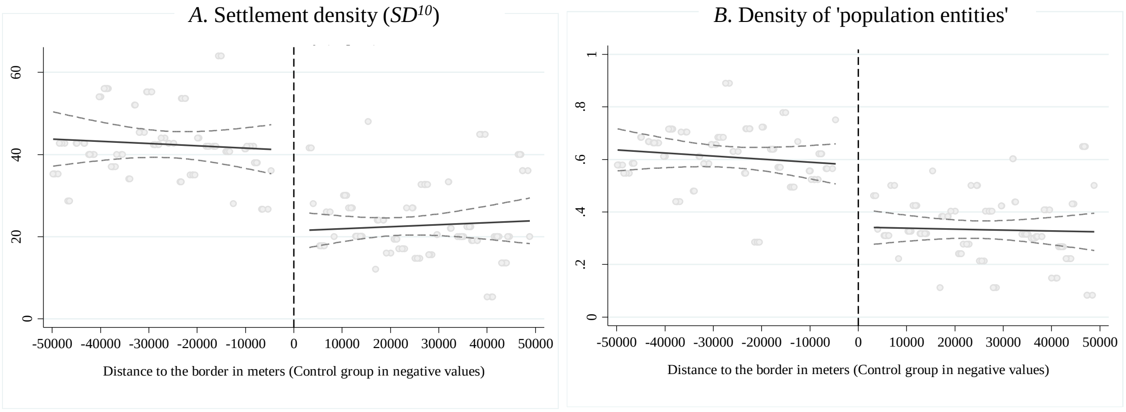


Figure 4. Area conquered by Leon and Castile approximately during the 12th and 13th centuries and settlement density



A. Settlement density (*SD*^{10})

B. Density of 'population entities'

C. Density of municipalities

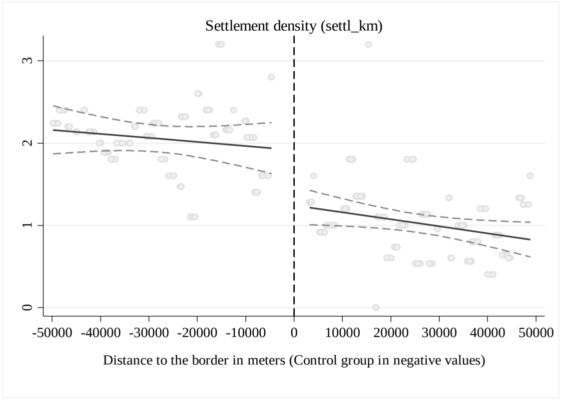
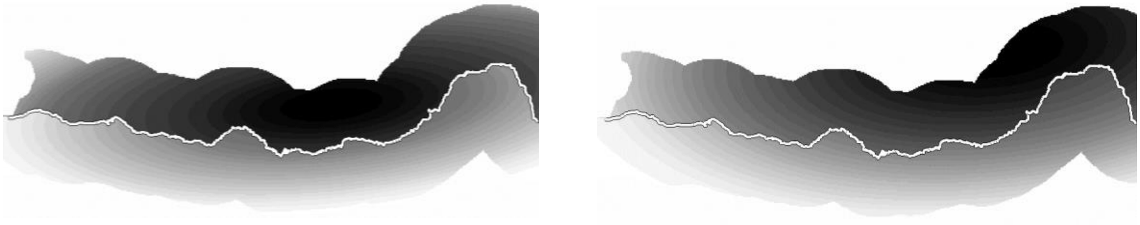
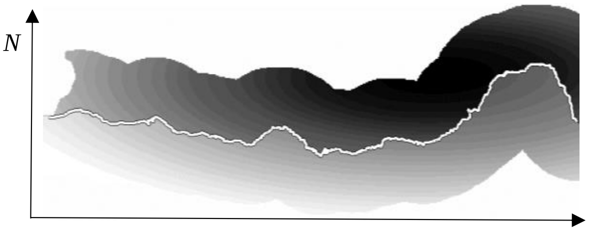


Figure 5. One-dimensional RD figures showing the discontinuity at the border



A. Settlement density (*SD*^{10})

B. Density of 'population entities'



C. Density of municipalities

Figure 6. Two-dimensional RD figures showing the discontinuity at the border

Note: A darker color indicates a higher density.

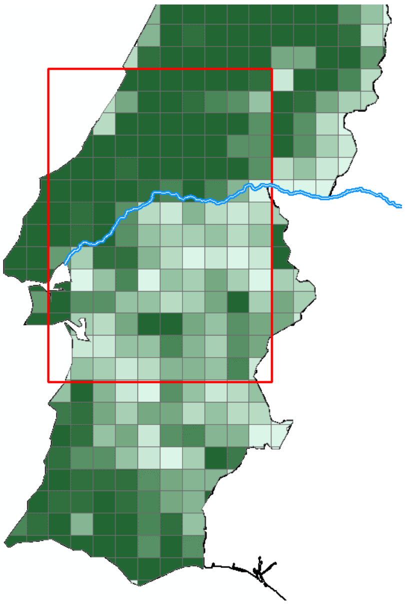


Figure 7. Discontinuity across the Tagus River in Portugal

Note: A darker color indicates higher settlement density.

Table A1

Description of variables

Variable	Description	Source
Variables measuring settlement patterns:		
Settlement density (SD^{10})	Percentage of 10km^2 grid cells that are inhabited in the observation unit.	Author's elaboration using data from GEOSTAT 2011.
Density of 'population entities'	Percentage of 25km^2 grid cells with at least one population entity.	Author's elaboration using data from Instituto Geográfico Nacional (2016).
Density of municipalities	Number of municipalities per 100 km^2.	Author's elaboration using data from Communes 2013 (EUROSTAT).
Settlement density: Census of "Floridablanca" (1787)	Number of settlements (i.e., population entities) in 1787 per km^2.	Author's elaboration using the 1787 population census (INE 1987).
Settlement density: Census of "Pecheros" of Carlos I (1528)	Number of settlements (i.e., population entities) in 1528 per km^2.	Author's elaboration using the 1528 census (INE 2008).
Pre-medieval settlements	Percentage of 25km^2 grid cells in the observation unit with pre-medieval settlements, where pre-medieval settlements refer to documented sites reflecting human settlements and existing before 500 CE.	Author's elaboration using data from Pleiades (2016).
Variables delimiting the area corresponding to the Spanish anomaly in settlement patterns:		
Treated region (conquered by Leon and Castile c.12-13th)	Dummy variable capturing the area conquered by Leon and Castile during (approx.) the 12th and 13th centuries. It is delimited by the Tagus River to the North, the Aragonese border to the East, the Portuguese border to the West and the Granada frontier to the South. This variable is used in Table 3.	Author's elaboration.
South of the Tagus (high military insecurity)	Dummy variable indicating whether the grid cell is located south of the Tagus River, and therefore, affected by high military insecurity from the 11th to 13th centuries (more precisely, from 1085 to 1212). This variable is used in tables 4 to 8.	Author's elaboration.
Geographic and climatic controls:		
Altitude	Average altitude of the surface area of the observation unit.	Author's elaboration using data from GTOPO30 (Data available from the U.S. Geological Survey).
Aridity Index	Average aridity of the grid cell surface area, corresponding to the period 1950–2000. Higher values indicate more humid conditions.	Author's elaboration using geo-referenced data from Trabucco and Zomer (2009).
Depth to rock	This variable refers to the cell's average depth of soil above rock, measured in centimeters.	Author's elaboration using geo-referenced data from Panagos et al. (2012), Liedekerke et al. (2006), and Panagos (2006).
Distance to the coast	Geodesic distance between the centroid of the observation units and the nearest point of the coast (in km).	Author's elaboration.

Table A1 (*continued*)

Description of variables

Variable	Description	Source
Geographic and climatic controls (continued):		
Distance to the Tagus River	Geodesic distance between the centroid of the grid cell and the Tagus River (in meters).	Author’s elaboration.
Distance to Madrid (to Lisbon)	Geodesic distance between the centroid of the grid cell and Madrid (Lisbon), in meters.	Author’s elaboration.
Distance to Roman roads	Geodesic distance between the centroid of the grid cell and the nearest Roman road, in kilometers.	Author’s elaboration using georeferenced data from McCormick et al. (2013)
Dominant surface textural class	Average dominant surface textural class of the grid cell. This variable ranges from 1 to 4 where 1 is coarse (clay < 18 % & sand > 65 %), 2 medium (18% < clay < 35% & sand > 15%,\or clay < 18% & 15% < sand < 65%), 3 medium fine (clay < 35 % & sand < 15 %), and 4 fine (35 % < clay < 60 %).	Author’s elaboration using georeferenced data from Panagos et al. (2012), Liedekerke et al. (2006), and Panagos (2006).
Island dummy	Dummy variable indicating whether the observation unit is within an island. The *mainland* part of countries that are themselves islands are not classified as islands (eg., UK, Iceland, etc.).	Author’s elaboration.
Latitude/ Longitude	The geographic coordinates of the grid cell centroids, in decimal degrees.	Author’s elaboration.
Precipitation	Annual precipitation, in hundred of mililiters. It corresponds to the average value of the surface area of the observation unit.	Author’s elaboration using data from WorldClim (Hijmans et al., 2005).
Ruggedness	Standard deviation of the altitude of the territory corresponding to the observation unit.	Author’s elaboration using data from GTOPO30 (Data available from the U.S. Geological Survey.).
Boundary fixed effects (Segments dummies)	Four dummy variables capturing the closest segment to the grid cell centroid. For instance, s*egment dummy 1* indicates whether the closest segment to the cell centroid is the first (most westerly) segment from four equal-length segments in which the Tagus River is divided.	Author's elaboration.
Soil erodibility class	Average soil erodibility class of the grid cell. This variable ranges from 1 to 5 where 1 is very weak erodibility and 5 very strong erodibility.	Author’s elaboration using georeferenced data from Panagos et al. (2012), Liedekerke et al. (2006), and Panagos (2006).
Soil quality	Average of seven key soil dimensions important for crop production: nutrient availability, nutrient retention capacity, rooting conditions, oxygen availability to roots, excess salts, toxicities, and workability. The average value for each component is calculated for the surface area corresponding to the observation unit .	Author’s elaboration using data from Fischer et al. (2008).

Table A1 (*continued*)

Description of variables

Variable	Description	Source
Geographic and climatic controls (continued):		
Subsoil (topsoil) available water capacity	Average available water capacity of the subsoil (topsoil) corresponding to the grid cell, measured in mm/m.	Author's elaboration using georeferenced data from Panagos et al. (2012), Liedekerke et al. (2006), and Panagos (2006).
Temperature	Annual average temperature. It corresponds to the average value of the surface area of the observation unit.	Author's elaboration using data from WorldClim (Hijmans et al., 2005).
Topsoil organic carbon content	Average topsoil organic carbon content of the grid cell, measured in percentage points, where a high content is considered above 6% and very low content below 1%.	Author's elaboration using geo-referenced data from Panagos et al. (2012), Liedekerke et al. (2006), and Panagos (2006).
Proxies for current economic development at the local level:		
Average number of vehicles per household	Number of vehicles for personal transport owned by households, divided by the number of households. The year of measurement is 2001. The value of the grid cell corresponds to the weighted average of the values of the municipalities overlapping with the cell, with weights being the overlapping municipalities' areas.	Author's elaboration using data from INE. Censos de Población y Viviendas 2001 (www.ine.es).
Average socioeconomic condition	Average of class marks of socioeconomic conditions of individuals (multiplied by 100). Socioeconomic condition is obtained by combining information from the variables occupation, activity and professional situation. The year of measurement is 2001. The value of the grid cell corresponds to the weighted average of the values of the municipalities overlapping with the cell, with weights being the overlapping municipalities' areas.	Author's elaboration using data from INE. Censos de Población y Viviendas 2001 (www.ine.es).
Labor force activity rate	Labor force activity rate of the population between 20 and 59 years old. The year of measurement is 2001. The value of the grid cell corresponds to the weighted average of the values of the municipalities overlapping with the cell, with weights being the overlapping municipalities' areas.	Author's elaboration using data from INE. Censos de Población y Viviendas 2001 (www.ine.es).
Log light density at night	Natural logarithm of 0.001 plus the average night light density from 2000 until 2005.	NOOA/National Centers for Environmental information, https://ngdc.noaa.gov/

Notes: The observation units are either NUTS 3-regions or 250km^2 grid cells. The basic layer with the administrative limits, used in the variables computed with GIS software, comes from EUROSTAT (NUTS 2010). This layer is combined with a layer defining the sub-national administrative borders of some Balkan countries not covered by the NUTS 2010 shapefile. This second shapefile comes from Natural Earth (http://www.naturalearthdata.com/).