Online Appendix

to

Geography and State Fragmentation

by

Shuhei Kitamura and Nils-Petter Lagerlöf

Abstract: This Online Appendix describes several results and robustness checks that did not make into the paper "Geography and State Fragmentation" due to space constraints. It is organized into several different sections. Section 1 discusses some illustrations and extensions of the model. Section 2 provides some correlation statistics. Section 3 considers a broader set of geography variables. Section 4 performs some robustness checks with respect to spatial controls. Section 5 examines some alternative methods and sources for measuring both borders as such, and stability of borders, as well as state fragmentation more generally. Section 6 shows the results when using smaller and larger cells. Finally, Section 7 considers different ways of using our night lights and population density data.

1 The model

1.1 Illustration of Y^*

Figure A.1 provides an illustration of how Y^* varies with $N = 1/s$ for the same numerical example as in Figure 2 in the paper.

1.2 Other ways to model spatial resource allocation

This section considers a version of the model where the elite allocate resources non-uniformly across the state's territory.

To that end, let output at distance $d \in [0, s/2]$ from the center of country i in period t, denoted $\widetilde{Y}_{i,t}(d)$, be given by

$$\widetilde{Y}_{i,t}(d) = \left[\widetilde{Z}(d)A_{i,t}\right]^{\alpha} \left[\widetilde{R}_{i,t}(d)\right]^{1-\alpha}, \tag{1}$$

where $\widetilde{R}_{i,t}(d)$ denotes the amount of resources allocated to a location at distance d from the center of country i, and $A_{i,t}$ is country i's provision of a public good, located at the center of a country, which benefits locations at distance d from the center by a factor $\widetilde{Z}(d)$, given by (2) in the paper. As in the paper, $A_{i,t}$ could represent country i's level of technology.

In each period, the elite first allocate the resources under their control to maximize total output. Denoting their total amount of resources by $R_{i,t}^{tot}$, the elite thus maximize

$2\int_0^{s/2} \widetilde{Y}_{i,t}(x)dx$, subject to $2\int_0^{s/2} \widetilde{R}_{i,t}(x)dx = R_{i,t}^{tot}$, taking $R_{i,t}^{tot}$ as given. Somewhat informally, ignoring that the control variable is continuous, the Lagrangian associated with this maximization problem can be written as $\mathcal{L} = 2\int_0^{s/2} \widetilde{Y}_{i,t}(x)dx + \Omega\left[R_{i,t}^{tot} - 2\int_0^{s/2} \widetilde{R}_{i,t}(x)dx\right]$, where Ω is the Lagrangian multiplier. The first-order condition can be written as

$$\frac{\partial \widetilde{Y}_{i,t}(d)}{\partial \widetilde{R}_{i,t}(d)} = (1-\alpha)\left[\widetilde{Z}(d)A_{i,t}\right]^{\alpha}\left[\widetilde{R}_{i,t}(d)\right]^{-\alpha} = \Omega, \tag{2}$$

for all $d \in [0, s/2]$, which states that the marginal productivity of resources is equalized across locations.

Using (2), we can write resources at each location as $\widetilde{R}_{i,t}(d) = ([1-\alpha]/\Omega)^{\frac{1}{\alpha}} \widetilde{Z}(d)A_{i,t}$. Using the budget constraint for resources gives

$$2\int_0^{s/2} \widetilde{R}_{i,t}(x)dx = 2\left(\frac{1-\alpha}{\Omega}\right)^{\frac{1}{\alpha}} A_{i,t}\left[\int_0^{s/2} \widetilde{Z}(x)dx\right] = R_{i,t}^{tot}. \tag{3}$$

Recall from (2) in the paper that $\widetilde{Z}(d) = 1-4\gamma d$, which implies that $\int_0^{s/2} \widetilde{Z}(x)dx = (s/2)(1-\gamma s)$. Inserted into (3), this gives

$$\left(\frac{1-\alpha}{\Omega}\right)^{\frac{1}{\alpha}} = \frac{R_{i,t}^{tot}}{A_{i,t}s(1-\gamma s)}, \tag{4}$$

which can be inserted into (2) to give resources per location as

$$\widetilde{R}_{i,t}(d) = \left(\frac{1-\alpha}{\Omega}\right)^{\frac{1}{\alpha}} \widetilde{Z}(d)A_{i,t} = \frac{R_{i,t}^{tot}}{s}\frac{\widetilde{Z}(d)}{1-\gamma s}. \tag{5}$$

Intuitively, resources allocated to locations at distance d from the center, relative to the average resources across the country, are proportional to each location's productivity, relative to the average productivity of the country. Substituting (5) into the production function in (1) shows that

$$\widetilde{Y}_{i,t}(d) = \widetilde{Z}(d)\left(\frac{1}{1-\gamma s}\right)^{1-\alpha} A_{i,t}^{\alpha}\left(\frac{R_{i,t}^{tot}}{s}\right)^{1-\alpha}. \tag{6}$$

Again using $\int_0^{s/2} \widetilde{Z}(x)dx = (s/2)(1-\gamma s)$, and recalling that average output per location equals $(2/s)\int_0^{s/2} \widetilde{Y}_{i,t}(d)dx = Y_{i,t}$, we get

$$Y_{i,t} = (1-\gamma s)\left(\frac{1}{1-\gamma s}\right)^{1-\alpha} A_{i,t}^{\alpha}\left(\frac{R_{i,t}^{tot}}{s}\right)^{1-\alpha} = (1-\gamma s)^{\alpha}A_{i,t}^{\alpha}\left(\frac{R_{i,t}^{tot}}{s}\right)^{1-\alpha}. \tag{7}$$

Finally, we can set total resources to $R_{i,t}^{tot} = sR_{i,t}$, where $R_{i,t}$ denotes resources per location, as given by (4) in the paper. This produces the same expression for $Y_{i,t}$ as in (3) in the paper, except that the factor $1-\gamma s$ is now replaced by $(1-\gamma s)^{\alpha}$.

2 Descriptive statistics

2.1 Cross-correlations between different border variables

Table A.1 presents cross-correlations between each of the six Euratlas border dummies ($b_{i,t}$) and the border frequency index, as well as a language dummy constructed from the World Language Mapping System, and what we call a current border dummy constructed from the Global Administrative Areas dataset; see Section 5.1 below for more details.

All border dummies show highly significant and large positive correlations with the border index. The border dummy for 2000 has the lowest correlation with border frequency, but even that coefficient is as high as 0.599. The border dummies also show positive correlation with each other, typically larger between closer years, suggesting that borders are not stationary but change gradually over time. Despite the rise and fall of several states and empires over these centuries, the locations of the borders between them are thus quite persistent. This is consistent with a theory where some underlying constant factor, such as geography, ultimately determines border locations.

Table A.1 shows a very high correlation coefficient (0.913) between the Euratlas border dummy for 2000 and the current border dummy, which also speaks to the reliability of the Euratlas data.

The language border dummy shows the highest correlation with the Euratlas dummy for 2000 (a correlation coefficient of 0.510) and the current border dummy (0.526). It thus seems that state formation today follows ethnic lines more closely than in preindustrial times. This may reflect the spread of democracy, making it easier for ethnic minorities living in a well defined territory to secede and form their own states (see, e.g., Alesina and Spolaore 2003). It could also be due to genocide, ethnic cleansing, and policies by state governments that make ethnic and linguistic minorities comply with the state's majority identity, as well as more voluntary forms of migration.

2.2 Cross-correlations between border frequency, geography, and modern outcomes

Table A.2 shows the unconditional correlation coefficients between border frequency, B_i, and each of our benchmark geography variables. The results confirm that the correlations in Table 2 in the paper hold without controls. It also shows that log night lights and log population density show positive and significant unconditional correlations with border frequency, confirming the results in Table 6 in the paper.

As would be expected, some of the geography variables capture the same variation. For example, the mountain dummy for 1000 meters and log ruggedness have a correlation coefficient of 0.474. However, most geography variables capture different types of variation. For example, river density and the mountain dummies both show stronger correlation with the border index than with each other, thus capturing different dimensions through which geography affects borders.

Note also that the two variables measuring suitability for rainfed and irrigated agriculture show high and strongly significant correlation with each other, but their respective correlations with border frequency carry opposite signs: positive for rainfed and negative for irrigated.

The geography variable showing the strongest correlation with border frequency is rainfall, with a correlation coefficient of 0.255. To illustrate what drives this, the map in Figure A.2 shows the locations of cells in the top and bottom quartiles of rainfall. As seen, rainfall is highest in the more fragmented Western Europe, and lowest in the relatively unified Middle East.

Figure A.3 illustrates the relationship between modern outcomes and border frequency in a bar graph diagram similar to that in Figure 8 in the paper, but here conditional on geography. For each border frequency outcome, the bars show the means of the residuals of log night lights and log population density, respectively, after regressing each of them on the benchmark geography variables used in column (9) of Table 2 in the paper. The patterns are similar to the unconditional ones in Figure 8 in the paper.

2.3 Adjusting for the Holy Roman Empire: cross-correlations with geography

As noted in Table 3 in the paper, treating the Holy Roman Empire as unified does not affect the correlations between border frequency and geography much. One way to understand why is to compute the difference between the border frequency index used in the main analysis (not treating the HRE as unified) and border frequency when treating it as unified, and then correlate this difference with the benchmark geography variables.

Table A.3 shows the correlation between our benchmark geography variables and the change in border frequency due to HRE adjustment. The correlations are relatively weak. Even where the coefficients are significantly different from zero, such as for river density and rainfall, their magnitudes are small. In other words, the HRE adjustment does not shift borders in ways that correlate with any of the geography measures we are looking at.

3 More geography controls

Tables A.4-A.10 show the results when using different sets of geography variables than those included in our benchmark specification.

3.1 Elevation

Table A.4 considers some specifications where we enter log elevation as control. When constructing this variable, in order not to drop cells with negative elevation (73 cells in total), we use elevation exceeding the lowest level in the sample. That is, if x_i denotes mean elevation of cell i (in meters) and $\widehat{x}$ is the minimum x_i across the 5202 cells (which in our baseline sample is -28 meters, located close to the Caspian Sea), then log elevation is constructed as $\ln(1 + x_i - \widehat{x})$, which equals zero for the cell with the lowest elevation.

Log elevation has a relatively high unconditional correlation with border frequency, as seen when it is entered alone in column (1), but it does not come out as significant when entered with our other benchmark geography controls. The reason is that log elevation is highly correlated with log ruggedness, which absorbs most of the same variation; their correlation coefficient is 0.81.

3.2 Coal, temperature, and lakes

Table A.5 considers three other geography variables. The coal dummy indicates presence of coal in the cell, as defined by the presence of rock of specific ages in maps provided by the Bundesanstalt für Geowissenschaften und Rohstoffe (BGR) in Hannover, Germany.[24]

Temperature refers to mean annual temperature measured in degrees Celsius averaged for the period 1961-90. The source is GAEZ, which we used also for agricultural suitability and rainfall.

The last variable measures the fraction of the cell's area covered by lake, based on Natural Earth data, which is the source used also for, e.g., coastlines and rivers.

Coal carries a positive coefficient significant at the 5% level when entered alone in column (1), and together with the other two in column (4). However, none of them comes out as significant when controlling for our benchmark set of controls.

All specifications in Table A.5 include latitude fixed effects, but the results for these three variables do not change qualitatively without fixed effects, or with both latitude and longitude fixed effects.

[24] We use the map IGME 5000 from BGR, and the file "age (chronostratigraphic).lyr" in a folder labelled "layer." The coal dummy indicates presence of rocks from the following geologic periods: Carboniferous (C), Carboniferous-Permian (C-P), Carboniferous-Middle Permian (C-P2), Early Carboniferous (C1), Late Carboniferous (C2), Late Carboniferous-Permian (C2-P), and Late Carboniferous-Middle Permian (C2-P2).

3.3 Alternative agricultural suitability variables

Recall that our two benchmark measures of agricultural suitability are based on the four most common grains (wheat, barley, oats and rye), and refer to potential yields when using rainfed and irrigated agriculture, respectively. Table A.6 examines two alternative measures of agricultural suitability.

Suitability for potato agriculture is constructed from GAEZ, the source used for our benchmark measures, and was originally used by Nunn and Qian (2011). The Caloric Suitability Index (CSI) comes from Galor and Özak (2016) and is also partly based on GAEZ, but is a calorie weighted measure of the yield a cell can generate if growing the crop with the highest caloric content. Here we use the definition that considers all crops available after 1500, i.e., in the wake of the Columbian exchange. Both are constructed under the assumption that rainfed agriculture is used.

Columns (1)-(3) of Table A.6 enter the potato measure and CSI, both separately and together, in lieu of the two benchmark agricultural suitability measures, keeping all other benchmark controls unchanged. There is no significant effect from CSI on borders, but the potato measure comes out as negative and significant. This pattern holds broadly when including our benchmark measure for suitability for rainfed agriculture as control in columns (4)-(6), and when entering both of our benchmark suitability measures, rainfed and irrigated suitability, in columns (7)-(9).

The two alternative measures are highly correlated with our benchmark measure of suitability for rainfed agriculture: the correlation coefficients are 0.73 and 0.61 for the potato measure and CSI, respectively. They are somewhat less correlated with the irrigated suitability measure, for which the corresponding correlation coefficients are 0.39 and 0.30. Since both the potato measure and CSI are constructed under the assumption that rainfed agriculture is used, this is not too surprising.

Because we want to be able to capture the possibly different effects of suitability for rainfed and irrigated agriculture, and because the potato and CSI measures do not have any irrigation based equivalents, we choose the measures based on the four common grains as our benchmark controls.

One possible argument against using the potato measure—the alternative measure that comes out as most significant in the regressions—could be that the four common grains may have been an overall more important source of nutrition than the potato for the region and period that we consider. According to Leff et al. (2004, Table 5), wheat is the currently most commonly grown crop by land area in the region that we consider (the Middle East, Europe, Central Asia, Asiatic Russia, and North Africa). The land most suitable for potato cultivation is concentrated in Europe (Nunn and Qian 2011, pp. 611-612).

3.4 River and coast dummies

Table A.7 shows the results when regressing border frequency on the benchmark set of geography controls, but using a river dummy instead of river density and a coast dummy instead of coastline density. These dummies are indicators of the presence of a river or coast in the cell. That is, they take the value one when the corresponding density variables are strictly positive, and zero otherwise. The specifications are otherwise identical to those in Table 3 in the paper. Not too surprisingly, the results are very similar.

Table A.8 shows some regressions in terms of local deviations, where the specifications correspond to those in Table 4 in the paper, but now replacing local deviations in river and coastline density with local deviations in the river and coast dummies, respectively. Again, the results are very similar, but local deviations in the coast dummy in Table A.8 come out as less significant than local deviations in coastline density in Table 4.

3.5 Modern outcomes with more geography controls

Table A.9 shows the results when regressing night lights and population density on border frequency, different sets of geography controls, and latitude fixed effects. Columns (1) and (5) replicate column (3) of Table 6 in the paper (both panels), but now also reporting the coefficients on the geography variables; columns (2) and (6) add the potato and CSI measures of agricultural suitability; columns (3) and (7) add a coal dummy, temperature, and the fraction of the cell's area covered by lake water; and columns (4) and (8) add all these variables together with the benchmark geography controls.

The coefficient on border frequency comes out as significant in all specifications. Some of the geography variables come out as significant as well, although the coefficients might be hard to interpret since many of the variables capture the same variation. For example, as discussed above, suitability for rainfed agriculture based on the four common grains is highly correlated with the potato measure and CSI.

Table A.10 shows the results corresponding to those in Table A.9, but in terms of local deviations. The results are similar to those in Table 7 in the paper, although local deviations in the border index here come out as slightly less significant, in particular in columns (3) and (7) where we use local deviations in the 2000 meter mountain dummy. The estimated coefficients are more significant when using the 1000 meter mountain dummy in columns (4) and (8).

4 Alternative spatial controls

4.1 Sparser latitude and longitude fixed effects

Many regressions in the paper enter fixed effects for every half-degree latitude and/or longitude, i.e., one dummy for each row and/or column in the grid. Table A.11 considers a somewhat sparser set of controls, namely dummies for each even integer degree latitude and/or longitude. For example, with these sparser controls, cells centered on latitudes 25, 25.5, 26, and 26.5 degrees are assigned the same latitude dummy, since their closest even integer latitude is 26 degrees. In that sense, our benchmark half-degree controls are four times as fine as these alternative controls.

Column (3) of Table A.11 enters fixed effects for even integer degree latitudes, and column (5) enters dummies for even integer degrees latitude and longitude. For comparison, column (1) shows the results without any fixed effects at all, while columns (2) and (4) show the results with our benchmark half-degree fixed effects, replicating some of the regressions in the paper; see column (9) of Table 2, and columns (1) and (5) of Table 3, respectively.

The differences in terms of coefficient estimates are relatively small both when comparing columns (2) and (3), and when comparing columns (4) and (5). In particular, the columns with sparse and fine controls differ less from each other than from column (1), where we drop the fixed effects altogether. That is, the sparser even-degree fixed effects absorb roughly the same variation as the more rigorous half-degree fixed effects. While not reported here, the results can be seen to differ even less when using fixed effects at an intermediate level of sparseness—i.e., with dummies for every integer latitude and/or longitude—which is not surprising.

4.2 Different Conley cutoffs

Most of our analysis adjusts the standard errors to control for spatial correlation using the Conley (1999) method, with cutoffs of 1.45 degrees. Larger cutoffs mean we allow for correlation in the error term across cells at longer distances from each other. Table A.12 shows that the results when setting these cutoffs to one or two degrees, instead of 1.45, do not change much qualitatively, although the standard errors expand a little with larger cutoffs.

The same holds when looking at local deviations, as shown in columns (1) and (2) of Table A.13, where we consider only specifications with local deviations in mountains over 1000 meters; the results for the 2000-meter mountain dummy (not reported) are insignificant, as in the benchmark case.

These local deviations are calculated as deviations from the average among the eight closest neighboring cells. We can do the same exercise instead using the 24 closest neighboring cells. As seen in columns (3) and (4) of Table A.13, the results are largely unchanged.

4.3 Fixed effects for clusters of cells

In the paper we explored several specifications with fixed effects for each half-degree latitude and/or half-degree longitude. This controls for any unobserved characteristics that vary either north-south or east-west, but cells can also share unobserved characteristics if they are close to each other without being on the exact same half-degree latitude or half-degree longitude. Running regressions in terms of local deviations serves to address this.

A closely related approach is to define dummies for square clusters of cells located at most two half-degree latitudes north or south from each other, and/or two half-degree longitudes east or west of each other, each cluster thus containing 9 cells. Similarly, we can create dummies for clusters of 16 or 25 cells located at most three or four half-degree latitudes north or south from each other, and/or three or four half-degree longitudes east or west of each other.

Obviously, no unobserved characteristic would be distributed exactly in square clusters. Moreover, where the squares are centered will always be somewhat arbitrary. However, these cluster fixed effects should arguably do a good job absorbing any factor that is *approximately* constant between closely neighboring cells.

Table A.14 presents the results when regressing modern outcomes on border frequency, with our benchmark set of geography controls, and entering fixed effects for clusters of either 9, 16, or 25, neighboring cells. Interestingly, border frequency now has a negative effect on modern outcomes in all specifications. The results thus mirror the patterns that we found when regressing local deviations in modern outcomes on local deviations in border frequency. Intuitively, both methods control for unobserved characteristics among cells that are close to each other.

5 Alternative measures and sources for borders

5.1 Border dummies

The analysis in the paper was based mostly on the average border frequency 1500-2000 defined in (13) in the paper, or its local-deviation equivalent in (17). Here we examine outcomes for each of the six dummies from Euratlas on which we computed border frequency.

We also construct two border dummies based on other sources than Euratlas. The first of these we call the current border dummy, which is based on maps from the Global Administrative Areas (www.gadm.org). These are supposed to show contemporary state borders. We do not know to which specific point in time that these refer, but the GADM Version 2 data that we use were posted in January 2012. The current border dummy is highly correlated with the Euratlas border dummy for 2000.

The other variable we call the language border dummy, which is calculated from the World Language Mapping System (www.worldgeodatasets.com/language).

Let $b_{i,j} \in \{0,1\}$ be the same border dummy as before, equal to one if cell i had a border in year j, or in this case a current border or a language border, and zero otherwise. That is, $j \in \{1500, ..., 2000, \text{CB}, \text{LB}\}$, with CB and LB indicating current and language borders, respectively. We can then define the local deviation in $b_{i,j}$ as $\Delta b_{i,j} = b_{i,j} - b_{-i,j}$, where now $b_{-i,j}$ is the average of $b_{i,j}$ in the eight closest neighboring cells to cell i, defined analogously to (16) in the paper. Since $b_{i,j} \in \{0,1\}$, it holds that $b_{-i,j} \in [0,1]$ and $\Delta b_{i,j} \in [-1,1]$.

5.1.1 Geography and border dummies

Table A.15 shows the outcomes when regressing each of the border dummies, $b_{i,j}$, on our benchmark set of geography controls. All specifications include latitude fixed effects. Even though the dependent variable is binary, we use ordinary least squares estimation to facilitate comparison to the regressions with border frequency as the dependent variable.

The coefficient estimates in columns (1)-(6) of Table A.15 carry the same signs as in the corresponding border frequency regression in column (1) of Table 3 in the paper, at least when they are significant. Some variables do come out as larger in magnitude and more significant in some years than others, but overall it does not seem that the results when using the border frequency index are driven by any particular set of years.

The outcomes when using current and language borders in columns (7) and (8) are also qualitatively similar, even though these borders come from different sources and, in the case of language borders, possibly measure something quite different. This suggests that the patterns documented earlier are not a reflection of any peculiarities in the Euratlas data.

Table A.16 shows the results in terms of local deviations, regressing $\Delta b_{i,j}$ on local deviation in geography variables. Columns (1)-(6) consider the Euratlas border dummies, and columns (7) and (8) the current or language border dummies, respectively. All specifications include latitude fixed effects. When comparing the results in columns (1)-(6) to those in column (1) of Table 4 in the paper, we see that all coefficient estimates that come out as significant carry the same sign across years.

In column (7), we see that the results for local deviations in current borders are similar to those for the Euratlas border dummy for 2000. Column (8) shows fewer significant estimates, one exception being local deviations in log ruggedness, which comes out as positive and significant.

5.1.2 Modern outcomes and border dummies

Table A.17 presents results from regressing modern outcomes on each of the six Euratlas border dummies 1500-2000. All specifications include the benchmark set of geography controls and latitude fixed effects. The correlations are positive in the earlier years and turn negative in 2000, although the positive correlations for the earlier years come out as more significant. That is, the positive effects on modern outcomes are mostly driven by historical borders.

Table A.18 presents the corresponding regressions in terms of local deviations, finding less of time trend in the significance or magnitude of the coefficients.

5.2 Variance as a measure of border stability

To study the effects on modern outcomes from the stability of borders we used border change, defined in (20) in the paper as $C_i = (1/5)\sum_{t=1600}^{2000}(b_{i,t} - b_{i,t-1})^2$. This simply measures the average number of times the cell has changed from having a border to not having one, or *vice versa*.

A closely related measure of border stability is the variance in the border dummy, measured for each cell over the six centuries 1500-2000. Since $B_i = (1/6)\sum_{t=1500}^{2000} b_{i,t}$ is the mean of the border dummy, we can write the variance as $V_i = (1/6)\sum_{t=1500}^{2000} (b_{i,t} - B_i)^2$. Like for the variance in any border dummy this simplifies to

$$V_i = B_i(1 - B_i). \tag{8}$$

Table A.19 shows the result from regressions identical to those in Table 8 in the paper, except that border change (C_i) is replaced by border variance (V_i). The results are qualitatively identical. Conditional on average border frequency, cells have better modern outcomes with less variance in borders, i.e., more border stability.

This formulation also suggests a useful economic interpretation of the result. First, we write a regression equation similar to (19) in the paper, but adding a control for border variance. Letting the coefficient on border variance be ϕ, this gives

$$\begin{aligned} \ln Y_i &= \gamma + \delta B_i + \phi V_i + \epsilon_i \\ &= \gamma + \delta B_i + \phi\,(B_i - B_i^2) + \epsilon_i \\ &= \gamma + (\delta + \phi)B_i - \phi B_i^2 + \epsilon_i. \end{aligned} \tag{9}$$

where we have used (8), and ignored the geography controls.

The estimates in Table A.19 where B_i and V_i enter together give $\widehat{\delta} > 0$ and $\widehat{\phi} < 0$, implying a convex relationship between modern outcomes, as measured by night lights and population density, and border frequency, B_i.

As an example, the specification in column (4) of Table A.19, with night lights as dependent variable, gives $\widehat{\delta} = 0.648$ and $\widehat{\phi} = -0.821$, implying that the lowest predicted levels of night lights can be found where border frequency equals $(\widehat{\delta}+\widehat{\phi})/(2\widehat{\phi}) = (0.821-0.648)/(2 \times 0.821) = 0.105$. It can also be seen that a cell which has a border in just one of the six years ($B_i = 1/6$) is predicted to have worse outcomes than one with zero border frequency ($B_i = 0$).[25] In other words, completely unified cells are better off than those with borders in only one year.

5.3 Other measures of state fragmentation than borders

Borders are intuitive variables to use when measuring state fragmentation. This is particularly true when we want to examine if features of the landscape such as rivers and mountains constitute natural borders. However, borders are not the only measure of state fragmentation that we can think of. Below we consider four related variables:

1. The log of the number of states that enter the cell each year, averaged from 1500 to 2000. That is, if $n_{i,t}$ is the number of states entering cell i in century t, then this variable is defined as $\left(\sum_{t=1500}^{2000} \ln n_{i,t}\right)/6$.

2. The log mean state size each year, averaged from 1500 to 2000. Let $s_{i,t}$ be the mean size in square kilometers of the states that enter cell i in century t (possibly only one state). Then this variable is defined as $\left(\sum_{t=1500}^{2000} \ln s_{i,t}\right)/6$. Note that $s_{i,t}$ here denotes the size of the states (or state) as measured by their (or its) total area, not the only the segments that intersect cell i.

3. The log of the distance to the border each year, averaged from 1500 to 2000. Let $x_{i,t}$ be the kilometer distance to the border, or more precisely to the boundary of the state polygon. Then this variable is defined as $\left(\sum_{t=1500}^{2000} \ln x_{i,t}\right)/6$.

4. Border density is defined as the total length of the boundaries of all states entering the cell, divided by the total areas of the state polygons that intersect the cell, adjusted to equal zero for non-border cells (i.e., cells with only one state). This measure is also averaged from 1500 to 2000. A more formal definition reads as follows. Let $b_{i,t}$ be the border dummy as defined in the paper (equal to one if $n_{i,t} \geq 2$, and zero otherwise, $n_{i,t}$ being the number of states entering cell i). Furthermore, let $l_{i,t}$ be the total length of all state boundaries intersecting cell i in year t, and let $q_{i,t}$ be the sum of the state territories which intersect cell i in year t. Then border density is defined as $\left(\sum_{t=1500}^{2000} b_{i,t} l_{i,t}/q_{i,t}\right)/6$.

[25] That is, $(\widehat{\delta}+\widehat{\phi}) - \widehat{\phi}(1/6) = 0.648 - 0.821 + 0.821/6 < 0$.

5.3.1 Unconditional correlations

Table A.20 shows the unconditional correlation coefficients between each of the five measures of state fragmentation (border frequency and the four others defined above). The correlation coefficients all carry the expected signs. Cells with higher border frequency also have shorter distance to the nearest border, belong to smaller states, and have higher border density.

The strongest correlation is between border frequency and the log number of states. This is perhaps not too surprising, since the border dummies from which border frequency is defined are just indicators that the number of states in a given year and cell exceeds one.

5.3.2 Geography and alternative measures of fragmentation

Table A.21 shows the results when regressing these alternative measures of state fragmentation on our benchmark set of geography controls and latitude fixed effects. For comparison, column (1) of Table A.21 replicates column (1) of Table 3 in the paper, using border frequency as the dependent variable.

The results for border frequency and the log number of states are very similar. This is not surprising, since they have high correlation. It is also reassuring, since the latter corresponds closely to measures used in some related studies. For example, Michalopoulos (2012) uses the log number of ethnic groups in a cell as a measure of ethnic diversity.

Using the other measures of fragmentation as dependent variables, the results are broadly similar to those with border frequency. That is, the coefficients mostly differ in sign when the correlations are negative. Some exceptions are worth noting. The negative effect on state size from rainfed agricultural suitability in column (3) seems to capture that many cells that record high suitability for rainfed agriculture have historically belonged to some of the largest states, such as Russia and Ukraine. The relatively low correlation between state size and border frequency also suggests that they do not capture the same things.

Column (4) of Table A.21 shows no significant effects on border density from mountains, ruggedness, or river density. This is due to the fact that borders at lower elevations and away from rivers tend to "meander" more than borders at higher elevations and where there are rivers. In other words, when borders are located in less mountainous and rugged terrain, and away from rivers, they tend to score very high levels of border density. The vast majority of these cells are located by the coast. This seems to reflect how the coastal segments of state polygons simply follow coastlines.

Overall, this suggests that border density, at least as we define it here, may not be the most suitable variable to uncover how geography forms "natural" borders.

5.3.3 Modern outcomes and alternative measures of fragmentation

Table A.22 regresses modern outcomes on each of these alternative measures of fragmentation. All regressions include the benchmark geography controls and latitude fixed effects. For comparison, columns (1) and (6) use the benchmark border frequency index as a measure of fragmentation, replicating the results for night lights and population density, respectively, as seen in column (3) of Table 6 in the paper. As seen, we observe more night lights, and higher population densities, in cells with a larger number of states, smaller log state size, and shorter distance to the border.

In columns (4) and (9), we also note a positive correlation between modern outcomes and border density. Recall from Section 5.3.2 above that border density partly measures how meandering a border is, and that borders tend to meander more along coastlines. Since night lights are also high in coastal areas (see Figure 9 in the paper), this might evoke suspicion that the correlation is partly spurious. This is another reason we prefer not to use border density as our benchmark outcome variable.

5.3.4 Local deviations in border distance

The model in Section 3 of the paper predicts that output is lower closer to borders; see Result 2 in the paper. To test this we correlated local deviations in modern outcomes with local deviations in border frequency. This approach does not utilize local variation across cells with zero border frequency, i.e., cells that have been located closer to the center of a state in all years 1500-2000.

Table A.23 shows a couple of regressions where we instead use border distance. Local deviations in log border distance indeed show a positive and significant correlation with local deviations in both log night lights and log population density. In other words, locations farther from state borders (and closer to the center of states) are more developed. This is consistent with Result 2 of the model, and the corresponding correlations when using local deviations in border frequency; see Table 7 in the paper.

5.4 The Abramson data

Our border variables were computed from the maps compiled by Euratlas (Nussli 2010). In this section we apply the same procedure to another set of maps used by Abramson (2017). These use as starting point the Centennia Historical Atlas, the original creator of which is Reed (2008). We refer to these as the Abramson data for short.

The original maps created by Reed (2008) are not geo-referenced, and the results shown here are based on digitized and adjusted maps shared by Scott Abramson, and not publicly available. They are described in more detail in Abramson (2017).

These data obviously measure something very similar to Euratlas. They also measure borders at a higher temporal frequency than the Euratlas data. On the other hand, they cover a smaller area, and only up to 1790.

Because these data are proprietary we do not use them in our benchmark regressions. Rather, the exercise undertaken here is to compare the results when using the Abramson data for the same, or adjacent, years as those for which we have Euratlas data.

To that end, and because the Abramson data end in 1790, we first compute border frequency across the years 1500, 1600, 1700, and 1790 from the Abramson data. This gives us a border frequency variable defined over a total of 3861 cells overlapping with our benchmark Euratlas data, which we can compare to the corresponding Euratlas border frequency index based on the years 1500, 1600, 1700, and 1800. The two border frequency measures have a correlation coefficient of 0.80 across these 3861 cells.

5.4.1 Geography and borders

Table A.24 shows the results when regressing the Abramson and Euratlas border frequency measures on our benchmark set of geography controls. All specifications include latitude fixed effects. Column (1) shows the results for the Euratlas measure based on all 5202 cells. The results are similar to those based on the same source for the years 1500-2000; see column (1) in Table 3 in the paper.

Column (2) again uses the Euratlas 1500-1800 measure as the dependent variable, but on a restricted sample of 3861 cells on which the Abramson measure is defined. The results do not change much compared to the baseline Euratlas sample in column (1), although the size of some estimates change a little and coastline density comes out as slightly more significant.

Column (3) uses the Abramson measure as the dependent variable. The results can be compared to those in column (2), based on the same sample but using the Euratlas measure. Both the magnitude and the precision of the estimated coefficients are very similar. The one exception is the coefficient on the dummy for mountains exceeding 2000 meters, which comes out as smaller and insignificant when using the Abramson measure.

The reason seems to be that there are differences in the coordinate alignments between the Euratlas maps and the digitized Centennia maps on which the Abramson measure is based. Therefore, the same border is sometimes allocated to different cells. This does not matter for most geography variables, since they are highly clustered: neighboring cells have very similar geography. However, it does matter for geographical features that apply to relatively few cells, such has the dummy for mountains (mean elevation of the cell) over 2000 meters. Intuitively, if a border gets pushed into a neighboring cell without such a mountain, the coefficient on that mountain indicator becomes less precisely estimated, but if it gets pushed into a cells with just a little less rainfall the precision of the estimate is not affected much.

Because of the variation in terms of coordinate alignments it is not meaningful to compare results in terms of local deviations.

We also want to explore how the results vary when HRE adjusting border frequency in the Euratlas and Abramson data, respectively. The Abramson data do not contain any maps of the HRE as a supranational entity, but we can adjust the Abramson data using the Euratlas HRE borders. This is not a clean comparison of the two datasets, but it might at least give us an idea about how robust our result is that the link between borders and modern development is weakened when treating the HRE as unified.

Columns (4)-(6) of Table A.24 present the regression results corresponding to those in columns (1)-(3), using this method to HRE adjust border frequency. Again, the estimates for the 2000 meter mountain dummy come out as significant when using the HRE adjusted Euratlas measure, both when using the full sample in column (4), and with the restricted sample in column (5), but not when using the HRE adjusted Abramson measure as the dependent variable. The other coefficients are overall relatively similar.

5.4.2 Borders and modern outcomes

Panel A of Table A.25 presents the outcomes when regressing modern outcomes on the Abramson and Euratlas border frequency measures, with our benchmark set of geography controls and latitude fixed effects. Both measures of border frequency show positive correlation with night lights and population density.

As seen in Panel B, we find no significant correlation between the HRE adjusted border measure and modern development when using the Euratlas border data. This is consistent with the results in the paper. However, when HRE adjusting the Abramson border data (using the Euratlas HRE borders, as described in Section 5.4.1 above) we do find a positive correlation between border frequency and both night lights and population density. In other words, keeping in mind the caveats discussed above, state fragmentation might have a positive correlation with modern development even when we interpret the Holy Roman Empire as unified.

6 Altering the size of the cells

The data considered in the paper used cells of size 0.5×0.5 degrees as unit of observation, where (recall) one degree is about 111 kilometers at the equator. This section presents results based on two alternative datasets, where each cell has size 0.1×0.1 and 1×1 degrees, respectively.

Following the same procedure as with the benchmark cell size data, we define a border as a cell with at least two states in a given year, and calculate a cell's border frequency as the average number of years it has a border 1500 to 2000, considering only cells with statehood in all years 1500 to 2000. For each cell, we also reconstruct our benchmark set of geography variables, and the two measures of modern outcomes used in the paper, log night lights and log population density, following the same procedures as when using 0.5×0.5 degree cells. For example, the variables that we call 1000 and 2000 meter mountain dummies are indicators of whether the mean elevation of a cell exceeds that height, and log ruggedness is the logarithm of one plus the standard deviation in elevation across the cell.

The 1×1 degree dataset consists of 1516 cells, with no missing observations.

In the 0.1×0.1 degree dataset, the number of observations varies across variables. Recall that our benchmark 0.5×0.5 degree dataset had 5202 observations for most variables (5201 for population density). Each cell with sides 0.5 degrees contains 25 cells with sides 0.1 degrees, so we could in principle have as many as $5202 \times 25 = 130,050$ cells. However, many of these smaller cells are either sea cells or lack statehood, so the number of observations is smaller for most variables. For example, for the border frequency index and most geography variables we have $108,374$ observations (about 17% less than the "potential" number), and for the the agricultural suitability variables the number of observations falls to $107,945$. When considering local deviations, the number of observations declines further, because the neighbors of some small cells have missing observations for all neighboring cells.

Tables A.26-A.28 report the results from a number of regressions using each of these datasets. We use ordinary least squares and report standardized coefficients, so as to facilitate comparison across regressions based on the different datasets. Adjusting the standard errors for spatial correlation is made practically impossible in the 0.1×0.1 degree dataset, due to the large number of observations. To keep the regression results comparable, we thus report robust standard errors for all regressions.

For the 0.1×0.1 and 0.5×0.5 degree datasets we include fixed effects for half-degree latitudes, i.e., a dummy for each row in the 0.5-degree cell grid, and thus only every fifth row in the 0.1-degree cell grid. This serves to keep the regressions as comparable as possible to those in the paper, and across these two datasets. For the 1×1 degree regressions we enter somewhat sparser fixed effects, namely every one-degree latitude, which is the finest fixed-effects structure we can use with those data. The results reported below do not change qualitatively if we use the sparser one-degree fixed effects across all regressions.

Table A.26 shows the results for each dataset when regressing border frequency on the benchmark set of geography controls, alternating between using a mountain dummy for 2000 and 1000 meters. Columns (1) and (2) report the same regressions as in columns (1) and (2) of Table 3 in the paper, but with standardized coefficients, and standard errors not being corrected for spatial correlation.

The results in Table A.26 are overall quite similar, but not identical, across datasets. Differences in the size of the estimated coefficients across datasets are significantly different from zero in some cases, at least when using the standard errors reported, which (recall) are not corrected for spatial correlation. However, the signs are mostly consistent.

One exception is that the coefficient on agricultural suitability comes out as negative with smaller cell size, and positive with larger cells. However, the agricultural suitability data from GAEZ are reported at a level of disaggregation of approximately 0.1 degrees, so we want to be careful about interpreting this coefficient when using 0.1×0.1 degree data. Similarly, for some of the smaller cells ruggedness (i.e., the standard deviation in elevation) is calculated on very few within-cell observation points, never more than four.

However, even disregarding how the geography data are constructed, it is hardly surprising that the results vary with cell size to some extent. First, some variation may occur within cells, but not across cells; what variation the regressions pick up depends on cell size. Second, some cells which were coded as having a border each year 1500-2000, such as those by the English Channel separating France and England, become coded as having no border in any year when using the smaller cell size. Vice versa, some non-border cells become border cells when using larger cells. In other words, relatively small changes in cell size can have a large effects in terms of the outcome variable. Third, recall again that the results reported here are not adjusted for spatial correlation, making some coefficient estimates appear more precisely estimated compared to the corresponding regressions in the paper.

Table A.27 presents results from the same regressions as in Table A.26, but in terms of local deviations. Again, the results are broadly similar across datasets, but not identical. For example, local deviations in ruggedness comes out as less significant with smaller cells. As discussed, this may not be too surprising.

Table A.28 presents some results when regressing modern outcomes on the border frequency index, latitude fixed effects (as explained), and our benchmark set of geography controls, in two panels: Panel A shows the cross section of all cells, and Panel B local deviations. Columns (1), (2), (7) and (8) correspond to columns (3) and (5) of Tables 6 and 7 in the paper, but here reporting standardized coefficients.

Border frequency comes out as positive and significant throughout in Panel A, both with larger and smaller cell size, and when dropping coastal cells and not dropping them. The results in Panel B show a negative local effect from borders, almost regardless of cell size and whether dropping coastal cells or not. The only exception is that local deviations in border frequency do not show a significant correlation with local deviations with modern outcomes in the 1×1 degree data, and when considering all 1516 cells, but the negative significant correlation is restored when dropping coastal cells; cf. columns (5) and (6), and columns

(11) and (12), in Panel B. The size of the coefficients also varies to some extent but this is arguably not strange.

7 Different treatments of night lights and population density

7.1 Effects on night lights per capita

In the paper modern outcomes were measured by either population density or night lights per land area. This makes sense, since by and large they capture the same dimension of economic and urban development. For example, it is easy to see from the map in Figure 9 in the paper that some large cities—we can identify, e.g., Moscow, Paris, and Madrid in the map—also record some of the highest levels of night lights.

We can also study the effects of borders on night lights per capita. The easiest way to do this is to use log night lights as the dependent variable and enter log population density as control. This approximately corresponds to using night lights per capita as the dependent variable, while controlling for population density.[26]

Table A.29 shows the results from a set of regressions identical to those in Table 6 in the paper, except that we now enter log population density as control. Panel A shows the global effects while Panel B shows the results in terms of local deviations.

Most of the correlations between border frequency and log night lights now come out as insignificant, but in both panels the more significant effects carry the same signs as in the paper. The global regressions in Panel A show a positive effect of border frequency in columns (2) and (3), i.e., when entering latitude fixed effects and geography controls. Panel B shows mostly negative effects, significant at the 1% level in column (5) when dropping coastal cells, and at the 10% level in column (6) when adding longitude fixed effects.

In a mechanical sense, it is not too surprising that these results are weaker, since population and night lights are so highly correlated, both globally across all cells, and in terms of local deviations from neighboring cells. However, whatever correlation there is seems to suggest that border frequency has qualitatively similar effects on night lights per capita as it has on night lights per land area and on population density.

[26]To be very precise, it is the same as using the ratio of one plus night lights over one plus population as the dependent variable, controlling for log of one plus population.

7.2 Alternative log transformations

Log night lights in the text actually refers to the log of *one plus* night lights (where night lights are measured as the average across the cell's pixels and over the years 1992-2013). Another commonly used measure is to take the log of 0.01 plus night lights; see, e.g., Michalopoulos and Papaioannou (2013, 2014) and Dickens (2017). The motivation is that it generates higher correlation with GDP per capita in cross-country data.

The regression results using this measure as the dependent variable are shown in Table A.30, applying the exact same specifications as in the top panels of Tables 6 and 7 in the paper. The results hardly change at all in terms of which coefficients come out as significant, but it is interesting to note that they all have larger magnitude.

The log transformation using the constant 0.01 may indeed be more common in the literature. In our benchmark regressions, we rather use one as the constant, to be consistent with the log transformation used for population densities, and because we feel uncertain about how often the other approach is used in grid-cell level analysis. However, it is comforting to know that the results for night lights are not sensitive to the choice of log transformation.

7.3 Controlling for urbanization

Section 5.3.4 in the paper correlated urbanization with borders in various panel regressions, using data from the History Database of the Global Environment (HYDE). In Table A.31, we examine how robust the correlations between modern outcomes and border frequency are to controlling for urbanization in 1500, or average urbanization 1500-2000, as well as the benchmark set of geography controls and latitude fixed effects. For comparison, columns (1) and (4) of Table A.31 report the coefficients without any urbanization controls, identical to those in column (3) of Panels A and B in Table 6 in the paper.

We do not know what the causal relationship between borders and urbanization is. It seems likely that both could affect one another, as well as our measures of modern outcomes. However, we note that border frequency stays significant when including either of these urbanization controls.

References

1. Abramson, S.F. (2017). "The Economic Origins of the Territorial State." *International Organization*, 71, 97-130.

2. Alesina, A. and E. Spolaore (2003). *The Size of Nations*. Cambridge: The MIT Press.

3. Conley, T. G. (1999). "GMM Estimation With Cross Sectional Dependence." *Journal of Econometrics*, 92, 1-45.

4. Dickens, A. (2017). "Ethnolinguistic Favoritism in African Politics." *Mimeo.* York University.

5. Galor, O. and Ö. Özak (2016). "The Agricultural Origins of Time Preference." *American Economic Review*, 106, 3064-3103.

6. Leff, B., N. Ramankutty, and J.A. Foley (2004). "Geographic Distribution of Major Crops Across the World." *Global Biogeochemical Cycles*, 18, 1-27.

7. Michalopoulos, S. (2012). "The Origins of Ethnolinguistic Diversity." *American Economic Review*, 102, 1508-1539.

8. Michalopoulos, S., and E. Papaioannou (2013). "Pre-Colonial Ethnic iInstitutions and Contemporary African Development." *Econometrica* 81, 113–152.

9. Michalopoulos, S., and E. Papaioannou (2014). "National Institutions and Subnational Development in Africa." *Quarterly Journal of Economics*, 129, 151–213.

10. Nunn, N., and N. Qian (2011). "The Potato's Contribution to Population and Urbanization: Evidence from a Historical Experiment." *Quarterly Journal of Economics*, 126, 593–650.

11. Reed, F. (2008). *Centennia Historical Atlas.* http://www.historicalatlas.com. Accessed: 1 June 2013.

Table A.1: Cross-correlations between different border variables.

	Border 1500	Border 1600	Border 1700	Border 1800	Border 1900	Border 2000	Border freq. 1500-2000	Current border	Language border
Border dummy 1500	1.000								
Border dummy 1600	0.499	1.000							
	(0.000)								
Border dummy 1700	0.538	0.630	1.000						
	(0.000)	(0.000)							
Border dummy 1800	0.419	0.395	0.581	1.000					
	(0.000)	(0.000)	(0.000)						
Border dummy 1900	0.230	0.224	0.342	0.385	1.000				
	(0.000)	(0.000)	(0.000)	(0.000)					
Border dummy 2000	0.228	0.217	0.291	0.305	0.458	1.000			
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)				
Border frequency 1500-2000	0.719	0.722	0.815	0.721	0.600	0.599	1.000		
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)			
Current border dummy	0.215	0.200	0.269	0.287	0.433	0.913	0.555	1.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		
Language border dummy	0.080	0.107	0.120	0.099	0.227	0.510	0.274	0.526	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Notes: Unconditional pairwise correlation coefficients between the different border variables, with p-values in parentheses. Border frequency is the average of the border dummies 1500-2000 (B_i).

Table A.2: Cross-correlations between border frequency, the benchmark geography variables, and modern outcomes.

	Border frequency	Mountain >2000m	Mountain >1000m	Log ruggedn.	River density	Ag. suit. rainfed	Ag. suit. irrigated	Rainfall	Log dist. to coast	Coastline density	Log night lights	Log pop. dens
Border freq.	1.000											
Mountain> 2000m	0.085	1.000										
	(0.000)											
Mountain> 1000m	0.086	0.382	1.000									
	(0.000)	(0.000)										
Log rugg.	0.147	0.204	0.474	1.000								
	(0.000)	(0.000)	(0.000)									
River dens.	0.114	−0.009	−0.024	−0.085	1.000							
	(0.000)	(0.498)	(0.084)	(0.000)								
Ag. suit. rainf.	0.110	−0.066	−0.076	−0.072	0.144	1.000						
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)							
Ag. suit. irr.	−0.038	−0.053	−0.123	−0.086	0.146	0.544	1.000					
	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)						
Rainfall	0.255	−0.010	−0.004	0.221	0.025	0.202	0.062	1.000				
	(0.000)	(0.468)	(0.761)	(0.000)	(0.067)	(0.000)	(0.000)					
Log dist. coast	−0.041	0.010	−0.038	−0.246	0.104	0.340	0.178	−0.217	1.000			
	(0.003)	(0.451)	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)				
Coastline dens.	−0.071	−0.029	−0.077	−0.161	−0.044	−0.257	−0.162	0.090	−0.194	1.000		
	(0.000)	(0.034)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)			
Log night lights	0.172	−0.088	−0.182	0.055	0.136	0.133	0.296	0.283	−0.263	0.012	1.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.391)		
Log pop. dens.	0.185	−0.025	−0.045	0.152	0.142	0.215	0.218	0.237	−0.278	0.081	0.818	1.000
	(0.000)	(0.076)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Notes: Unconditional pairwise correlation coefficients between border frequency (B_i), the benchmark geography variables, and modern outcomes, with p-values in parentheses.

Table A.3: Cross-correlations between the change in border frequency when HRE adjusting and the benchmark geography variables.

Mountain >2000m	0.006
	(0.660)
Mountain >1000m	0.023
	(0.099)
Log ruggedness	−0.005
	(0.718)
River density	−0.062
	(0.000)
Ag. suit. rainfed	−0.030
	(0.030)
Ag. suit. irrig.	0.091
	(0.000)
Rainfall	−0.129
	(0.000)
Log distance to coast	−0.007
	(0.635)
Coastline density	0.037
	(0.007)

Notes: Unconditional pairwise correlation coefficients between the benchmark geography variables and the change in border frequency when treating the Holy Roman Empire as unified; *p*-values in parentheses.

Table A.4: Geography and border frequency: controlling for elevation.

	Dependent variable: Border frequency 1500-2000					
	(1)	(2)	(3)	(4)	(5)	(6)
Log elevation	0.039***	0.035***	0.030***	−0.004	0.005	0.011
	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
Mountain >2000m		0.112***		0.125***	0.122***	0.142***
		(0.041)		(0.041)	(0.040)	(0.034)
Mountain >1000m			0.053**			
			(0.022)			
Log ruggedness				0.046***	0.024***	0.012*
				(0.008)	(0.008)	(0.007)
River density					1.808***	1.391**
					(0.649)	(0.670)
Ag. suit. rainfed					0.002	0.081***
					(0.031)	(0.030)
Ag. suit. irrig.					−0.103***	−0.098***
					(0.022)	(0.019)
Rainfall					0.040***	0.039***
					(0.013)	(0.013)
Log dist. to coast					−0.129**	−0.060
					(0.053)	(0.057)
Coastline density					0.021*	0.022**
					(0.011)	(0.011)
Log land area	0.002	0.003	0.003	0.006	0.027***	0.026***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)
R^2	0.14	0.14	0.14	0.16	0.19	0.34
Number of obs.	5202	5202	5202	5202	5202	5202
Fixed effects	Latitude	Latitude	Latitude	Latitude	Latitude	Lat./Long.

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. * indicates $p<0.10$, ** $p<0.05$, and *** $p<0.01$.

Table A.5: Geography and border frequency: coal, temperature, and fraction lake.

	Dependent variable: Border frequency 1500-2000							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coal dummy	0.036**			0.036**	0.002			0.001
	(0.017)			(0.017)	(0.016)			(0.016)
Temperature		−0.005*		−0.005*		0.004		0.004
		(0.003)		(0.003)		(0.003)		(0.003)
Fraction lake			0.016	0.032			0.116	0.118
			(0.088)	(0.088)			(0.083)	(0.082)
Mountain >2000m					0.127***	0.149***	0.127***	0.149***
					(0.040)	(0.042)	(0.040)	(0.042)
Log ruggedness					0.027***	0.031***	0.027***	0.031***
					(0.007)	(0.007)	(0.007)	(0.007)
River density					1.805***	1.802***	1.817***	1.812***
					(0.644)	(0.637)	(0.648)	(0.642)
Ag. suit. rainfed					0.004	0.010	0.006	0.012
					(0.030)	(0.031)	(0.030)	(0.031)
Ag. suit. irrig.					−0.104***	−0.108***	−0.104***	−0.108***
					(0.022)	(0.022)	(0.022)	(0.022)
Rainfall					0.040***	0.038***	0.041***	0.039***
					(0.013)	(0.013)	(0.013)	(0.013)
Log dist. to coast					−0.116**	−0.096*	−0.117**	−0.097*
					(0.052)	(0.051)	(0.052)	(0.051)
Coastline density					0.023**	0.026**	0.023**	0.026**
					(0.010)	(0.011)	(0.011)	(0.011)
Log land area	0.019***	0.015***	0.019***	0.015***	0.028***	0.029***	0.027***	0.028***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)	(0.006)
R^2	0.12	0.12	0.12	0.12	0.19	0.19	0.19	0.19
Number of obs.	5202	5202	5202	5202	5202	5202	5202	5202

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. All specifications include latitude fixed effects. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.6: Geography and border frequency: alternative measures of agricultural suitability.

	Dependent variable: Border frequency 1500-2000									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Mountain >2000m	0.123***	0.136***	0.134***	0.125***	0.137***	0.134***	0.121***	0.133***	0.129***	0.149***
	(0.039)	(0.040)	(0.041)	(0.039)	(0.041)	(0.041)	(0.040)	(0.040)	(0.041)	(0.033)
Log ruggedness	0.027***	0.030***	0.025***	0.025***	0.030***	0.025***	0.022***	0.027***	0.022***	0.016***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)
River density	1.668***	1.636***	1.623***	1.638***	1.659***	1.615***	1.770***	1.791***	1.747***	1.368**
	(0.590)	(0.576)	(0.579)	(0.584)	(0.582)	(0.578)	(0.641)	(0.639)	(0.635)	(0.661)
Ag. suit. potato	−0.108***		−0.149***	−0.154***		−0.165***	−0.160***		−0.169***	−0.098**
	(0.033)		(0.039)	(0.041)		(0.042)	(0.041)		(0.042)	(0.041)
Caloric suit. index		−0.003	0.091*		0.062	0.082		0.054	0.074	−0.027
		(0.041)	(0.049)		(0.052)	(0.053)		(0.052)	(0.053)	(0.050)
Ag. suit. rainfed				0.053	−0.078**	0.023	0.111***	−0.021	0.083**	0.157***
				(0.035)	(0.036)	(0.038)	(0.037)	(0.037)	(0.040)	(0.037)
Ag. suit. irrig.							−0.105***	−0.103***	−0.104***	−0.100***
							(0.022)	(0.022)	(0.022)	(0.019)
Rainfall	0.041***	0.043***	0.042***	0.041***	0.042***	0.042***	0.040***	0.041***	0.040***	0.039***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Log dist. to coast	−0.074	−0.141***	−0.060	−0.071	−0.111**	−0.060	−0.068	−0.111**	−0.059	0.009
	(0.056)	(0.053)	(0.055)	(0.056)	(0.053)	(0.056)	(0.054)	(0.052)	(0.053)	(0.057)
Coastline density	0.020*	0.025**	0.018*	0.018*	0.025**	0.017*	0.015	0.023**	0.015	0.021**
	(0.010)	(0.011)	(0.011)	(0.010)	(0.011)	(0.011)	(0.010)	(0.011)	(0.010)	(0.010)
Log land area	0.023***	0.023***	0.023***	0.019***	0.028***	0.022***	0.020***	0.030***	0.023***	0.023***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
R^2	0.19	0.18	0.19	0.19	0.18	0.19	0.20	0.19	0.20	0.34
Number of obs.	5202	5202	5202	5202	5202	5202	5202	5202	5202	5202
Fixed effects	Latitude	Latitude	Latitude	Latitude	Latitude	Latitude	Latitude	Latitude	Latitude	Lat./Long.

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. * indicates $p<0.10$, ** $p<0.05$, and *** $p<0.01$.

Table A.7: Geography and border frequency using river and coast dummies.

	Dependent variable: Border frequency 1500-2000					
	(1)	(2)	(3)	(4)	(5)	(6)
Mountain >2000m	0.122***		0.129***	0.111***	0.146***	
	(0.040)		(0.037)	(0.039)	(0.033)	
Mountain >1000m		0.041**				0.063***
		(0.019)				(0.018)
Log ruggedness	0.025***	0.023***	0.016***	0.028***	0.017***	0.014**
	(0.007)	(0.007)	(0.006)	(0.009)	(0.006)	(0.005)
River dummy	0.076***	0.076***	0.052***	0.069***	0.070***	0.070***
	(0.010)	(0.010)	(0.009)	(0.010)	(0.009)	(0.009)
Ag. suit. rainfed	−0.006	−0.011	0.034	−0.065*	0.076***	0.074**
	(0.030)	(0.030)	(0.025)	(0.036)	(0.029)	(0.029)
Ag. suit. irrig.	−0.106***	−0.102***	−0.044***	−0.125***	−0.104***	−0.097***
	(0.022)	(0.022)	(0.017)	(0.024)	(0.018)	(0.018)
Rainfall	0.043***	0.042***	0.032***	0.092***	0.042***	0.042***
	(0.013)	(0.013)	(0.010)	(0.020)	(0.012)	(0.012)
Log dist. to coast	−0.203***	−0.196***	−0.114**	−0.154**	−0.110**	−0.103*
	(0.056)	(0.056)	(0.047)	(0.061)	(0.056)	(0.056)
Coast dummy	−0.053***	−0.049***	−0.028**		−0.043***	−0.035***
	(0.013)	(0.014)	(0.012)		(0.012)	(0.012)
Log land area	0.009**	0.009**	0.006		0.011**	0.012***
	(0.004)	(0.004)	(0.004)		(0.004)	(0.004)
R^2	0.21	0.20	0.14	0.26	0.35	0.35
Number of obs.	5202	5202	5202	3869	5202	5202
Fixed effects	Latitude	Latitude	Latitude	Latitude	Lat./Long.	Lat./Long.
HRE adjustment	No	No	Yes	No	No	No
Drop coastal cells	No	No	No	Yes	No	No

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.8: Geography and border frequency: local deviations using river and coast dummies.

	Dependent variable: Δ Border frequency 1500-2000					
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Mountain >2000m	0.049					
	(0.037)					
Δ Mountain >1000m		0.071***	0.072***	0.075***	0.078***	0.073***
		(0.018)	(0.018)	(0.018)	(0.019)	(0.019)
Δ Log ruggedness	0.019***	0.018***	0.017***	0.015***	0.020***	0.017***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.005)
Δ River dummy	0.036***	0.037***	0.038***	0.038***	0.036***	0.038***
	(0.008)	(0.007)	(0.008)	(0.008)	(0.007)	(0.008)
Δ Ag. suit. rainfed	−0.104***	−0.098***	−0.109***	−0.101***	−0.133***	−0.099***
	(0.039)	(0.037)	(0.036)	(0.036)	(0.046)	(0.035)
Δ Ag. suit. irrig.	0.007	0.014	0.015	0.022	0.004	0.017
	(0.018)	(0.019)	(0.018)	(0.019)	(0.021)	(0.018)
Δ Rainfall	0.071**	0.072***	0.073***	0.068**	0.086*	0.080***
	(0.028)	(0.028)	(0.028)	(0.027)	(0.044)	(0.029)
Δ Log dist. to coast	−0.169	−0.242	−0.259	−0.182	−0.036	−0.232
	(0.320)	(0.312)	(0.313)	(0.312)	(0.332)	(0.318)
Δ Coast dummy	−0.002	0.002	0.004	0.004		0.004
	(0.010)	(0.011)	(0.010)	(0.010)		(0.011)
Δ Log land area	0.011***	0.010***	0.011***	0.010***		0.011***
	(0.004)	(0.004)	(0.004)	(0.004)		(0.004)
R^2	0.02	0.03	0.04	0.04	0.05	0.06
Number of obs.	5202	5202	5202	5202	3869	5202
Fixed effects	None	None	Latitude	Latitude	Latitude	Lat./Long.
HRE adjustment	No	No	No	Yes	No	No
Drop coastal cells	No	No	No	No	Yes	No

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. * indicates $p<0.10$, ** $p<0.05$, and *** $p<0.01$.

Table A.9: Borders and modern outcomes: more geography controls.

	Dependent variable:							
	Log night lights				Log population density			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Border frequency 1500-2000	0.371***	0.375***	0.340***	0.344***	0.540***	0.534***	0.503***	0.500***
	(0.084)	(0.084)	(0.081)	(0.081)	(0.120)	(0.115)	(0.116)	(0.114)
Mountain >2000m	−0.429***	−0.284***	0.384***	0.478***	−0.302***	0.032	0.665***	0.893***
	(0.096)	(0.087)	(0.103)	(0.098)	(0.115)	(0.114)	(0.135)	(0.140)
Log ruggedness	0.054**	0.067***	0.177***	0.179***	0.110***	0.128***	0.267***	0.263***
	(0.023)	(0.024)	(0.025)	(0.025)	(0.037)	(0.038)	(0.041)	(0.040)
River density	6.966***	6.761***	6.859***	6.656***	11.746***	11.175***	11.685***	11.104***
	(1.466)	(1.373)	(1.260)	(1.189)	(3.436)	(3.173)	(3.184)	(2.954)
Ag. suit. rainfed	−0.017	−0.890***	0.192*	−0.592***	1.290***	−0.496**	1.547***	−0.143
	(0.116)	(0.160)	(0.109)	(0.157)	(0.178)	(0.217)	(0.171)	(0.217)
Ag. suit. irrig.	1.151***	1.175***	0.987***	1.016***	0.700***	0.752***	0.506***	0.572***
	(0.077)	(0.076)	(0.075)	(0.075)	(0.109)	(0.106)	(0.107)	(0.105)
Ag. suit. potato		0.525***		0.439***		0.793***		0.711***
		(0.173)		(0.167)		(0.239)		(0.235)
Caloric suit. index		1.126***		1.032***		2.709***		2.559***
		(0.175)		(0.171)		(0.262)		(0.255)
Rainfall	0.191***	0.201***	0.118**	0.128***	0.181***	0.205***	0.107*	0.130**
	(0.046)	(0.043)	(0.047)	(0.044)	(0.064)	(0.056)	(0.064)	(0.057)
Log dist. to coast	−2.108***	−2.164***	−1.392***	−1.463***	−3.180***	−3.176***	−2.336***	−2.392***
	(0.196)	(0.200)	(0.204)	(0.205)	(0.281)	(0.285)	(0.296)	(0.295)
Coastline density	−0.504***	−0.477***	−0.389**	−0.372**	−0.227**	−0.185*	−0.093	−0.067
	(0.148)	(0.144)	(0.161)	(0.157)	(0.102)	(0.095)	(0.114)	(0.105)
Coal dummy			0.104**	0.134***			−0.037	0.039
			(0.046)	(0.045)			(0.068)	(0.064)
Temperature			0.131***	0.125***			0.157***	0.143***
			(0.011)	(0.011)			(0.016)	(0.015)
Fraction lake			0.312	0.275			0.456	0.251
			(0.227)	(0.251)			(0.335)	(0.377)
Log land area	−0.238***	−0.176***	−0.198***	−0.145***	−0.351***	−0.223***	−0.304***	−0.188***
	(0.038)	(0.038)	(0.039)	(0.039)	(0.046)	(0.042)	(0.046)	(0.042)
R^2	0.35	0.37	0.40	0.42	0.39	0.43	0.42	0.45
Number of obs.	5202	5202	5202	5202	5201	5201	5201	5201

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. All specifications include latitude fixed effects. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.10: Borders and modern outcomes: local deviations with more geography controls.

	Dependent variable:							
	Δ Log night lights				Δ Log population density			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Border frequency 1500-2000	−0.179***	−0.161***	−0.080	−0.085*	−0.280***	−0.258***	−0.139*	−0.146**
	(0.053)	(0.052)	(0.051)	(0.051)	(0.076)	(0.076)	(0.073)	(0.073)
Δ Mountain >2000m	−0.205***		0.237***		−0.370***		0.261***	
	(0.059)		(0.058)		(0.082)		(0.086)	
Δ Mountain >1000m		−0.262***		0.130***		−0.345***		0.196***
		(0.041)		(0.045)		(0.059)		(0.066)
Δ Log ruggedness	−0.044*	−0.039*	0.023	0.024	0.031	0.036	0.116***	0.118***
	(0.023)	(0.022)	(0.022)	(0.022)	(0.029)	(0.029)	(0.029)	(0.030)
Δ River density	5.032***	4.956***	4.622***	4.665***	6.515***	6.411***	5.883***	5.927***
	(1.377)	(1.346)	(1.112)	(1.121)	(2.299)	(2.258)	(1.923)	(1.928)
Δ Ag. suit. rainfed	−0.023	−0.040	−0.594***	−0.611***	0.114	0.105	−0.644***	−0.663***
	(0.106)	(0.104)	(0.130)	(0.129)	(0.150)	(0.150)	(0.172)	(0.171)
Δ Ag. suit. irrig.	0.698***	0.670***	0.587***	0.601***	0.623***	0.585***	0.476***	0.492***
	(0.065)	(0.065)	(0.062)	(0.062)	(0.096)	(0.095)	(0.090)	(0.090)
Δ Ag. suit. potato			0.688***	0.718***			0.724***	0.760***
			(0.151)	(0.150)			(0.201)	(0.199)
Δ Caloric suit. index			0.346	0.266			0.996***	0.909**
			(0.227)	(0.227)			(0.375)	(0.371)
Δ Rainfall	−0.081	−0.084	−0.046	−0.049	−0.139	−0.137	−0.113	−0.114
	(0.078)	(0.077)	(0.070)	(0.071)	(0.116)	(0.115)	(0.108)	(0.109)
Δ Log dist. to coast	−11.150***	−10.681***	−7.661***	−7.756***	−15.965***	−15.392***	−11.318***	−11.411***
	(1.137)	(1.136)	(1.100)	(1.102)	(1.837)	(1.841)	(1.742)	(1.744)
Δ Coastline density	−0.253***	−0.248***	−0.199**	−0.200**	0.102*	0.109**	0.172***	0.172***
	(0.091)	(0.090)	(0.080)	(0.080)	(0.053)	(0.053)	(0.055)	(0.055)
Δ Coal dummy			0.064**	0.062**			0.089*	0.086*
			(0.030)	(0.030)			(0.046)	(0.046)
Δ Temperature			0.150***	0.155***			0.195***	0.206***
			(0.013)	(0.013)			(0.019)	(0.020)
Δ Fraction lake			−0.544**	−0.538**			−0.981***	−0.971***
			(0.243)	(0.244)			(0.317)	(0.320)
Δ Log land area	−0.041	−0.037	0.021	0.021	−0.097***	−0.092***	−0.013	−0.012
	(0.029)	(0.028)	(0.026)	(0.026)	(0.029)	(0.028)	(0.028)	(0.028)
R^2	0.13	0.14	0.20	0.19	0.08	0.08	0.13	0.13
Number of obs.	5202	5202	5202	5202	5201	5201	5201	5201

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. All specifications include latitude fixed effects (not in local deviations). * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.11: Geography and border frequency: sparser latitude and longitude fixed effects.

	Dependent variable: border frequency 1500-2000				
	(1)	(2)	(3)	(4)	(5)
Mountain >2000m	0.122***	0.127***	0.123***	0.153***	0.148***
	(0.042)	(0.040)	(0.040)	(0.033)	(0.033)
Log ruggedness	0.009*	0.027***	0.027***	0.019***	0.019***
	(0.005)	(0.007)	(0.007)	(0.006)	(0.006)
River density	1.884***	1.806***	1.785***	1.382**	1.420**
	(0.682)	(0.644)	(0.641)	(0.658)	(0.641)
Ag. suit. rainfed	0.090***	0.004	0.009	0.084***	0.076***
	(0.028)	(0.030)	(0.030)	(0.029)	(0.029)
Ag. suit. irrig.	−0.108***	−0.104***	−0.101***	−0.101***	−0.098***
	(0.022)	(0.022)	(0.022)	(0.019)	(0.019)
Rainfall	0.068***	0.040***	0.042***	0.039***	0.039***
	(0.011)	(0.013)	(0.013)	(0.013)	(0.013)
Log dist. to coast	−0.069	−0.116**	−0.110**	−0.026	−0.010
	(0.047)	(0.052)	(0.052)	(0.053)	(0.053)
Coastline density	0.023*	0.023**	0.024**	0.026**	0.027***
	(0.014)	(0.010)	(0.010)	(0.010)	(0.010)
Log land area	0.035***	0.028***	0.028***	0.028***	0.029***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)
R^2	0.12	0.19	0.18	0.33	0.31
Number of obs.	5202	5202	5202	5202	5202
Fixed effects	None	Latitude	Latitude	Lat./Long.	Lat./Long.
Sparseness	NA	Half degree	Even degree	Half degree	Even degree

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. Columns (2) and (4) enter fixed effects for every half-degree latitude, and/or longitude, and columns (3) and (5) enter fixed effects for every even integer degree latitude and/or longitude. Column (1) enters no fixed effects at all. * indicates $p<0.10$, ** $p<0.05$, and *** $p<0.01$.

Table A.12: Geography and border frequency: different Conley cutoffs.

	Dependent variable: Border frequency 1500-2000			
	(1)	(2)	(3)	(4)
Mountain >2000m	0.127***		0.127***	
	(0.037)		(0.042)	
Mountain >1000m		0.052***		0.052**
		(0.016)		(0.022)
Log ruggedness	0.027***	0.025***	0.027***	0.025***
	(0.006)	(0.006)	(0.008)	(0.008)
River density	1.806***	1.794***	1.806***	1.794***
	(0.636)	(0.638)	(0.655)	(0.656)
Ag. suit. rainfed	0.004	−0.000	0.004	−0.000
	(0.025)	(0.025)	(0.035)	(0.035)
Ag. suit. irrig.	−0.104***	−0.098***	−0.104***	−0.098***
	(0.018)	(0.018)	(0.026)	(0.027)
Rainfall	0.040***	0.040***	0.040***	0.040***
	(0.011)	(0.011)	(0.015)	(0.015)
Log dist. to coast	−0.116***	−0.117***	−0.116*	−0.117*
	(0.041)	(0.041)	(0.063)	(0.063)
Coastline density	0.023**	0.020*	0.023**	0.020*
	(0.010)	(0.010)	(0.011)	(0.011)
Log land area	0.028***	0.026***	0.028***	0.026***
	(0.005)	(0.005)	(0.006)	(0.006)
R^2	0.19	0.19	0.19	0.19
Number of obs.	5202	5202	5202	5202
Conley cutoffs	1 degree	1 degree	2 degrees	2 degrees

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1 or 2 degrees of each other, as indicated. All specifications also include latitude fixed effects. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.13: Geography and border frequency: local deviations and different Conley cutoffs.

	Dependent variable: Δ Border frequency 1500-2000			
	(1)	(2)	(3)	(4)
Δ Mountain >1000m	0.070***	0.070***	0.065***	0.065***
	(0.017)	(0.020)	(0.016)	(0.019)
Δ Log ruggedness	0.019***	0.019***	0.025***	0.025***
	(0.005)	(0.005)	(0.005)	(0.005)
Δ River density	0.717***	0.717***	0.550*	0.550*
	(0.240)	(0.244)	(0.312)	(0.325)
Δ Ag. suit. rainfed	−0.109***	−0.109***	−0.080***	−0.080**
	(0.032)	(0.040)	(0.028)	(0.033)
Δ Ag. suit. irrig.	0.022	0.022	0.015	0.015
	(0.017)	(0.020)	(0.017)	(0.019)
Δ Rainfall	0.065**	0.065**	0.057***	0.057**
	(0.027)	(0.029)	(0.019)	(0.022)
Δ Log dist. to coast	−0.352	−0.352	−0.133	−0.133
	(0.271)	(0.275)	(0.213)	(0.213)
Δ Coastline density	0.015**	0.015*	0.011	0.011
	(0.008)	(0.008)	(0.007)	(0.007)
Δ Log land area	0.015***	0.015***	0.009**	0.009**
	(0.004)	(0.005)	(0.004)	(0.004)
R^2	0.04	0.04	0.04	0.04
Number of obs.	5202	5202	5202	5202
Conley cutoffs	1 degree	2 degree	1 degrees	2 degrees
Number of neighbors	8 cells	8 cells	24 cells	24 cells

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1 or 2 degrees of each other, as indicated. The dependent and independent variables are measured as local deviations from the 8 or 24 closest neighboring cells, as indicated. All specifications include latitude fixed effects (not in local deviations). * indicates $p<0.10$, ** $p<0.05$, and *** $p<0.01$.

Table A.14: Borders and modern outcomes: fixed effects for clusters of cells.

	Dependent variable:					
	Log night lights			Log population density		
	(1)	(2)	(3)	(4)	(5)	(6)
Border frequency 1500-2000	−0.124**	−0.143***	−0.119**	−0.207***	−0.245***	−0.218***
	(0.052)	(0.051)	(0.056)	(0.078)	(0.078)	(0.082)
R^2	0.73	0.68	0.63	0.73	0.67	0.62
Number of obs.	5202	5202	5202	5201	5201	5201
Cluster size	9 cells	16 cells	25 cells	9 cells	16 cells	25 cells

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. The sample is divided into different clusters containing 9, 16, and 25 neighboring cells, as explained in the text. Each regression enters a dummy for the cluster each cell belongs to, with the cluster size indicated. All specifications also include latitude fixed effects and the benchmark set of geography controls. * indicates $p<0.10$, ** $p<0.05$, and *** $p<0.01$.

Table A.15: Geography, borders by year, and current/language borders.

	Dependent variable:							
	Border dummy by year:						Dummy for:	
	1500	1600	1700	1800	1900	2000	Current bdr.	Language bdr.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mountain >2000m	0.086*	0.087	0.128**	0.173***	0.040	0.246***	0.228***	0.251***
	(0.050)	(0.053)	(0.053)	(0.062)	(0.055)	(0.071)	(0.069)	(0.053)
Log ruggedness	0.040***	0.030***	0.033***	0.024**	0.022***	0.013	0.004	0.020
	(0.009)	(0.010)	(0.009)	(0.010)	(0.008)	(0.010)	(0.011)	(0.013)
River density	2.381***	2.210***	1.688***	1.791***	1.338**	1.428**	1.454**	0.452
	(0.848)	(0.805)	(0.630)	(0.688)	(0.536)	(0.667)	(0.702)	(0.524)
Ag. suit. rainfed	0.007	0.009	0.002	0.009	−0.035	0.033	0.006	−0.008
	(0.048)	(0.044)	(0.044)	(0.039)	(0.035)	(0.044)	(0.046)	(0.060)
Ag. suit. irrig.	−0.155***	−0.120***	−0.077**	−0.152***	−0.025	−0.094***	−0.102***	−0.182***
	(0.035)	(0.033)	(0.035)	(0.030)	(0.022)	(0.030)	(0.031)	(0.042)
Rainfall	0.070***	0.043**	0.055***	0.023	0.023*	0.027*	0.021	0.050**
	(0.019)	(0.018)	(0.018)	(0.015)	(0.014)	(0.015)	(0.016)	(0.022)
Log dist. to coast	−0.254***	−0.266***	−0.214***	−0.050	0.049	0.039	0.015	0.187
	(0.082)	(0.081)	(0.083)	(0.066)	(0.058)	(0.080)	(0.081)	(0.121)
Coastline density	0.053**	0.019	0.014	0.011	0.016	0.026**	0.014	0.031
	(0.025)	(0.012)	(0.013)	(0.010)	(0.011)	(0.012)	(0.013)	(0.024)
Log land area	0.043***	0.028***	0.023***	0.021***	0.016***	0.035***	0.032***	0.052***
	(0.009)	(0.008)	(0.008)	(0.006)	(0.006)	(0.008)	(0.008)	(0.012)
R^2	0.16	0.14	0.14	0.11	0.08	0.08	0.08	0.11
Number of obs.	5202	5202	5202	5202	5202	5202	5202	5202

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. All specifications include latitude fixed effects. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.16: Geography, borders by year, and current/language borders: local deviations.

	Dependent variable:							
	Δ Border dummy by year:						Δ Dummy for:	
	1500	1600	1700	1800	1900	2000	Current bdr.	Language bdr.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Mountain >2000m	0.029	0.003	0.045	0.099*	−0.013	0.140***	0.098*	0.016
	(0.054)	(0.049)	(0.048)	(0.052)	(0.053)	(0.052)	(0.052)	(0.058)
Δ Log ruggedness	0.006	0.021***	0.016**	0.014*	0.029***	0.035***	0.026**	0.034***
	(0.010)	(0.008)	(0.007)	(0.008)	(0.008)	(0.010)	(0.010)	(0.012)
Δ River density	0.836**	0.981***	0.416*	0.909***	0.484*	0.577*	0.386	0.484
	(0.330)	(0.364)	(0.243)	(0.334)	(0.272)	(0.334)	(0.383)	(0.329)
Δ Ag. suit. rainfed	−0.095*	−0.128**	−0.109**	−0.117**	−0.139***	−0.100*	−0.123**	−0.074
	(0.056)	(0.055)	(0.047)	(0.050)	(0.049)	(0.056)	(0.053)	(0.057)
Δ Ag. suit. irrig.	−0.001	−0.017	0.044	−0.020	0.024	0.057	0.080**	0.069
	(0.032)	(0.028)	(0.029)	(0.027)	(0.026)	(0.039)	(0.036)	(0.042)
Δ Rainfall	0.065*	0.091**	0.081**	0.041	0.059*	0.050	0.072	0.067
	(0.039)	(0.036)	(0.037)	(0.033)	(0.035)	(0.041)	(0.044)	(0.047)
Δ Log dist. to coast	−0.026	−0.076	−0.499	−0.717**	−0.068	0.028	−0.751	−1.030
	(0.521)	(0.440)	(0.472)	(0.342)	(0.371)	(0.494)	(0.503)	(0.631)
Δ Coastline density	0.009	0.015	0.031**	0.015*	0.007	0.022**	0.006	0.039***
	(0.026)	(0.013)	(0.013)	(0.009)	(0.008)	(0.009)	(0.010)	(0.014)
Δ Log land area	0.035***	0.017**	0.022***	0.017**	−0.002	0.007	−0.000	0.020**
	(0.009)	(0.007)	(0.008)	(0.007)	(0.005)	(0.007)	(0.007)	(0.009)
R^2	0.03	0.03	0.03	0.03	0.02	0.03	0.03	0.03
Number of obs.	5202	5202	5202	5202	5202	5202	5202	5202

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. All specifications include latitude fixed effects (not in local deviations). * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.17: Border dummies and modern outcomes.

Panel A	Dependent variable: Log night lights					
	(1)	(2)	(3)	(4)	(5)	(6)
Border dummy	0.177***	0.330***	0.251***	0.320***	0.067	−0.113**
	(0.050)	(0.053)	(0.055)	(0.060)	(0.057)	(0.044)
R^2	0.34	0.35	0.35	0.35	0.34	0.34
Number of obs.	5202	5202	5202	5202	5202	5202
Panel B	**Dependent variable: Log population density**					
Border dummy	0.197***	0.385***	0.348***	0.487***	0.094	−0.015
	(0.067)	(0.072)	(0.076)	(0.082)	(0.076)	(0.059)
R^2	0.38	0.39	0.39	0.39	0.38	0.38
Number of obs.	5201	5201	5201	5201	5201	5201
Year to which border dummy refers	1500	1600	1700	1800	1900	2000

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. All specifications include latitude fixed effects and the benchmark set of geography controls. * indicates $p<0.10$, ** $p<0.05$, and *** $p<0.01$.

Table A.18: Border dummies and modern outcomes: local deviations.

Panel A	Dependent variable: Δ Log night lights					
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Border dummy	−0.058**	−0.046	−0.080**	−0.028	−0.084**	−0.106***
	(0.027)	(0.031)	(0.034)	(0.039)	(0.033)	(0.025)
R^2	0.13	0.13	0.13	0.13	0.13	0.13
Number of obs.	5202	5202	5202	5202	5202	5202
Panel B	**Dependent variable: Δ Log population density**					
Δ Border dummy	−0.115***	−0.078*	−0.098**	−0.047	−0.145***	−0.148***
	(0.039)	(0.046)	(0.049)	(0.058)	(0.048)	(0.039)
R^2	0.08	0.08	0.08	0.08	0.08	0.08
Number of obs.	5201	5201	5201	5201	5201	5201
Year to which Δ border dummy refers	1500	1600	1700	1800	1900	2000

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. All specifications include local deviations in the benchmark set of geography controls and latitude fixed effects (not in local deviations). * indicates $p<0.10$, ** $p<0.05$, and *** $p<0.01$.

Table A.19: Borders and modern outcomes: border variance as measure of border stability.

	Dependent variable:							
	Log night lights				Log population density			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Border frequency 1500-2000	0.563***		0.824***	0.648***	0.856***		0.796***	0.605***
	(0.096)		(0.125)	(0.104)	(0.139)		(0.195)	(0.161)
Border variance 1500-2000		0.519***	−0.766***	−0.821***		1.416***	0.174	−0.192
		(0.188)	(0.229)	(0.200)		(0.264)	(0.356)	(0.304)
R^2	0.28	0.26	0.28	0.35	0.28	0.27	0.28	0.39
Number of obs.	5202	5202	5202	5202	5201	5201	5201	5201
Geography controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	None	None	None	Latitude	None	None	None	Latitude

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.20: Cross-correlations between modern outcomes and alternative measures of state fragmentation.

	Log night lights	Population density	Border frequency	Log number of states	Log mean state size	Border density	Log border distance
Log night lights	1.000						
Log population density	0.818	1.000					
	(0.000)						
Border frequency 1500-2000	0.172	0.185	1.000				
	(0.000)	(0.000)					
Log number of states 1500-2000	0.213	0.210	0.935	1.000			
	(0.000)	(0.000)	(0.000)				
Log mean state size 1500-2000	−0.392	−0.246	−0.297	−0.351	1.000		
	(0.000)	(0.000)	(0.000)	(0.000)			
Border density 1500-2000	0.160	0.159	0.534	0.533	−0.237	1.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)		
Log border distance 1500-2000	−0.312	−0.367	−0.789	−0.757	0.503	−0.437	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	

Notes: Unconditional pairwise correlation coefficients between log night lights, log population density, and different measures of state fragmentation, with p-values in parentheses.

Table A.21: Geography and alternative measures of state fragmentation.

	Dependent variable:				
	Border frequency	Log number of states	Log mean state size	Border density	Log border distance
	(1)	(2)	(3)	(4)	(5)
Mountain >2000m	0.127***	0.054***	−0.119	0.002	−0.664***
	(0.040)	(0.018)	(0.102)	(0.002)	(0.156)
Log ruggedness	0.027***	0.017***	−0.121***	0.001	−0.090***
	(0.007)	(0.004)	(0.029)	(0.001)	(0.031)
River density	1.806***	1.095***	−2.594***	0.061	−7.924***
	(0.644)	(0.387)	(0.787)	(0.045)	(2.728)
Ag. suit. rainfed	0.004	−0.014	0.873***	−0.006***	−0.225
	(0.030)	(0.017)	(0.121)	(0.002)	(0.143)
Ag. suit. irrig.	−0.104***	−0.057***	−0.281***	−0.006***	0.414***
	(0.022)	(0.014)	(0.088)	(0.001)	(0.107)
Rainfall	0.040***	0.023***	−0.322***	0.002	−0.154**
	(0.013)	(0.007)	(0.047)	(0.001)	(0.061)
Log dist. to coast	−0.116**	−0.054*	2.737***	−0.013***	2.448***
	(0.052)	(0.031)	(0.283)	(0.004)	(0.278)
Coastline density	0.023**	0.012**	−0.043	0.001	−0.042
	(0.010)	(0.006)	(0.055)	(0.004)	(0.073)
Log land area	0.028***	0.015***	0.057**	−0.003***	−0.052*
	(0.006)	(0.003)	(0.027)	(0.001)	(0.030)
R^2	0.19	0.20	0.42	0.09	0.37
Number of obs.	5202	5202	5202	5202	5202

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.22: Modern outcomes and alternative measures of state fragmentation.

	Dependent variable:									
	Log night lights					Log population density				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Border frequency 1500-2000	0.371***					0.540***				
	(0.084)					(0.120)				
Log number of states 1500-2000		0.984***					1.353***			
		(0.142)					(0.205)			
Log mean state size 1500-2000			−0.309***					−0.286***		
			(0.027)					(0.041)		
Border density 1500-2000				3.250***					4.539***	
				(0.581)					(0.819)	
Log border distance 1500-2000					−0.144***					−0.249***
					(0.025)					(0.038)
R^2	0.35	0.35	0.40	0.35	0.35	0.39	0.39	0.40	0.39	0.40
Number of obs.	5202	5202	5202	5202	5202	5201	5201	5201	5201	5201

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. All specifications include latitude fixed effects and the benchmark set of geography controls. * indicates $p<0.10$, ** $p<0.05$, and *** $p<0.01$.

Table A.23: Modern outcomes and distance to borders: local deviations.

	Dependent variable:							
	Δ Log night lights				Δ Log population density			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Log border distance 1500-2000	0.090**	0.092**	0.207***	0.099**	0.169***	0.168***	0.335***	0.180***
	(0.042)	(0.042)	(0.034)	(0.041)	(0.058)	(0.058)	(0.049)	(0.058)
R^2	0.11	0.13	0.18	0.14	0.06	0.08	0.12	0.09
Number of obs.	5202	5202	3869	5202	5201	5201	3869	5201
Geography controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	None	Latitude	Latitude	Lat./Long.	None	Latitude	Latitude	Lat./Long.
Drop coastal cells	No	No	Yes	No	No	No	Yes	No

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.24: Geography and border frequency using Abramson data.

	Dependent variable: Border frequency 1500-1800.					
	(1)	(2)	(3)	(4)	(5)	(6)
Mountain >2000m	0.118***	0.337***	0.121	0.124***	0.324***	0.089
	(0.037)	(0.095)	(0.108)	(0.033)	(0.087)	(0.080)
Log ruggedness	0.032***	0.049***	0.048***	0.017***	0.025***	0.006
	(0.008)	(0.011)	(0.012)	(0.006)	(0.009)	(0.010)
River density	2.017***	2.094***	2.469***	1.028**	0.969**	1.342**
	(0.708)	(0.757)	(0.865)	(0.421)	(0.432)	(0.570)
Ag. suit. rainfed	0.007	−0.002	−0.000	0.060**	0.058	0.071*
	(0.035)	(0.046)	(0.048)	(0.028)	(0.035)	(0.038)
Ag. suit. irrig.	−0.126***	−0.135***	−0.109***	−0.032*	−0.029	−0.024
	(0.027)	(0.031)	(0.034)	(0.019)	(0.021)	(0.024)
Rainfall	0.048***	0.043**	0.045**	0.033***	0.033**	0.035**
	(0.015)	(0.018)	(0.019)	(0.011)	(0.013)	(0.014)
Log dist. to coast	−0.196***	−0.252***	−0.257***	−0.121***	−0.157***	−0.280***
	(0.062)	(0.078)	(0.085)	(0.045)	(0.057)	(0.066)
Coastline density	0.024*	0.053***	0.052***	0.017*	0.027*	0.015
	(0.012)	(0.020)	(0.017)	(0.010)	(0.015)	(0.014)
Log land area	0.029***	0.041***	0.045***	0.014***	0.015**	0.021**
	(0.007)	(0.010)	(0.011)	(0.005)	(0.008)	(0.008)
R^2	0.19	0.21	0.20	0.12	0.13	0.13
Number of obs.	5202	3861	3861	5202	3861	3861
Source for borders	Euratlas	Euratlas	Abramson	Euratlas	Euratlas	Abramson
Sample	Euratlas	Abramson	Abramson	Euratlas	Abramson	Abramson
HRE adjustment	No	No	No	Yes	Yes	Yes

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. Border frequency is measured over 1500-1800 for the Euratlas measure, and 1500-1790 for the Abramson measure. The HRE adjusted border frequency measures use Euratlas as source for the HRE borders, and either Euratlas or Abramson as source for sovereign state borders, as indicated (see also explanations in the text). All specifications include latitude fixed effects. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.25: Borders and modern outcomes using Abramson data.

Panel A	Dependent variable:					
	Log night lights			Log population density		
	(1)	(2)	(3)	(4)	(5)	(6)
Border frequency 1500-1800	0.443***	0.450***	0.493***	0.574***	0.531***	0.556***
	(0.073)	(0.078)	(0.071)	(0.103)	(0.112)	(0.102)
R^2	0.35	0.30	0.30	0.39	0.32	0.33
Number of obs.	5202	3861	3861	5201	3860	3860
Panel B						
Border freq. 1500-1800 (HRE-adj.)	0.045	0.019	0.253***	0.099	–0.021	0.244**
	(0.072)	(0.078)	(0.076)	(0.100)	(0.106)	(0.103)
R^2	0.34	0.28	0.28	0.38	0.31	0.31
Number of obs.	5202	3861	3861	5201	3860	3860
Border variable	Euratlas	Euratlas	Abramson	Euratlas	Euratlas	Abramson
Sample	Euratlas	Abramson	Abramson	Euratlas	Abramson	Abramson

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. Border frequency is measured over 1500-1800 for the Euratlas measure, and 1500-1790 for the Abramson measure. All specifications include latitude fixed effects. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.26: Geography and border frequency with smaller and larger cell size.

	Dependent variable: Border frequency 1500-2000					
	(1)	(2)	(3)	(4)	(5)	(6)
Mountain >2000m	0.073***		0.071***		0.074***	
	(0.029)		(0.004)		(0.082)	
Mountain >1000m		0.071***		0.052***		0.033
		(0.012)		(0.002)		(0.027)
Log ruggedness	0.146***	0.133***	0.055***	0.052***	0.197***	0.196***
	(0.004)	(0.004)	(0.000)	(0.000)	(0.010)	(0.010)
River density	0.109***	0.108***	0.090***	0.090***	0.164***	0.163***
	(0.631)	(0.632)	(0.021)	(0.021)	(1.002)	(1.010)
Ag. suit. rainfed	0.005	−0.000	−0.050***	−0.054***	0.067	0.060
	(0.018)	(0.018)	(0.002)	(0.002)	(0.045)	(0.045)
Ag. suit. irrig.	−0.127***	−0.119***	−0.041***	−0.039***	−0.165***	−0.162***
	(0.013)	(0.013)	(0.001)	(0.001)	(0.032)	(0.033)
Rainfall	0.143***	0.144***	0.132***	0.130***	0.102**	0.100**
	(0.007)	(0.007)	(0.001)	(0.001)	(0.016)	(0.016)
Log dist. to coast	−0.065***	−0.065***	−0.023***	−0.022***	−0.144***	−0.141***
	(0.028)	(0.027)	(0.003)	(0.003)	(0.063)	(0.064)
Coastline density	0.029**	0.025**	0.008**	0.008**	0.021	0.020
	(0.010)	(0.010)	(0.000)	(0.000)	(0.014)	(0.014)
Log land area	0.116***	0.109***	0.047***	0.046***	0.135***	0.132***
	(0.004)	(0.004)	(0.001)	(0.001)	(0.008)	(0.008)
R^2	0.19	0.19	0.10	0.09	0.24	0.24
Number of obs.	5202	5202	107945	107945	1527	1527
Cell size (degrees)	0.5×0.5	0.5×0.5	0.1×0.1	0.1×0.1	1×1	1×1

Notes: Ordinary least squares regressions with standardized coefficients and robust standard errors in parentheses (not adjusted for spatial correlation). The specifications include fixed effects for half-degree latitudes in columns (1)-(4), and one-degree latitudes in columns (5)-(6). * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.27: Geography and border frequency with smaller and larger cell size: local deviations.

	Dependent variable: Δ Border frequency 1500-2000					
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Mountain >2000m	0.034*		0.052***		0.062***	
	(0.028)		(0.005)		(0.039)	
Δ Mountain >1000m		0.079***		0.032***		0.080***
		(0.015)		(0.002)		(0.026)
Δ Log ruggedness	0.072***	0.068***	0.002	0.001	0.120***	0.111***
	(0.004)	(0.004)	(0.001)	(0.001)	(0.010)	(0.010)
Δ River density	0.062***	0.064***	0.070***	0.070***	0.035	0.033
	(0.239)	(0.246)	(0.015)	(0.015)	(0.730)	(0.728)
Δ Ag. suit. rainfed	−0.083***	−0.079***	−0.022***	−0.023***	−0.007	0.000
	(0.028)	(0.027)	(0.002)	(0.002)	(0.065)	(0.064)
Δ Ag. suit. irrig.	0.015	0.022	0.013***	0.013***	−0.020	−0.022
	(0.016)	(0.016)	(0.001)	(0.001)	(0.039)	(0.038)
Δ Rainfall	0.046**	0.046**	−0.003	−0.003	0.079**	0.080**
	(0.025)	(0.025)	(0.026)	(0.026)	(0.029)	(0.029)
Δ Log dist. to coast	−0.011	−0.018	−0.003	−0.003	−0.021	−0.027
	(0.262)	(0.260)	(0.216)	(0.217)	(0.293)	(0.291)
Δ Coastline density	0.030**	0.027**	0.003***	0.003***	0.042**	0.037*
	(0.008)	(0.007)	(0.000)	(0.000)	(0.010)	(0.010)
Δ Log land area	0.076***	0.069***	0.013***	0.013***	0.079**	0.074*
	(0.004)	(0.004)	(0.001)	(0.001)	(0.008)	(0.008)
R^2	0.03	0.04	0.01	0.01	0.05	0.06
Number of obs.	5202	5202	107929	107929	1527	1527
Cell size (degrees)	0.5×0.5	0.5×0.5	0.1×0.1	0.1×0.1	1×1	1×1

Notes: Ordinary least squares regressions with standardized coefficients and robust standard errors in parentheses (not adjusted for spatial correlation). The specifications include fixed effects for half-degree latitudes in columns (1)-(4), and one-degree latitudes in columns (5)-(6). * indicates $p<0.10$, ** $p<0.05$, and *** $p<0.01$.

Table A.28: Borders and modern outcomes with smaller and larger cell size.

Panel A	Dependent variable:											
	Log night lights						Log population density					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Border frequency 1500-2000	0.094***	0.092***	0.062***	0.062***	0.099***	0.081**	0.087***	0.088***	0.072***	0.074***	0.079***	0.069**
	(0.049)	(0.052)	(0.021)	(0.021)	(0.069)	(0.087)	(0.073)	(0.077)	(0.028)	(0.029)	(0.105)	(0.134)
R^2	0.35	0.39	0.23	0.25	0.40	0.49	0.39	0.45	0.30	0.31	0.45	0.54
Number of obs.	5202	3869	107596	99681	1527	949	5201	3869	107815	99868	1527	949
Panel B	Dependent variable:											
	Δ Log night lights						Δ Log population density					
Δ Border frequency 1500-2000	−0.053***	−0.089***	−0.032***	−0.037***	−0.037	−0.085***	−0.054***	−0.087***	−0.014***	−0.018***	−0.036	−0.099***
	(0.045)	(0.045)	(0.017)	(0.017)	(0.067)	(0.067)	(0.067)	(0.068)	(0.019)	(0.019)	(0.105)	(0.107)
R^2	0.13	0.18	0.04	0.04	0.12	0.19	0.08	0.11	0.03	0.01	0.10	0.15
Number of obs.	5202	3869	107579	99681	1527	949	5201	3869	107799	99868	1527	949
Cell size (degrees)	0.5×0.5	0.5×0.5	0.1×0.1	0.1×0.1	1×1	1×1	0.5×0.5	0.5×0.5	0.1×0.1	0.1×0.1	1×1	1×1
Drop coastal cells	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes

Notes: Ordinary least squares regressions with standardized coefficients and robust standard errors in parentheses (not adjusted for spatial correlation). In Panel A, columns (1)-(6) use log night lights as the dependent variable, and columns (7)-(12) use log population density as the dependent variable. In Panel B, columns (1)-(6) use local deviations in log night lights as the dependent variable, and columns (7)-(12) use local deviations in log population density as the dependent variable. All specifications include latitude fixed effects and the benchmark set of geography controls in Panel A, and their local deviations in Panel B. The latitude fixed effects refer to half-degree latitudes in columns (1)-(4) and (7)-(10), and one-degree latitudes in columns (5)-(6) and (11)-(12). * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.29: Borders and modern outcomes: night lights controlling for population density.

Panel A	Dependent variable: Log night lights					
	(1)	(2)	(3)	(4)	(5)	(6)
Border frequency 1500-2000	0.082	0.134***	0.087**	−0.058	0.064	−0.043
	(0.056)	(0.048)	(0.042)	(0.045)	(0.048)	(0.037)
Log population density	0.515***	0.502***	0.526***	0.528***	0.509***	0.499***
	(0.009)	(0.010)	(0.010)	(0.010)	(0.011)	(0.009)
R^2	0.67	0.73	0.77	0.77	0.77	0.81
Number of obs.	5201	5201	5201	5201	3869	5201
Panel B	**Dependent variable: Δ Log night lights**					
Δ Border frequency 1500-2000	−0.022	−0.041	−0.043	−0.039	−0.081***	−0.046*
	(0.028)	(0.028)	(0.028)	(0.027)	(0.028)	(0.028)
Δ Log population density	0.500***	0.487***	0.487***	0.487***	0.462***	0.487***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)
R^2	0.58	0.62	0.63	0.63	0.65	0.64
Number of obs.	5201	5201	5201	5201	3869	5201
Geography controls	No	Yes	Yes	Yes	Yes	Yes
Fixed effects	None	None	Latitude	Latitude	Latitude	Lat./Long.
HRE adjustment	No	No	No	Yes	No	No
Drop coastal cells	No	No	No	No	Yes	No

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. In Panel A geography controls refer to the benchmark controls used in the paper, and in Panel B their local deviations. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.30: Borders and modern outcomes: using log of 0.01 plus night lights.

Panel A	Dependent variable: Log (0.01+night lights)					
	(1)	(2)	(3)	(4)	(5)	(6)
Border frequency 1500-2000	1.296***	0.806***	0.563***	0.015	0.460***	0.068
	(0.150)	(0.147)	(0.132)	(0.129)	(0.141)	(0.117)
R^2	0.03	0.36	0.41	0.41	0.49	0.49
Number of obs.	5202	5202	5202	5202	3869	5202
Panel B	**Dependent variable: Δ Log (0.01+night lights)**					
Δ Border frequency 1500-2000	−0.217**	−0.283***	−0.277***	−0.263***	−0.454***	−0.296***
	(0.088)	(0.084)	(0.084)	(0.083)	(0.081)	(0.083)
R^2	0.00	0.11	0.13	0.13	0.15	0.15
Number of obs.	5202	5202	5202	5202	3869	5202
Geography controls	No	Yes	Yes	Yes	Yes	Yes
Fixed effects	None	None	Latitude	Latitude	Latitude	Lat./Long.
HRE adjustment	No	No	No	Yes	No	No
Drop coastal cells	No	No	No	No	Yes	No

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. In Panel A geography controls refer to the benchmark controls used in the paper, and in Panel B their local deviations. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table A.31: Borders and modern outcomes: controlling for urbanization.

Panel A	Dependent variable:					
	Log night lights			Log population density		
	(1)	(2)	(3)	(4)	(5)	(6)
Border frequency 1500-2000	0.371***	0.344***	0.306***	0.540***	0.491***	0.450***
	(0.084)	(0.067)	(0.078)	(0.120)	(0.102)	(0.114)
Fraction urban 1500-2000		3.103***			4.036***	
		(0.098)			(0.156)	
Fraction urban 1500			2.373***			2.570***
			(0.154)			(0.220)
R^2	0.35	0.56	0.42	0.39	0.52	0.40
Number of obs.	5202	5055	5036	5201	5055	5036

Notes: Ordinary least squares regressions with Conley standard errors in parentheses assuming spatial autocorrelation among observations within 1.45 degrees of each other. All specifications include the benchmark set of geography controls and latitude fixed effects. * indicates $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Figure A.1: How average output per location, Y^*, depends on the number of countries, N, for the same numerical example as in Figure 2 in the paper.

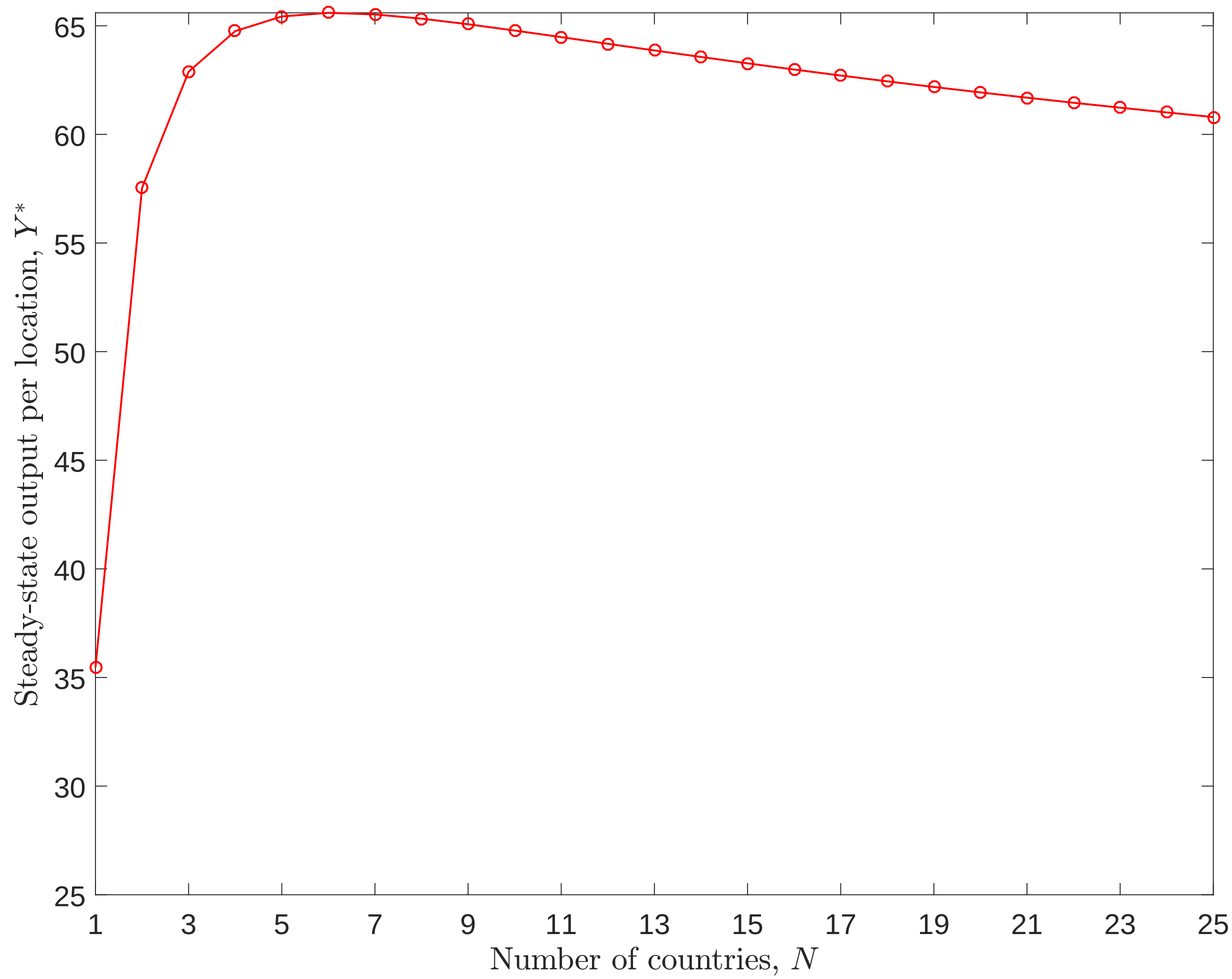

Figure A.2: Map showing locations of cells with high and low rainfall.

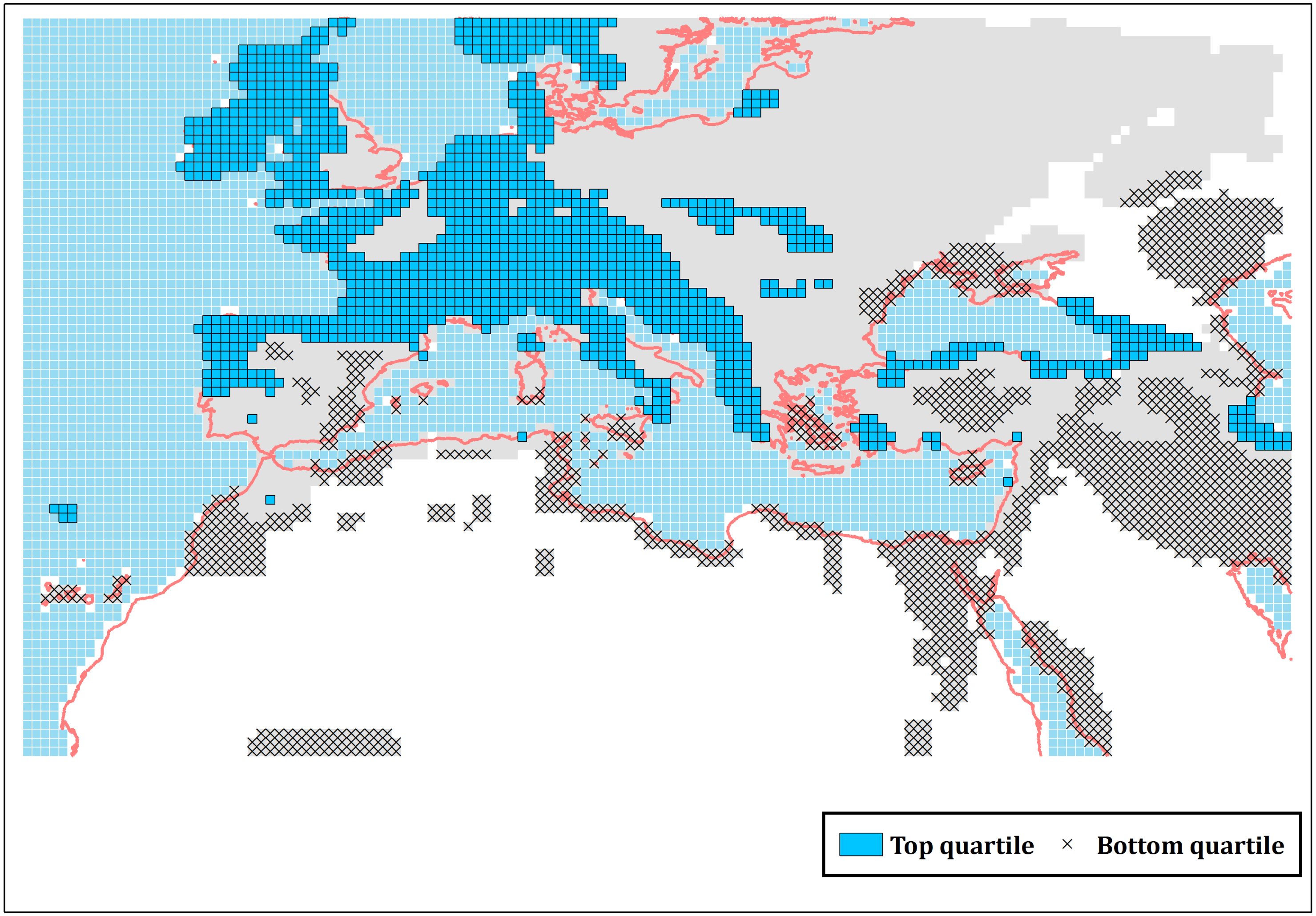

Figure A.3: Border frequency and modern outcomes after controlling for geography.