The Origins of the Territorial State*

Scott Abramson†
Princeton University
Department of Politics

Abstract

Building on both economic and martial theories of state formation, this paper develops and tests a model explaining temporal and spatial variation in the number and size of states. In doing so it provides the first systematic empirical tests of theories of state formation and draws several novel insights into the origins of the modern territorial state. In several steps I demonstrate a conditional relationship between war and the origins of the state. I begin by charting the evolution of the European system of states between 1100 and 1789. In doing so I show that contrary to bellicist theories, the typical state declined in size and the number of states increased during the period associated with the military revolution. However, once the data are disaggregated across space, I find that these trends are concentrated within a central band of European geography running roughly from the Low-Countries, through the Rhineland, and into Northern Italy. Exploiting random climatic variation in the propensity of certain pieces of geography to support large populations, I am able to show via an instrumental variables approach that this pattern was caused by variation in urban development. Where towns and cities formed, small independent political communities emerged. In the absence of urban growth large centralized states came to exist.

*Thanks to Carles Boix, Kris Ramsay, and Kosuke Imai. Please do not cite or circulate
†sabramso@princeton.edu
For the nearly forty years since the publication of *The Formation of National States in Western Europe* (Tilly 1975) questions of the territorial state’s origin have aroused the interest of scholars spanning comparative politics, international relations, and historical sociology. This paper addresses these questions by explaining why in some times and in some places large states came to dominate their smaller counterparts and, conversely, why in other times and places small states were capable of staving off the unifying efforts of would-be consolidators. In doing so, it advances scholarship on the topic in two ways. First, this paper develops a novel, micro-founded, formal narrative of European state formation. Second, in taking the empirical implications of this model to data, it represents the first set of statistical tests of theories of state formation making use of systematically collected data.

The theory I develop and test argues in favor of a conditional relationship between war-making and the creation and survival of states. Viewing states as wealth maximizing agents in anarchic competition for control of valuable territory, I show that more states will form when the costs of statehood, those associated with maintaining a military-bureaucratic apparatus, are low and in places where there is a substantial economic surplus to extract. In places and times when the opposite is true, when and where the costs of statehood are high and the economic surplus to be extracted is limited, fewer and larger states will form. Because the commercial revolution concentrated economic and urban development in a geographic corridor running through Central Europe, it is there that small states could match the increasing costs statehood and thus where political fragmentation ensued. Outside of this urban band, driven by the pressures identified by martial theories, large centralized states formed.

This simple approach captures the often competing insights of war-making and economic theories of state formation. The former, what I will call the bellicist approach, has its modern origins in the work of Charles Tilly though dates even earlier to German sociologists Otto Hintze and Max Weber. Scholars in this group view changes in war and war-making as having been determinative for the development the European system of states (Hintze 1906; 1975, Weber 1968, Tilly 1975; 1985; 1990, Downing 1992, Ertman 1997).
For this group “war made states” through an explicitly Darwinian process. In their view large states could far easier raise the manpower and finance required to field the increasingly large standing armies and increasingly dear technologies of coercion necessary to survive in an era of endemic warfare. Shifts in the scale and costs of war produced by technological shocks to the production of violence required ever increasing numbers of soldiers and ever more expensive armaments (Roberts 1956, Parker 1976; 1996, Black 1991, Rogers 1995). Though a number of possible military innovations are identified, the general logic of these bellicist theories is that technological and tactical changes in the production of force, by increasing the costs and frequency of war, selected those states most fit to survive (Bean 1973, McNeill 1984, Tilly 1975; 1985; 1990). Here, large states maintained an advantage in the form of substantial populations, larger tax bases, and greater access to natural resources. Because of these endowments the bellicists argue large states were more capable than their smaller counterparts of meeting the demands of war and, therefore, were more likely to survive.

The second camp, what I will call the economic school, reemerged with the work of Hendrik Spruyt in the early 1990s (Spruyt 1994a;b). This group, heavily influenced by scholars like Joseph Strayer, Perry Anderson, and Stein Rokkan, understand the development of the European system of states as the consequence of new social coalitions that formed from changing patterns of trade and economic development (Strayer 1973, Eisenstadt and Rokkan 1973, North and Thomas 1973, Anderson 1974). Where these new coalitions arose new and smaller states similarly formed.

Just as bellicist scholars rely upon a revolution in arms, those in the economic camp rely upon a series of historical shocks to patterns of trade and commerce in order to explain observed patterns of European state formation. The re-emergence of the Eastern trade at the end of the tenth century created new urban groups that altered the trajectory of state formation (Pirenne and Clegg 1937, Pirenne 1969, Lopez 1976, Cipolla 1994). Where towns formed, where burghers could bargain for or force their rights upon princes and kings, smaller political units came into existence.

Despite the extensive scholarship on the topic, little empirical work has been conducted to test these theories.¹ Having collected data describing the entire universe of European states over the

¹There have been several attempts at using agent based approaches to examine theories of state formation. For two examples, see Cederman (1997) and Boix et al. (2011). For recent empirical work examining the relationship
six hundred and ninety years preceding the French Revolution, I turn to the empirical evaluation of existing theories of state formation against my own theoretical contribution and, in the process of doing so, provide evidence in favor of my model.

I conduct this empirical analysis in two steps. I begin by charting the evolution of the European system of states between 1100 and 1789. I show that contrary to bellicist theories, the typical state declined in size and the number of states increased during the period associated with the military revolution. That is, during the period that many military historians associate with substantial and dramatic changes in the production of violence, the period in which large states are predicted by some to have come to dominate the international system, I find the exact opposite. However, once the data are disaggregated by geography, parcelling out Central Europe from the continent’s periphery, I find that these trends are concentrated within central band of European geography running roughly from the Low-Countries, through the Rhineland, and into Northern Italy. Outside of this central cordon, as predicted by bellicist theories, the number of states declined and their average size increased during the period associated with changes in the production of military force.

In the second step I show that the aforementioned pattern of political fragmentation in the continent’s central corridor and consolidation in it’s periphery was caused by the uneven reemergence of urban life. Where trade and commerce regained a foothold, where cities reemerged from the Dark-Age nadir, small independent political communities formed. Where cities and urban life did not emerge, large territorial states took shape. Taking an instrumental variables approach, exploiting random climatic variation in the ability of some places to support large populations, I show that this relationship was causal.

The remainder of this paper is organized as follows. First, I present a new dataset describing the boundaries, location, and existence of every European state between 1100 and 1790. From this, several facts emerge supporting the notion that the pressures related to increasing costs and frequency of war only forced major changes upon those states outside of the most productive parts of the continent. Next, I develop a simple formal model of state formation that generates predictions in line with the descriptive analysis presented in the prior section. Last, I test these predictions between state size and the origins of systems of public debt see Stasavage (2011a;b).
within the framework causal inference, taking an instrumental variables approach to identify the effect of changes in urban growth on the number of states that formed.

1 The Military Revolution and The European System of States

The idea, made famous by Charles Tilly, that “war made states” and reciprocally that “states made war (Tilly 1985)” has become so inculcated in the study of comparative politics and international relations that it is taken almost doctrinally. Indeed, a great deal of wide ranging and sometimes disparate historical evidence supports the claim that changes in the patterns of war caused the centralization, bureaucratization, and rationalization of many activities thought to typify modern statehood. With respect to the size of political organization, much of this literature either directly or indirectly asserts that the pressures of endemic war brought on by or, in turn, causing changes in the production of violence led to the creation of geographically large states.

Specifically, these changes are described as an early-modern “military revolution,” a notion first promulgated by Roberts (1956) who viewed changes in infantry tactics devised in the mid sixteenth century by Maurice of Orange and Gustavus Adolphus as fundamentally altering the manner in which armies were trained, raised, and paid for. Others view alternative technological and tactical changes as having marked the sea change between premodern and modern ways of waging war. Parker (1976; 1996), and McNeill (1984) for example, sees as crucial the late-fifteenth century development of siege artillery, a technological innovation that engendered an even more costly defensive response, the trace italliene.2

These accounts are not exhaustive. For example, Black (1991), places a greater significance on developments that took place between 1660 and 1710, namely, the invention of the flintlock musket. Rogers (1995), on the other hand dates an almost continuous series of military innovations beginning in the Hundred Years War. Regardless, whether or not the exact timing of the military revolution is clearly defined, a recurrent theme in the study of early-modern military history is that

---

2Bean (1973) similarly judges the development of powerful artillery as the crucial technological change demarcating the onset of the military revolution and is explicitly economic in linking the adoption of the cannon to the rise of large territorial states. In his logic, the innovation of these technologies drove up the fixed costs of war, making local political rule untenable, leading to geographically large states.
a fundamental change in the production of force took place sometime between the mid fifteenth and the end of the eighteenth centuries.

Social scientists have taken these historical accounts and built upon them theories of the modern state’s origins, arguing that there existed a causal relationship between the military innovations described above and the rise of geographically large, territorial states (Tilly 1975; 1985; 1990, Downing 1992, Ertman 1997). What is common from them is that changes in the costs of war selected those states that could pay them, leaving those less fit to fail. In the most famous of these theories, the states most fit for survival were those existing at the intersection of “capital” and “coercion” (Tilly 1990). These were states with both sizable populations necessary to provide the manpower for large standing armies and access to the finance necessary to fund them. The political units exhibiting these traits were large “territorial” states.

1.1 Data

Data measuring every state’s geographic scale is required to evaluate the claim that the military revolution caused a fundamental change in the size of political organization. As such, I have constructed a longitudinal data set measured in five year panels capturing all European political boundaries between 1100 to 1790. I code states as geographically defined political communities, distinct from kin or tribal groups, that maintain a quasi-monopoly on the use of force and use the following three criteria to operationalize this:

1. Direct Military Occupation

If a nominally independent state is militarily occupied by a foreign power, according to my coding scheme it ceases to exist. Similarly, if a political unit successfully conquers a piece of territory, this newly occupied territory is treated as a part of the conquering state. For example, when the Ezzelino or Pallavicini families were able to effectively wield military control over several Italian city-states I code the amalgamation of these units as a single state. Or when military orders like the Teutonic Knights or the Knights Hospitaller conquered well defined territories I treat these new units as independent states. Similarly, when the Castilian-Aragonese state drove the Moors from Grenada, the Emirate of Grenada ceases to be coded
Figure 1: The boundaries of every European state in the year 1450.

as an independent state and its territory gets coded as part of Castile.

2. The Capacity To Tax

Expropriative power, the ability to take from another that which she has produced or owns, is the coercive authority most associated with statehood. The formal exposition of states as wealth maximizing actors to be described in Section 2 as well as similar treatments of states as “stationary bandits” or organized criminal organizations underscores this crucial aspect of state violence: states use their ability to coerce in order to steal from those they
govern. One need not to ascribe Marxian theories of the state to adhere to an expropriative definition of statehood. For example, the ability to extract is the key feature of state power driving several recent and influential theories of political transitions (Boix 2003, Acemoglu and Robinson 2006). In these theories it is precisely the ability of the state to extract that actors - economic classes in these models - enter into conflict to control. As such, I take the capacity to tax as evidence of the state’s quasi-monopoly of coercion. So, for example, when Worms (1184) or Lubeck (1226) ceased to pay imperial taxes and (demonstrably) gained rights to collect taxes and tolls within their boundaries I code them as independent states.

3. Appointment of The Executive

It is possible for those who wield the coercive capacity of the state to be the agents of a foreign principal. That is, although the executive of a given polity may nominally wield the preponderance of force, her survival may depend upon the will of a foreign power. As such, the territory over which this dependent executive rules would not represent an independent political unit but rather the puppet or fief of some other power. Accordingly, I do not code units of this type as independent states. For example the heads of a large number of ecclesiastical units, for periods of time, were dependent upon papal authorities for their continued survival and as such do not get treated as independent states but get coded as part of the larger “Papal State.”

With this coding scheme I treat as distinct political units some states that more traditional historiography might consider unified. For example, when Boleslaw III of Poland divided his Kingdom between his sons creating the Masovian, Seniorate, and Sandomierz provinces along with Greater Poland, my coding scheme treats each as distinct units. Similarly feudal territories like Toulouse, Provence, or Brittany, though seigniorial dependents of the French King get coded independent until they are integrated into France proper. Likewise, Imperial city states, prince-bishoprics, free-cities, and imperial abbeys that effectively show independence - by demonstrating the characteristics outlined above - get coded as independent units. So, for example, when Fredrick Barabarossa attempted to assert Imperial rule in Lombardy, the various states that composed the Lombard League
and which successfully resisted begin to be coded as independent.

Similarly, during periods of civil conflict I identify the territories held by the various involved factions and, if they satisfy my coding scheme, I treat them as states. For example, during the English Wars of the Roses, I am able to define the territory controlled by the Yorkists and Lancastrians and treat them each as independent states during the periods of conflict beginning in 1455 and lasting until Henry Tudor’s victory at Bosworth in 1485 after which, despite sporadic rebellion by Yorkists in the North of the country, the rebellious factions do not satisfy my definition of statehood and do not get coded as independent states.

Following this coding scheme the data are constructed by manually geo-referencing several sets of historical maps. Two of the main sources from which the base GIS boundaries are constructed are the Centennia Historical Atlas (Reed 2008) and the Euratlas (Nussli 2010) Digital Atlas. The Nussli data is measured in one-hundred year panels whereas the Reed atlas utilizes a much more high frequency approach, taking observations in tenths of years. I use the boundaries as defined by both datasets aligning them at every hundred year mark based upon the coding scheme defined above. The Nussli data matches the Reed data nearly perfectly at these points. Where there are discrepancies it is usually because the Nussli dataset takes observations from a window surrounding each panel and not a snapshot exactly at the one-hundred year point. Because the Reed data is not geo-referenced I construct shape files that are compatible with GIS analysis by manually constructing the boundaries from re-projected images provided by the atlas and then referencing each observation using the European Alpers Equal-Area projection system.

The Nussli data have been used in several recent publications appearing in the American Political Science Review and are considered highly accurate (Stasavage 2011a;b, Blaydes and Chaney 2011). Nevertheless, even after combining the data from these digital sources there are still a number of imperfections; units I code as independent states are absent from the reconstructed shape files. These tend to be small independent principalities, ecclesiastical units, and city-states that were not picked up by the historical geographers who created the digital reproductions from which my maps are constructed. In order to rectify these flaws and prevent the ensuing selection problems that would plague the subsequent statistical analysis, I turn to a number of historical and
contemporary primary source maps to create high frequency boundary changes for these missing units.

From this combination of secondary and primary cartographic sources I am able to project the boundaries for all political units that meet the coding criteria but which are not found on the initial basemaps. Using known pieces of physical geography, known political boundaries, and the location of cities and towns to properly reference these maps, I create shape files that, with a high degree of accuracy, reflect the geographic scale of each unit. For each of these units I track the history of their boundary changes - expansions and contractions - and adjust the shape-files accordingly. In order to provide a more detailed description of the procedure I walk through the creation of the shape file for part of Nassau between 1159 and 1328.

The town of Nassau dates to at least 915 and was founded Robert the son of Dudo-Henry of the House of Laurenburg. The Laurenburgs built Nassau Castle in 1125 and established the County of Nassau in 1159 - effectively claiming rights of taxation, toll collection, and justice. As such, Nassau enters the dataset as an independent state following 1159. The County (later Principality) of Nassau exists in the digital base maps from this point, giving an accurate measure of its boundaries and size. However, the digital maps fail to record the dissolutions and mergers of various component units of Nassau, of which nearly all meet my coding as independent states. I manually make these corrections as described below.

For ninety-six years Nassau existed as a single independent state, however, upon the sale of Weilburg to the Count of Nassau in 1255 the territory was split amongst the two sons of Henry II with Otto I receiving the territory north of the river Lahn and his older brother Walram II receiving all of that to the south. Using this geographic boundary as the dividing line I create the Counties of Nassau-Dillenburg and Nassau-Weilburg, respectively. Dillenburg remained a single state until Otto’s death in 1303 after which Nassau-Dillenburg was divided into three units, splitting off Siengen and Hadamar from the initial unit. The boundaries of these new states are constructed using the boundaries as they exist on several historical reproductions and one primary source map (Blaeu et al. 1990, Franz 1952, Hockmann 2005). Using the known latitudes and longitudes of the cities of Siengen and Hadamar I then can reference the projected images from the historical
Figure 2: The Construction of Nassau between 1160 and 1303
reproductions using points representing the locations of these cities. From here, using these points and the boundaries of pre-existing Nassau-Dillenburg we can create the boundaries and subsequent shape-files for each of these new units. These three states remained independent until Siengen conquered Dillenburg in 1328 and then Hadamar in 1394.

1.2 Analysis

In this section I show that two trends predicted by bellicist theories do not appear in the data: 1.) during the period associated with the military revolution the typical state declined in size and, in the same period, 2.) the number of states was roughly constant. First, however, I define what I mean by the “typical” state.

To begin examine Figure 3 which plots the trend over time for every decile of state size measured in square kilometers. If one were to only consider the mean, it appears that the bellicist hypothesis matches the general trend; between 1400 and 1790 the average size of states more than doubled from approximately 33,000 to 71,000 square kilometers. Although a believer in the military revolution might take away from this some sense of confirmation, in the presence of extreme values the mean is a poor indicator of central tendency. This point is made apparent not only by the large spread between the median of state size and its mean but by the relationship between the mean and even the 90th percentile. We see that, following the late fifteenth century the size of the state at the ninetieth percentile was less than the mean size.

The reason for this is apparent. There are several extremely large states distorting what we might otherwise view as typical. To see this one need only examine the kernel density plot of state size as measured in square kilometers. The left hand panel of Figure 4 shows just this. The distribution of the non-transformed data is heavily skewed, with far more large states relative to what one would expect if size of states took a normal distribution. That is, the size of states are drawn from a “heavy tailed” distribution. Consequently, the mean is a poor measure of central tendency, yielding biased notions of what is typical.

An oft used and simple corrective to this type of problems is to assume a log-normal distribution,
that is, $\log(Area) \sim \mathcal{N}(\mu, \sigma^2)$. This effectively weights the outlying cases and, if the distributional assumption is correct, allows for better descriptive inference. That is, by making this transformation we can better describe the trend in state size for the “typical” state. It turns out in this instance the assumption of log normality is likely justified. The right panel of Figure 4 gives the histogram and kernel density estimates of logged state size. Visual inspection indicates that, indeed, log transforming the data yields a distribution that is much more closely approximated as normal.\footnote{Using a similar approach but examining only 1500 and 1998 Warren et al. also find that in both years state sizes are distributed log-normally.}

\footnote{One might consider another transformation of the data designed to preserve rank, properly weight outlying cases, and create an approximately normal distribution, the Box-Cox power-transformation (Box and Cox 1964). This transformation is as follows:}

$$y_i^{(\lambda)} = \begin{cases} 
\frac{y_i^{\lambda - 1}}{\lambda} & \text{if } \lambda \neq 0 \\
\log(y_i) & \text{if } \lambda = 0.
\end{cases}$$

(1)
Figure 4: Kernel density and histogram plots of state size across all time periods. The left-hand panel is that of the untransformed data, measured in square-kilometers. The right-hand panel presents those for the logged transformed data. The dashed vertical line in each plots the top quantile for each. Note the remarkable skew in the untransformed data driven by the few extremely large states within the top decile.

To gauge the appropriateness of this distributional assumption I plot the true sample quantiles of the log-transformed data against quantiles from a hypothetical normally distributed random variable with the same mean and variance as the observed data. If state size were normally distributed we would expect the sample quantiles to be on the line $y = x$. Figure 5 gives these Quantile-Quantile (QQ) plots of both the untransformed state size in the upper panel and the log-transformed QQ plot in the lower panel. As expected when we take either the logarithmically

Using the data pooled across all time periods I estimate $\lambda$ to be -0.099, very close to a value of zero. This provides some suggestive evidence that a log normal transformation is appropriate. None of the substantive empirical conclusions drawn from this chapter are altered if I utilize a Box-Cox instead of the logarithmic transformation
Figure 5: The top row plots for the untransformed data theoretical normal quantiles on the $x$ axis against the sample quantiles on the $y$. The bottom row plots the same but for the log transformed data. Note that the transformed data sits nearly perfectly on the line $x = y$ whereas the untransformed data does not.
transformed state size the QQ plots rest almost perfectly on the 45 degree line, a good indication that the data is, indeed, lognormal.

Although, a naive interpretation of the left hand panel of Figure 3 would indicate a revolution in state size coinciding with known changes in military technology, once we properly weight outlying cases this upward trend disappears. Having shown that the data in each period is likely drawn from a log-normal distribution, it is apparent that once transformed these data yield the opposite trend; the typical state between eleven-hundred and seventeen-ninety declined in size.

Moreover, it is obvious from simple visual inspection of the right-hand panel of Figure 3 that both the mean and median state size are declining over time and, as one would expect from a normally distributed random variable, they move in near perfect tandem. As such, we can use them as good descriptors of central tendency in the distribution of state size. The decline in both measures is substantial; the mean and median logged area decreasing between 1100 and 1790 from 9.03 to 6.32 and 9.62 to 5.67, respectively. Re-transforming these results gives declines of 7,818 and 14,816 square kilometres from initial values of 8,372 and 15,106. By these measures the “typical” state, though quite small in 1100, became even smaller over time, a pattern in direct contrast to the predictions made by those who see war as selecting against small units.

As with the average size of states, bellicist theories predict a decline in the number of independent units over time, a decline expected to be particularly sharp during the period associated with the military revolution. However, this hypothesis, once confronted with data, does not seem to hold. Again, I find that instead of declining over time, the number of independent units increased, expanding rapidly between the twelfth and thirteenth centuries and remaining relatively constant for the period after that. That is, during the period in which bellicist theories predict a decline we see no dramatic change in the number of states.

To summarize, although the bellicist literature describes a military revolution taking place at some point between 1450 and 1700, its predicted consequences fail to materialize when the data is examined systematically. During the period associated with systemic changes in the production of

---

5 A Engle-Granger two step procedure indicates that the two series are cointegrated. Estimating the following relationship Mean_t - β-Median_t = μ_t where β is estimated to be 1.04, a Dickey-Fuller test yields a test statistic of -3.69 allowing us with a high degree of confidence to reject the null hypothesis that μ_t is a non-stationary series.
violence two facts emerge: 1.) The typical state declined in size and 2.) The number of independent states increased.

1.3 Regional Variation

This section examines the trends in the number and size of states at the regional level. In doing so it yields some initial insight into the conditional relationship between war-making and state formation, providing evidence that only in the most productive places in Europe, a central regional band extending, roughly, in an arc from London, through the Low Countries and into Northern Italy, could small political communities persist. That is, as regional inequality became more pronounced those places that became increasingly wealthy saw a concomitant increase in the number of latent states willing to claim territory - despite the increase in costs territory. However, in less productive regions, political units facing increasing costs of statehood consolidated.

As initial evidence Figure 7 plots the average logged size and number of states, first dividing the map into two broad regions: Central Europe, the area described above, and the remainder of the map, what I will call the periphery. In both regions the number of states is increasing before the early thirteen century. After this period the upward trend in the number of states continues in
Figure 7: The average size and number of states separating out Central Europe from the rest of the continent. In both plots the black line represents Central Europe and the dashed line represents the periphery - everywhere that is not Central Europe. The top panel plots against time the number of states in these two regions. The lower panel plots for each region, once more against time, the difference between the mean of log-transformed data in a given year and the mean for the entire (regional) series.

Central Europe whereas in the periphery it plateaus and then begins to decline in the early fifteenth century. Similarly the average size is initially declining in both groups. However, in the periphery, beginning in the early sixteenth century, the average size starts to increase whereas it continues to decline in the center.

Nevertheless, by aggregating the entire periphery I am combing geographic locations with disparate historical experiences and thereby run the risk of masking some alternative trend. To guard against this I further break down the peripheral states into six regions. Since “region” is itself a constructed notion reflecting the political history of state formation, it is difficult to avoid a bit of arbitrariness with regard to where one region begins and another ends. Still, I construct “nat-
Figure 8: This map represents the classification of arbitrary grid-squares into seven regional groups: Iberia, Britain and France, Central Europe, Italy, Scandinavia and the Baltics, Eastern Europe (including Russia), and Ottoman Lands, including the Balkans.

“natural” regions fitting their political, social, and cultural histories. They are as follows: The British Isles and France, Iberia, Italy, Scandinavia and the Baltics, Eastern Europe including Russia, and the Ottoman Territories which includes the Balkans. These are shown in Figure 8. Each region is constructed using a line of latitude or longitude as an artificial boundary, those political units whose centroid falls within these demarcating lines gets counted as belonging to it. As a robustness check I vary the lines of latitude and longitude used to construct these regions, moving them in the cardinal directions by as much as 100 kilometers. None of the trends are substantively altered.

Again we see that outside of Central Europe the general trend was towards fewer and increasingly large states. However, within this central band the pattern is the opposite. Figure 9 plots the number and demeaned size of states by region across time. From this, we see that this number is declining in all regions outside of Central Europe in all years following the mid fourteenth century. Similarly the size is increasing in each peripheral location during this time period. Within Central Europe, however, the opposite is true as more and ever smaller states form between the beginning of the series and the end of the thirteenth century. Given that geographic space is fixed, the average
Figure 9: The top panel plots the number of states dividing the map into the Western Periphery (Iberia, Britain/France, and Italy) on the right hand side, and the Eastern Periphery (Scandinavia, Eastern Europe, and Ottoman Territories) on the right. The lower panel plots the demeaned trend in logged state size. From the early fifteenth century a clear decline in the number and increase in the size of states in all peripheral regions is apparent.

size of political organization in each region is a function of both the size of the region and the number of units. It seems as if the average size of states in peripheral regions increased because the number of units occupying them declined. Conversely, in Central Europe the average size of states declined because the number of states is increased.

With respect to the causes of these patterns of state formation the timing of the changes in trend is illuminating. The expansion of the overall number and the decline in the typical size of states coincides with the reemergence of trade and a general revival of commercial and urban life during the first third of the millennium. Like the regional pattern in the number and size of states these economic trends were not uniform across space but were geographically concentrated in a manner nearly identical to the patterns of state formation described herein. The emergence of
towns and cities - a product of the geographically concentrated revival of commerce and economic development - created groups in these places with the resources necessary to maintain or claim independence, thereby preventing the consolidation of political rule.

2 A Simple Model

In order to explain the patterns of state formation described above I develop a simple model that obtains results comporting with the facts as presented. Inspired by Tilly’s metaphor comparing states to organized criminal organizations - gangs in conflict over the control of neighborhoods where they establish themselves as racketeers - I model the choice of a wealth maximizing unitary individual dividing effort between extraction and conquest. That is, I capture the two basic strategies Tilly claims states have available to themselves: war making - fighting other states - and state making - expropriating revenue from the territories they control.

The model differs from existing formal treatments in two ways. First, it differs from the class of models that seek to explain the size and number of states (Friedman 1977, Bolton et al. 1996, Bolton and Roland 1997, Alesina and Spolaore 1997; 2005a). Instead of viewing these outcomes as the product of an optimization problem facing a social planner, decisive voter, or autocrat I more realistically treat these outcomes as the consequence of anarchic competition wherein combat between states determines territorial boundaries. Despite the fact that borders are the product of costly competition between states rather than a choice made by a planner, voter, or autocrat weighing welfare gains and losses of changes in size, by treating the optimal size of states as a choice made by some decisive actor, even sometimes in the shadow of interstate competition (Alesina and Spolaore 2005b; 2006), previous models fail to treat the number and size of states as outcomes directly determined by conflict.

Second, unlike formal models of state formation that allow for the development of states in anarchy (Skaperdas 1992, Konrad and Skaperdas 2005, Hirshleifer 2001), this model explicitly considers the spatial component of statehood. Although these models treat the number of states as the outcome of a violent processes, the competition they describe is not geographically bound. These models view the economic prize latent states attempt to claim as being amorphously defined
in space. States simply fight over some pie that is divided amongst a number of agents through some sort of grand mêlée. Their disregard of the territorial nature of state formation is similarly unrealistic, states fight other states for control of specific territorially fixed resources. The model I present combines the anarchic-competitive nature of interstate competition found in the models presented by this second group with the spatial component of the first group and, in doing so, moves towards a more complete model of state formation.

I show that variation in the size of states is driven by the ability of latent states, political groups who might plausibly seek statehood, to claim territory, constrained only by the costs of operating as an independent political entity. In geographic locations more suitable for producing output these potential leviathans are going to be more willing pay the costs of statehood and claim territory than they would in less productive regions. Consequently, a greater number of, and therefore on average smaller, states will form in productive regions than they will in less productive regions.

2.1 Setup

I model the decisions of states that maximize revenue with respect to effort devoted to extraction and effort devoted to the acquisition of territory from which economic output is extracted. The total amount of revenue extracted by a state is treated as a function of effort devoted to extracting economic surplus and the amount of economic surplus, \( R_i = R((1 - E_i), Y_i(E_i)) \). For simplicity we can use the following production function:

\[
R_i = (1 - E_i)\gamma Y(E_i)
\]

By assumption we have:

\[
1 = E_i + (1 - E_i)
\]

Where \( E_i \) is the effort state \( i \) devotes to combat with other states and \((1 - E_i)\) the effort it devotes to extraction. The parameter \( \gamma \) captures technology by which effort devoted to extraction is translated into revenue, measuring the elasticity of extractive effort.\(^6\) That is, it tells us for a

\(^6\)To see this simply take logs of the production function \( \log(R_i) = \gamma_i \log((1 - E_i)) + \log(Y(E_i)) \)
1% change in effort devoted to extraction what percentage change in total revenue is yielded. For high values of $\gamma$ states are efficient at extraction and for low values they are inefficient.

Economic output, $Y$, is distributed over land, $L$, on $[0, 1]$ according to a distribution function $F(\cdot)$ and corresponding density, $f(\cdot)$, where $F(\cdot)$ is strictly increasing on $[0, 1]$ and satisfies $F(0) = 0$ and $F(1) = 1$. I assume that $f(\cdot)$ is strictly quasi-concave. Letting $\overline{M}$ denote the mode, this implies that $F(\cdot)$ is strictly convex on $[0, \overline{M}]$ and strictly concave on $[\overline{M}, 1]$.

We can order states based upon their location in space, $l_1, \ldots, l_n$, from lowest to highest with lower subscripts representing locations closer to 0 and higher subscripts representing locations closer to 1. The border between states $i$ and $i + 1$ is defined by the contest success function $p_i = \frac{(\delta_i E_i)^m}{(\delta_i E_i)^m + ((1 - \delta_{i+1}) E_{i+1})^m}$ such that $(l_{i+1} - l_i)p_i$ is equal to the amount of territory between $i$ and $i + 1$ held by state $i$. Where $m$ is the decisiveness parameter, capturing the returns to scale of effort devoted to combat. Since $m$ cannot be meaningfully zero or negative we assume $m > 0$. If $m \leq 1$ diminishing returns from competitive effort result. For $m > 1$ then increasing returns to combat hold throughout. The decisiveness parameter has an immediate interpretation with respect to military conflict. A low $m$ corresponds to situations in which the defense has the upper hand whereas a high level of $m$ implies that the offense is advantaged. I define $\delta_i \in [0, 1]$ as the fraction of effort devoted to conquest aimed at defining the right border and $1 - \delta_i$ is the fraction of effort devoted to conquest aimed at defining the left border.

$Y_i$ is the amount of economic output controlled by state $i$ located in the interval defined by the border demarcating the territory held by state $i$ and $i + 1$ and the border demarcating the territory held by state $i$ and $i - 1$. It is defined as:

$$Y_i = F(l_i + (l_{i+1} - l_i)p(E_i, \delta_i, E_{i+1}, \delta_{i+1})) - F(l_{i-1} + (l_i - l_{i-1})p(E_{i-1}, \delta_{i-1}, E_i, \delta_i))$$

and each entrant state faces the following maximization problem:

$$\max_{\delta_i, E_i} R_i = (1 - E_i)\gamma \cdot Y_i(E_i, \delta_i) - C$$

Where $Y_i$ is defined as above and $C$ is the fixed cost of statehood. Defining $a_i = \delta_i E_i$, the amount
of effort devoted defining the right border, and $b_i = (1 - \delta_i)E_i$, the amount devoted to defining the left border, it also allows us to rewrite the maximization problem as

$$\max_{a_i, b_i} R_i = (1 - (a_i + b_i))Y(a_i, b_i) - C$$

Each of these “latent leviathans” must choose whether or not to “become” a state by locating somewhere on the interval $[0, 1]$. So, some subset of the total population, $n \in N$, will become a state and the remainder will remain latent. Unlike political parties or candidates for election two entrant states, restricted by the laws of physics, cannot realistically occupy the same physical space. As such, I restrict location strategies to exclude multiple entrants at the same location. Conditional on the choice to locate, states then face the decision regarding distribute their effort between extraction and conquest, described above.

The game proceeds in three stages:

1. Latent states simultaneously decide whether to enter and a location, $l_i$ on $[0, 1]$.

2. States that have entered make investment decisions about effort devoted to the conquest and production

3. Rents to statehood are determined

The equilibrium concept is subgame perfection and the game is solved by backwards induction. In the last period states will take their location as fixed and make investment decisions over conquest and production.

### 2.2 Size in the Long Run

During the period studied, at least in the long run, the technologies of extraction and war-making diffused across all states. Given that states had roughly similar access to the same technology we might expect actors in the international arena to expend similar, if not identical, amount of effort or resources on interstate competition. That is, in the long run we might expect states to balance against each-other. As such, I will focus the set of equilibria in which strategies are symmetric in both effort and direction.
To see this symmetric equilibrium note the following.

For entrants 2 through n-1, we can solve for the following reaction function:

\[
\frac{a_i}{b_{i+1}} \frac{1 - E_{i+1}}{Y_{i+1}} = \frac{b_i}{a_{i-1}} \frac{1 - E_{i-1}}{Y_{i-1}}
\]  

(2)

The proof of this is left to the appendix.

Since entrants 1 and n only face a competitor in one direction they simply maximize with respect to \(E_1\) and \(E_n\) directly. From the first order conditions for these two entrants we can solve for the following

\[
\frac{E_1}{1 - E_1} \cdot Y_1 = \frac{b_2}{1 - E_2} \cdot Y_2 \quad \& \quad \frac{E_n}{1 - E_n} \cdot Y_n = \frac{a_{n-1}}{1 - E_{n-1}} \cdot Y_{n-1}
\]  

(3)

Taking these each of these and plugging them into the reaction functions for players N-1 and 2 we obtain the following system of equations

\[
\begin{align*}
    e_1 \cdot Y_1 & = (1 - \delta_2) e_2 \cdot Y_2 \\
    \delta_2 e_2 \cdot Y_2 & = (1 - \delta_3) e_3 \cdot Y_3 \\
    \ldots & = \ldots \\
    \delta_{n-2} e_{n-2} \cdot Y_{n-2} & = (1 - \delta_{n-1}) e_{n-1} \cdot Y_{n-1} \\
    \delta_{n-1} e_{n-1} \cdot Y_{n-1} & = e_n \cdot Y_n
\end{align*}
\]

**Proposition 2.1.** For the subset of symmetric equilibria in which states devote equal effort to conquest, that is \(E_i = E_j \forall i, j\) and in which states devote equal effort in both directions \(\delta_i = (1 - \delta_i) = \frac{1}{2} \forall i \in \{2...n - 1\}\) we can conclude that states will be progressively smaller as we approach the mode of the distribution of economic output.
Proof. Rearranging 2.2 we get the following system.

\[
\frac{1}{2} = \frac{y_1}{y_2} \\
1 = \frac{y_2}{y_3} \\
. & = . \\
. & = . \\
. & = . \\
1 = \frac{y_{n-2}}{y_{n-1}} \\
\frac{1}{2} = \frac{y_{n-1}}{y_n}
\]

When states are devoting equal effort and splitting their effort evenly in both directions, it follows that \(Y_i\) must equal \(Y_{i+1}\) for all \(i \in \{2...n-1\}\). In the set of symmetric equilibria the only strategy available for states to satisfy this is their location decision. In order for the above system to hold states must locate progressively closer to each other as they approach the mode. To see this note that the amount of economic output is increasing as you approach the mode of the distribution. Since in the symmetric equilibrium entrants are devoting the same effort to conquest in both directions, they are equally dividing all territory located between any given pair of states. Moreover, since the amount of economic surplus available to expropriate is increasing as you approach the mode, in order for \(Y_i\) to equal \(Y_{i+1}\) the territory that states extract from must be declining and therefore entrants must locate increasingly close to each other as they approach this peak. More succinctly, when all states are expending equal effort in both directions they are going to locate progressively closer together in the most productive places. They must split the same amount of output spread over an increasingly diminishing amount of territory. As a consequence, states in these productive places are going to be smaller.\(^7\)

\[\square\]

The implication of this result is straightforward. In the most productive places more states will form, doing so because the economic surplus available for them is greatest in these locations.

\(^7\)Entrants 1 and \(n\) may or may not be larger than entrants 2 and \(n-1\), respectively. Because they only face one competitor the amount of resources they hold is equal to \(Y_1 = 2Y_2\) & \(Y_n = 2Y_{n-1}\) and their size is going to be determined by the behaviour of \(f()\) at the tails.
Because states will form progressively close together as the economic surplus available to capture increases they will, in these more productive places, be smaller. Furthermore, from this result we can learn about the manner in which changes to the distribution of economic output affect the of state sizes across space.

**Corollary 2.2.** In the set of symmetric subgame perfect Nash equilibria, as the variance of \( f(\cdot) \) declines, states will locate increasingly close together as they approach the mode. That is, economic output becomes increasingly concentrated around the peak, the number of states will decline far away from the mode of the distribution and increase closer to the mode.

To see why Corollary 2.2 holds note that if all states are committing equal effort to combat then \( Y_i = Y_{i+1} \). As the variance of \( f(\cdot) \) decreases then the fraction of resources located proximate to the mode is increasing. As such, if \( Y_i = Y_{i+1} \), then states must locate even closer to each-other as they approach the mode when the variance is low. In the extreme case, where the distribution is approximately uniform, states will locate close to equidistant from each-other.

This implies that states will be increasingly small as territory becomes more productive. As an example, when new trade routes open they increase incomes in some places relative to others. In these newly wealthy places a greater number of latent states are expected to locate. Conversely, in less productive regions fewer latent states will choose to locate. This divergence in the number and size of states is exacerbated by the degree of regional inequality. As the distribution of economic output becomes more unequal across space, states will locate even more frequently and closer to each other as the most productive places grow increasingly wealthy.

### 2.3 The Number of States

Knowing how players will behave once they have located in space we now must consider the entry decision. States will enter so long as they receive a positive payoff from doing so. That is, they will be indifferent between remaining latent and paying the entry cost at the point where cost equals Rents, when \( C = (1 - E_i^*)^\gamma Y(E_i^*) \). Where \( E_i^* \) and \( Y_i^* \) are the equilibrium effort and output for
state $i$. We get from this two main conclusions about the number of states.

**Proposition 2.3.** *As the fixed costs of statehood rise, the number of latent states willing to enter declines.*

To see this note that the equilibrium payoff to state $i$ is $R^*_i = (1 - E^*_i)Y(E^*_i) - C$ and that latent states will only enter if there is a positive return from doing. So, for some $C$, $(1 - E^*_i)Y(E^*_i) - C = 0$ and all latent states are indifferent between entering and staying out. For $C > C$ there exists at least one state who would prefer to remain latent.

**Proposition 2.4.** *In the subset of symmetric equilibria, in those where greater effort is devoted to conquest fewer states will be willing to enter*

To show this consider any two equilibria, $E^*$ and $E^{**}$ such that $E^* > E^{**}$. Note that the equilibrium payoff $(1 - E^*_i)Y(E^*_i) - C$ is declining in $E$. It follows that when each state is devoting a large fraction of effort to combat the total number of latent states capable of receiving a positive return to statehood is lower than when states are devoting low levels of effort to combat. In other words, in equilibria with “high” effort devoted to combat fewer states will be willing to enter because they will be less capable of obtaining a positive payoff.

Propositions 2.3 and 2.4 capture the effects of the military revolution on the size and number of states. First I show that as the fixed costs of statehood increase the number of states is expected to decline. This is analogous to the standard bellicist hypothesis in that the costs related to being a state, for example the maintenance of large standing armies and bureaucracies, drive states that cannot remain profitable from the business of being states. Similarly, We get the effect of changes in the marginal costs of war. That is, low $E^*$ captures periods of relative peace where states are expending little effort on conquest and high $E^*$ represents periods of endemic war where states are expending substantial portions of national resources on war. When war is more frequent and more costly fewer latent states will be willing to enter and states will be on average larger.
2.4 Empirical Implications

The long run model yields two countervailing mechanisms affecting the size of states. The first is that which bellicist theories of state formation have proffered: as the costs of statehood increase, costs they typical associate with maintaining a bureaucratic-military apparatus, fewer states will survive. By driving up costs of statehood the military revolution of the sixteenth and seventeenth centuries selected against states that could not pay these costs. Similarly, as war becomes more frequent and requires greater marginal investment, states that cannot match these costs will be driven from the market.

The second mechanism matches the conclusions of the economic school of state formation; the creation of new states is a function of their ability to obtain a positive payoff from forming. That is, state formation is a function of not only the costs of statehood but variation in the economic potential of geography needed meet these costs. In locations where there is a great deal of economic surplus latent groups will be more willing to form because it is easier for them to turn a profit. The effect of the re-emergence of trade, urban life, and the divergence of European incomes associated with the commercial revolution are likewise captured by the model (Pirenne 1969, Anderson 1974, Lopez 1976, Spruyt 1994b). The unequal re-emergence of trade and urban life, outcomes concentrated in Central Europe, provided economic incentives that led to a concomitant increase in the number of independent states by providing groups in some places the resources needed to effectively monopolize coercion. However, the pace of economic development only outpaced the increasing costs of statehood in those regions fortunate enough to grow. In those places that failed to develop we see a decline in the number of states. In peripheral Europe the decline in the number of states is explained by the rising costs of statehood. In these locations economic development did not outpace the pressures of war and the ever large, complex, and costly state apparatuses needed to match them.
3 Urban Growth and State Formation

This section directly estimates the causal relationship between development and state formation. I do this using arbitrarily defined pieces of geography, 10,000 km² grid squares, as my unit of analysis. As the dependent variable I count the number of states that occupy any space on each unit. This operationalization is directly related to the theoretical model presented above. Although the theory predicts entry of new states on a single dimension, geographic space is two dimensional. As such, I test the hypothesis derived from the model, that in the most productive places more states will form, by examining the relationship between levels of urban population and proto-industry on the number of “entrant” states on each arbitrarily defined grid-square. I estimate the following model:

\[ \text{Entrants}_{it} = \alpha + \beta \cdot \log(\text{Urban Population}_{it}) + \epsilon_{it} \] (4)

Where the parameter of interest \( \beta \) captures the effect of urban population on the number of states on a given grid-square. I estimate this relationship via pooled OLS and, since the number of entrants is by definition a positive integer, negative binomial regression. The results are summarized in Table 1. I estimate three models, the marginal effects from each of which are extremely close in size. For each model I successively add in year effect and, latitude and longitude. The estimated effects of a one-hundred percent change in urban population range from 9% to 20%, a substantial effect when the differences between the maximum urban population in any given year range from forty-seven to sixty-six times the mean.

These results provide some initial evidence that the most urban quadrants in any year had the most states “enter” on them. As a consequence, the average size of these states in these places was by definition smaller. Nevertheless, these results need not signify a causal relationship. That is, the assumptions necessary for the above regression analysis to estimate causal effects are unlikely to be satisfied. The size of the urban population and the number of states within a given piece of territory are likely driven by a number of common but unobservable confounders.

\[ \text{Entrants}_{it} = \alpha + \beta \cdot \log(\text{Urban Population}_{it}) + \epsilon_{it} \] (4)

---

\(^{8}\) I have estimated all of the subsequent analysis using an alternative dependent variable, a Herfindahl like index of fragmentation, defined for each grid-square \( j \) as \( H_j = \sum_i N \left( \frac{\text{Area}_i}{\text{Area}_j} \right)^2 \). Where \( \text{Area}_i \) is the total area held by state \( i \) in grid-square \( j \) and \( \text{Area}_j \) is the total area in grid-square \( j \). This measure is highly correlated with the number of entrants (\( \rho = -.72 \)) and all of the results remain qualitatively the same if this measure is used.
<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>Neg Bin</td>
<td>OLS</td>
</tr>
<tr>
<td>log(Urban Population(_{it}))</td>
<td>.16</td>
<td>.16</td>
<td>.17</td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year Effects</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Lat/Long</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R(^2)</td>
<td>.10</td>
<td>.11</td>
<td>.11</td>
</tr>
<tr>
<td>(\phi)</td>
<td>6.55</td>
<td>6.94</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>9357</td>
<td>9357</td>
<td>9357</td>
</tr>
</tbody>
</table>

Table 1: The marginal effect of a one-hundred percent change in urban population on the number of entrant states on a given grid-square. Ninety-five percent confidence intervals in brackets estimated from robust standard errors clustered by grid-square. The confidence interval for the negative binomial regressions are estimated via quasi-Bayesian simulation.

In order to show that the relationship between changed in patterns of urban development and state formation is causal, I take an instrumental variables approach, exploiting random climatic variation in the ability of arbitrary pieces of geography to support large urban populations. Using a set of paleoclimatological sources I construct an estimate of the propensity of land to feed large populations by growing cereals like wheat. I will show that the ease with which some places could produce calorically dense and easily storable foods like wheat is a strong cause of urban population and, I argue has no direct effect on state formation processes. As such, this measure is an ideal instrument, satisfying the necessary exclusion restriction needed to estimate causal effects.

### 3.1 Urban Growth and Agricultural Productivity

Cities as centers of economic specialization can only exist once populations are able to devote effort to activities other than subsistence. As such, the places that had some advantage in feeding large populations, places that could produce food most easily, were also those more likely to develop as urban centers. In pre-modern Europe these locations were those that could grow a specific set of crops that were superior to other alternatives in terms of their ability to support feed sizable populations over an extended period.

The sentiment that premodern urban life required an agricultural surplus to sustain itself
is echoed throughout the literature on the commercial revolution and pre-modern city-growth. De Vries (1984) and Bairoch and Braider (1991) provide but two prominent examples of the view that increases agricultural productivity are a necessary preconditions for growth of cities. Nicholas (1997) is rather succinct in describing this logic.

_Cities could not develop until the rural economic could feed a large number of people who, instead of growing their own food, compensated the farmer be reconsigning his products and later by manufacturing items that the more prosperous peasants desired. The ‘takeoff’ of the European economy in the central Middle Ages is closely linked to changes in the rural economy that created an agricultural surplus that could feed large cities [p. 104]_

To identify in an instrumental variables framework the causal effects of urban development on state formation I exploit random climactic variation in the ability of a given location to produce key agricultural outputs necessary to support large populations. The natural predisposition for some places to feed large groups has been directly related to the development of urban life and the revival of commerce by a number of economic historians. Pirenne (1969), for example, argues that the location of towns in the premodern Europe was strictly a function of natural geography, that “In a more advanced era, when better methods would permit man to conquer nature and to force his presence upon her despite handicaps of climate or soil, it would doubtless have been possible to build towns anywhere the spirit of enterprise and the quest of gain might suggest a site.” This was, however, not the case. Rather, “...the first commercial groups were formed in neighborhoods which nature had disposed to become...the focal points of economic circulation.”

I focus on the ability of some places to produce cereals like wheat and barely for two reasons. First, the European diet of the premodern era was centered around the consumption of complex carbohydrates derived from cereals. So important to the typical European’s life were foods derived from these plants that Economic historian Robert Lopez notes that “in the form of bread, porridge, or mush, cereals were almost everywhere the basis of human alimentation...(Lopez 1976).” Not only were cereals central to diets across European geography but across classes as well, being integral
in the consumption patterns of the aristocracy and peasantry alike although certainly in unequal proportions (Duby et al. 1974).

Second, the ability to grow cereals has been directly linked to the support of large populations. Cereals like wheat, unlike other plants, are most capable of feeding large populations with minimal effort. This is because cereal crops, unlike fruits, pulses, or nuts, are extremely fast growing, high in calories from carbohydrates, and have extremely high yields per hectare (Diamond 1997). Moreover, unlike other crops cereals can be stored for long periods of time enabling communities to smooth consumption over extended periods. To summarize, the ability to feed large populations was key to the development of cities. Since in pre-modern Europe the principle component of diets were cereals like wheat, foods that are particularly good at supporting large populations, climatic variation across time and space in the ability to grow these crops serves as a good encouragement for urban growth.

The instrument is constructed in two steps:

1. I take spatially referenced temperature data from two paleo-climatological sources, both measured at half-degree by half-degree latitude/longitude intervals. The first measure from Mann et al. (2009) records temperature anomalies for the past 1500 years. A temperature anomaly captures the deviation at each point from the 1961-2000 mean temperature. I then construct a measure of absolute temperature by adding back the 1961-2000 baseline mean temperature as calculated from Jones et al. (1999)’s twentieth century data. This yields a half degree by half degree grid of temperatures for every year over the past 1500 years. Hundred year averages of these yearly measures are then taken.

2. Next, using tension weighted splines I take these estimates, measured at fixed intervals, and construct a smoothed measure of temperature for the entire continent. From this continuous measure the average for each grid-square is taken yielding an estimate of temperature across our fixed but arbitrary pieces of geography. All of these operations are taken using the interpolation and zonal averaging tools found in ArcGIS 10.

To show that the logic of the encouragement holds, I use twentieth century data on temperature
Figure 10: The FAO wheat suitability index is plotted on the y-axis against average annual temperature on the x-axis. The FAO measure is the “Agro-climatically attainable yield for rain fed wheat” is from the Global Agro-ecological Assessment for Agriculture in the 21st century. It captures the ability of land to produce wheat absent of modern irrigation techniques. A clear parabolic relationship with a peak at approximately 10.5 degrees Celsius is observed.

and two measures of the ability to grow wheat to demonstrate the robust relationship between the two. I employ two data sets from the FAO. The first, the “Agro-climatically attainable yield for rain fed wheat,” is from the Global Agro-ecological Assessment for Agriculture in the 21st century. It captures the ability of land to produce wheat absent of modern irrigation techniques. I estimate the optimal climate to grow wheat (at around 10.5 degrees Celsius). A clear a parabolic relationship between temperature and this FAO measure is observed simply by plotting it against average annual temperature between 1960 and 2000. The peak of this curve is approximately at 10.5 degrees Celsius.

Regressing the FAO measure of wheat suitability on the absolute deviation from 10.5 degrees we see that, indeed there is an extremely strong relationship between the two. The results from this regression are summarized in first column of Table 2. The effect of a one degree deviation from the optimal temperature is substantial, decreasing the FAO measure by .61 units. This is a particularly large effect since the FAO measure is on a fourteen point scale. Moreover, an extremely large
Table 2: The relationship between temperature and the suitability to produce wheat. The first column regresses the FAO wheat suitability index against the absolute average distance from 10.5 degrees Celsius between 1960 and 2000. The second column regresses the average wheat yield on this measure.

amount of the variation in the FAO wheat suitability measure is explained by deviation from this optimal growing temperature, the estimated $R^2$ statistic is calculated to be .55. Additionally, regressing average annual wheat yields between 1960 and 1990, measured in hectograms produced per hectare, on deviation from the optimal growing temperature again shows a similarly robust relationship. A one degree deviation from the optimal temperature has a large effect on average annual wheat yields – approximately 1600 hectograms per hectare.

### 3.2 A Valid Instrument

To estimate causal effects in an instrumental variables framework a series of assumptions must be satisfied (Imbens and Angrist 1994, Angrist et al. 1996). They are as follows:

1. Random Assignment of the Instrument
2. Non Zero Effect of the Instrument on the Treatment
3. Monotonicity of the Instrument
4. Exclusion

The first assumption, random assignment, is almost certainly satisfied. Until the 19th century there has been no human effect on climate, the changes to which have been determined by naturally occurring phenomena ignorablely determined with respect to the treatment considered here, urban population. The second assumption that the instrument is strong is also satisfied. In each of the
first-stage regressions all tests against weak instrumentation satisfy conventional levels of statistical significance.

In order for the third assumption, monotonicity, to be violated there would need to be a set of “defier” quadrants. For a unit to “defy” the encouragement, the size of the urban population would have to decline in response to more favorable climatic conditions or conversely increase in response to less favorable conditions. It is reasonable to think of reasons for which there would be a class of “always takers” - imperial cities who, by virtue of their position at the center of an extractive empire, would always maintain large urban populations. Similarly, we can think of possible “never takers” quadrants that because they are distant from trade routes or for reasons other than climate cannot feed large populations and will consequently never develop large urban centers. However, for a unit to defy the encouragement would require an extremely unlikely response in the exact opposite direction. Since the instrument is perturbing an, essentially, biological response this is unlikely.

The last assumption, the exclusion restriction, is also likely satisfied. For this to be violated the instrument, deviation from the optimal growing temperature for cereals like wheat, would have to effect the number of states entering on given quadrant in some way other than through my measure of economic development, urban population. It seems unlikely that changes in this measure of climate will effect the choice to form a new state or for states to disappear other than through their effect on economic incentives as proxied by urban development.

As discussed in the theoretical model, we can divide the constraints facing statemakers into two component parts. The first are the costs of statehood. Substantively we take these to mean the size of militaries and bureaucracies necessary to maintain statehood. Shocks to the propensity for a given piece of geography to support urban populations should have no direct effect on this particular incentive to form or dissolve as an independent state. The instrument perturbs the major component of the second constraint, the economic surplus available for latent or existing states to claim, urban population. The question then becomes do changes in this measure of temperature affect economic incentives other than through urban population?
To answer this we must distinguish between weather shocks that create short term variation in agricultural production and long term climatic trends that affect the ability to support large populations. Short term weather shocks of the sort described by Miguel et al. (2004) do affect agricultural economic output but by their very nature (at least until the industrial revolution) are random both across time and space. Since we are taking an average over a hundred year period and because these shocks are distributed in a random way these short term perturbations should both average out over time and space. Moreover, since the changes in optimal growing temperature over a hundred year panel are extremely slow, the long term trend in which would be very difficult to perceive at any given point in time in an era before meteorological data was systematically collected. Because of this these changes would only effect the economic incentive to form states through their long term effects - the ability to sustain large populations.

Nevertheless, in case climate may have some direct effect on urban population I attempt to control for alternative channels through which the optimal growing temperature might effect state formation. I “control” for the ways in which climate, other than though the deviation from this optimal temperature for cereals, might effect the number of states on a given square. I do this in two ways. The first is by controlling for both latitude and longitude. Since climate is strongly correlated with geographic location, controlling for the position in space should similarly control for the effects of climate other than through the optimal growing temperature.

By including grid square fixed effects we get similar results. Here identification comes from within unit variation, deviation from the mean of the unit’s distance from the optimal growth temperature for cereals. In this sense we are again controlling for long term climatic conditions and identification is only coming from the random changes from this long-term trend. A Hausman test fails to reject the null hypothesis that the 2SLS fixed effects parameter estimates are different from the pooled 2SLS estimates, indicating that the controlling for latitude and longitude accounts for all time invariant aspects of climate. Similarly, I estimate the same instrumental variables model in first differences, where identification is coming exclusively from century-on-century changes in the propensity to support urban populations. The results are similar though, as expected, less precisely estimated. Although the results are nearly identical, in the first-differenced 2SLS estimates the
instrument fails to meet “rule of thumb” levels of strength (Stock and Yogo 2005).

3.3 Instrumental Variables Results

The instrumental variables estimates of $\beta$, the effect of a 100% change in urban population on the number of states forming on an arbitrary piece of land, are shown in Table 3. The effect sizes are rather large, a one hundred percent increase in total urban population on a given grid-square is expected to increase the number of states locating within that same unit by between one quarter and just over eight-tenths of a new state depending upon specification. Again, this is a rather large effect size as the inequality between geographic units is quite large. Taking the smallest 2SLS estimated effect size, 0.24, it would only take slightly more than a four fold increase in the total urban population to increase the number of states on a given unit by one.

The estimates in which identification is coming off of changes in optimal growing temperature after accounting for other possible climatic channels, those controlling for latitude and longitude as well as for unit specific heterogeneity (Models 3 through 6) are even larger, ranging from 0.67 to 0.88. Of these, the fixed effects (Models 4 & 5) and first-differencing (Model 6) estimates have a ready interpretation. Because they are identified off of within unit changes, they tell us how the number of states on a given geographic unit changes over time with changes in urban population. As an additional specification Model 5 includes the lag of the number of states. Again instrumenting for the Urban Population, the effect is roughly similar. Controlling for the number of states in a territory during the last century, a one-hundred percent change in the urban population still causes 0.58 of a new state to form on the same piece of geography in the next century.

Interpreted this way these results provide an explanation for why the central European core became simultaneously more urbanized and politically fragmented. As I showed in the second section of this chapter, the spatial distribution of urban population peaks in the central band of European geography running from Holland through Germany and into Northern Italy. Distance from this central corridor is highly predictive of the size of the urban population living within the bounds of a given geographic unit. For example, in 1800 at a distance of 2000 kilometers from this band a given unit is expected to have an urban population of approximately seven thousand. In
<table>
<thead>
<tr>
<th>Model:</th>
<th>2SLS</th>
<th>IV Neg Binomial</th>
<th>2SLS</th>
<th>IV Neg Binomial</th>
<th>2SLS</th>
<th>IV Neg Binomial</th>
<th>2SLS</th>
<th>2SLS</th>
<th>2SLS-FD</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(Urbanization$_{it}$)</td>
<td>.37</td>
<td>.15</td>
<td>.36</td>
<td>.16</td>
<td>.67</td>
<td>.24</td>
<td>.79</td>
<td>.58</td>
<td>.88</td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year Effects</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lat/Long</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Entrants$_{t-1}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First Stage Regressions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Temp$_{it}$ − 10.5]</td>
<td>-.26</td>
<td>-.27</td>
<td>-.09</td>
<td>.40</td>
<td>.23</td>
<td>-.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F statistic for Excluded Instruments</td>
<td>338</td>
<td>352</td>
<td>22</td>
<td>22</td>
<td>9.7</td>
<td>3.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R$^2$</td>
<td>.70</td>
<td>.71</td>
<td>.77</td>
<td>.002</td>
<td>.003</td>
<td>.003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1684</td>
<td>1684</td>
<td>1684</td>
<td>1684</td>
<td>1684</td>
<td>1684</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: 2SLS and IV Poisson parameter estimates of the effect of urban population on the number of states existing on a given 10,000$^2$ km piece of geography. Controls for latitude and longitude, year effects, and fixed effects, are continued in both the first and second stage regressions. 95% confidence intervals estimated from robust standard errors clustered by grid-square are in brackets.
the same year, directly on the line connecting Northern Italy and Holland, the urban population is expected to be more than ten times this amount. Making use of the lowest estimated effect of urban population, the grid square directly on this central band is expected to have two and a half more states on it than the quadrant 2,000 kilometres away. Using the largest estimated effect this difference doubles, with five more states expected to be found in the centrally located quadrant.

To better demonstrate the geographic distribution of political fragmentation as a function of urban population, Figure 11 plots the predicted values derived from the same model across space. Estimates are derived from the most conservative two-stage lease squares model from Table 3. We see that the break-up of Central Europe during the first half of the last millennium was caused by variation in urban growth.

3.4 Changes in The Costs of Statehood

The model presented in Section 2. argues in favor of a conditional relationship between the costs of statehood and the size of the economic surplus available for states to expropriate. In periods when the costs of statehood are high, the effect of an economic surplus on the number of states forming on a given piece of territory is expected to be attenuated. That is, even when there is a large economic surplus to extract, if the costs of statehood are high we should expect fewer states to form than would otherwise in periods when the costs are low.

To examine this set of hypotheses I consider the changing effect of urban population across time, showing that in periods associated with high costs of statehood the effect of urban population on state formation is less than in periods when the costs are low. To begin, I divide the data into two periods. The first, up to and including fifteen hundred, captures the period of low fixed costs of statehood, before the military revolution, and the second, after fifteen hundred, contains the period associated with the military revolution and high costs of statehood. If the prediction of model holds true, then the effect of urban population should be greater in the period before 1500 than it is after.

From Panel A. in Table 4 we see that, as expected, across specifications the magnitude of the effect in the period associated with high fixed costs of statehood is greater than that in the
Figure 11: The Fragmentation of Central Europe: The expected number of states as derived from the 2SLS estimate of Model 1 in Table 3, plotted across space. Darker colors denote a greater number of predicted states for a given piece of geography, lighter colors represent fewer predicted states.
### A.) Instrumental Variables Results

<table>
<thead>
<tr>
<th></th>
<th>Pre 1500</th>
<th>Post 1500</th>
<th>Pre 1500</th>
<th>Post 1500</th>
<th>Pre 1500</th>
<th>Post 1500</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(Urban Population)$_{it}$</td>
<td>.65</td>
<td>.16</td>
<td>1.21</td>
<td>.35</td>
<td>.57</td>
<td>-.05</td>
</tr>
<tr>
<td>Generalized Hausman Test:</td>
<td>(417.14)</td>
<td>(2.358)</td>
<td>(5.83)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F statistic for Excluded Instruments</td>
<td>226</td>
<td>436</td>
<td>10.4</td>
<td>30.5</td>
<td>10.0</td>
<td>45.7</td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lat/Long</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year Effects</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N = 1684, T = 4/3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### B.) Total Effect of By Year

<table>
<thead>
<tr>
<th>log(Urban)$_{1200}$</th>
<th>log(Urban)$_{1300}$</th>
<th>log(Urban)$_{1400}$</th>
<th>log(Urban)$_{1500}$</th>
<th>log(Urban)$_{1600}$</th>
<th>log(Urban)$_{1700}$</th>
<th>log(Urban)$_{1800}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.10</td>
<td>.62</td>
<td>.72</td>
<td>.43</td>
<td>.22</td>
<td>.16</td>
<td>.11</td>
</tr>
</tbody>
</table>

N=1684, T=7  Angrist-Pischke multivariate F test of excluded instruments: 507.5

Table 4: Panel A. shows the effect of urban population on the number of states forming on a 10,000 square kilometer grid, dividing the data by time period. The first period includes all years before and including fifteen-hundred. The second includes all years after fifteen-hundred. Panel B. shows the estimated total effect of urban population as it varies by year. The estimates are derived from the estimation of Equation 5. Ninety-five percent confidence intervals calculated from heteroskedasticity robust standard errors clustered by grid-square in brackets.
period associated with low costs. Moreover, based upon a generalized Hausman test, we see that
the parameter estimates on the log of urban population from each pair of models are statistically
different from each-other. To further show how the effect of urban population changes across time
I estimate the following model:

\[
\text{Entrants}_{it} = \alpha + \beta \cdot \log(\text{Urban Population}_{it}) + \sum_{t=1200}^{T=1700} \gamma_t \cdot T_i \times \log(\text{Urban Population}_{it}) + \eta_t + \epsilon_{it} \tag{5}
\]

Where \( T_i \) is a century indicator, \( \eta_t \) the century effect, and \( \eta_{it} \) a disturbance term. The total
effect of logged urban population in a given century is given by \( \beta + \gamma_t \). These estimated total effects
along with the ninety-five percent confidence interval are presented in the lower panel of Table 4.
Again, we see that there is a general decline across time in the effect of urban population on the
number of states forming. However, one should be cautious in interpreting these results. Although
the parameter estimates are declining across time, concurrently, inequalities in urban population
are increasing. Consequently, the changes across space in political fragmentation that one might
read into these results are in fact much less pronounced.

Within the theoretical context of the model and with the general empirical pattern presented
earlier, these results suggest that the fragmentation of Central Europe in the early half of the last
millennium was caused by the early revival of commerce and urban life. However, as the costs of
statehood increased over time, as the need to maintain ever complex bureaucracies and increasingly
large and expensive militaries, this effect diminished in magnitude.

In all, I have shown that the geographical coincidence of political fragmentation and urban
growth is causal. That is, Central Europe was divided into a large number of small states because
it was urban. Peripheral Europe was politically unified because it was less urbanized. Variation in
the revival of commerce and urban life caused variation in the number of states forming on a given
piece of geography.
4 Conclusion

Did war make states? This paper has provided evidence that if it did, it did so conditionally. In it I developed a formal narrative of European state formation that takes into account the contemporaneous effects of changes in both patterns of commercial activity and the ways in which wars were fought. The predictions made by this model are two: First, in places and times where the costs of statehood are high and the economic surplus available to states low the consolidation of political rule is expected. Second, when the opposite is true, when the costs of statehood are low and the available economic surplus is high fragmented political authority is predicted.

The predictions match two major historical trends scholars believe had major consequences for the origins of the state. First, changes in the production of violence associated with the military revolution have been argued to increase the costs of operating as an independent state. Second, the reemergence of commerce in some places in the early half of the last millennium has been argued to allow small political communities the resources necessary to claim and sustain independence. Moreover, I showed that the predictions made by the model are supported by the general patterns observed in the data. Within the productive European core fragmented political authority resulted from the expansion of trade and the reemergence of urban life. Outside of this region, in the continent’s less urban periphery, the increasing costs of maintaining the apparatus of the state forced the consolidation of large, territorial states.

Exploiting random climatic variation in the propensity of some places to sustain large populations, I then showed the proposed relationship between urban development and fragmented political authority to be causal. That is, I showed that the central European core was divided into many small political communities because it was urban and the periphery of the continent was dominated by large, territorial, states because it was comparatively less urban. In all, the number and size of states before the French Revolution has been explained as the outcome of a competitive process in which rational, revenue maximizing, agents compete over land of varying economic worth constrained by the costs of competition.
Appendix

Proof of Equation 2. From taking logs of the utility function and first order conditions with respect to \(a_i\) and \(b_i\) we can define the ratio:

\[
\frac{a_i}{b_i} = \frac{f(l_i + (l_{i+1} - l_i)p_i) \cdot (l_{i+1} - l_i) \cdot p_i(1 - p_i)}{f(l_{i-1} + (l_{i-1} - l_{i-2})p_{i-1}) \cdot (l_{i-1} - l_{i-2}) \cdot p_{i-1}(1 - p_{i-1})}
\] (6)

Similarly we can write

\[
\frac{a_i}{b_i} \cdot \frac{a_{i-1}}{b_{i-1}} = \frac{f(l_i + (l_{i+1} - l_i)p_i) \cdot (l_{i+1} - l_i) \cdot p_i(1 - p_i)}{f(l_{i-2} + (l_{i-1} - l_{i-2})p_{i-2}) \cdot (l_{i-1} - l_{i-2}) \cdot p_{i-2}(1 - p_{i-2})}
\] (7)

From the first order conditions we also get

\[
\frac{a_{i-1}}{b_i} = \frac{Y_i}{Y_{i-1}} \cdot \frac{1 - E_{i-1}}{1 - E_i}
\] (8)

Plugging this into Equation 8 we get

\[
\frac{Y_i}{Y_{i-1}} \cdot \frac{1 - E_{i-1}}{1 - E_i} \cdot \frac{a_i}{b_i} = \frac{f(l_i + (l_{i+1} - l_i)p_i) \cdot (l_{i+1} - l_i) \cdot p_i(1 - p_i)}{f(l_{i-2} + (l_{i-1} - l_{i-2})p_{i-2}) \cdot (l_{i-1} - l_{i-2}) \cdot p_{i-2}(1 - p_{i-2})}
\] (9)

Again, using the first order conditions we can re-write the right hand side of Equation 9 and rearrange terms to get

\[
\frac{a_i}{1 - E_i} \cdot Y_i = \frac{b_{i+1}}{1 - E_{i+1}} \cdot Y_{i+1}
\] (10)

Combining Equations 8 and 10 and rearranging we get

\[
\frac{a_i}{b_{i+1}} \cdot \frac{1 - E_{i+1}}{Y_{i+1}} = \frac{b_i}{a_{i-1}} \cdot \frac{1 - E_{i-1}}{Y_{i-1}}
\] (11)
References


L. Blaydes and E. Chaney. The feudal revolution and europe,Â’s rise: Political divergence of the christian and muslim worlds before 1500 ce. 2011.


G. Parker. The military revolution: Military innovation and the rise of the West, 1500-1800. Cambridge Univ Pr, 1996.


