State History and Economic Development: Evidence from Six Millennia*

Oana Borcan
University of Gothenburg

Ola Olsson
University of Gothenburg

Louis Putterman †
Brown University

July 3, 2014

Abstract

All since the rise of the first civilizations, economic development has been closely intertwined with the evolution of states. In this paper, we contribute to the literature on state history and long-run economic development in four ways. First, we extend and complete the state history index from Bockstette, Chanda and Putterman (2002) by coding the experience with states from the first state origins, 3500 BCE, up until 2000 CE. Second, we explore empirically the relationship between time since transition to agriculture and state age, as well as subsequent state history. Our estimated unconditional correlation implies that a 1000 year earlier transition to agriculture is associated with a 470 years earlier emergence of state institutions. We show how this relationship differs between indigenously- and externally- originated states. Third, we show that the relationship between our extended state history index and current levels of economic development has the shape of an inverted u. The results reflect the fact that countries that were home to the oldest states, such as Iraq, Egypt and China, are poorer today than younger inheritors of their civilizations, such as Germany, Denmark and Japan. This pattern was already in place by 1500 CE and is robust to adjusting for migrations during the colonial era. Finally, we demonstrate a very close relationship between state formation and the adoption of writing.

Keywords: State history, comparative development
JEL Codes: O11, O43, O50, N00

1 Introduction

History has shown that economic development often thrives in states where governments guarantee the rule of law and provide public goods for their citizens. In order to reach a deeper understanding of why some countries have good government and others do not, social scientists have become increasingly interested in studying the long-run patterns of institutional development within states. The roots of countries'

*We are grateful for useful comments from Carl-Johan Dalgaard, Jakob Gerner Hariri, and from seminar participants at University of Copenhagen and Brown University. We also thank Taewan Roh and Nicholas Carter for valuable research assistance.

†Corresponding author: Louis_Putterman@brown.edu

contemporary failures or successes have often been traced back to "critical junctures" far back in history.[1]

In this paper, we study the emergence of states from their first origin around 3500 BCE up until the present day and analyze how state development has interacted with economic development. More specifically, we attempt to make four distinct contributions to the literature. First, we complete the state history index initially developed by Bockstette, Chanda and Putterman (2002) for 159 countries. We extend the index from 1 CE backwards in time to the first origins of states around 3500 BCE and also code the 1950-2000 CE period, which was previously missing from the time series. Second, we use the complete index of state histories to study the determinants of the timing of state emergence and experience. Our estimates indicate a very strong and robust positive link between the time since the transition to agriculture and state emergence, as well as state history. Moreover, we explore the role of transition to agriculture and geographical characteristics in states that emerged indigenously as opposed to by conquest. Third, we analyze how our extended state history index correlates with various indicators of economic development. In particular, we show that the relationship between our extended state history index and current levels of economic development has the shape of an inverted u, implying that the very young and very old states have the least developed economies whereas the richest countries have intermediate state history scores. Lastly, our analysis is probably the first to document for a large cross-section of countries a very strong connection between state emergence and the adoption of writing.

The first of these objectives - the creation of a state history index for the BCE-period - is perhaps the most important contribution that we make. In line with the methodology in the original effort by Bockstette el al. (2002), we combine three dimensions of state development: 1) The existence of a state above tribal level; 2) Whether rule is internally or externally based; 3) The territorial coverage of the state in relation to current national borders. Our main source of information is Encyclopedia Britannica Online and the three indicators were coded for each of the 159 countries in our sample and for each 50-year period from the origin of the first states around 3500 BCE, yielding a panel data set with 17,490 country-period observations. The details of the construction of the index are described further below.

The work clearly involves several methodological challenges. For instance, how should a state be defined? In this regard, we follow the tradition of Service (1962), Carneiro (1981), Johnson and Earle (2000) and others, distinguishing between bands, chiefdoms, and full-fledged states. Unlike the other forms of governments, states are

[1] See for instance North (1990), Acemoglu at al (2006 and 2012), and Besley and Persson (2009).

further characterized by a centralized government with the ability to collect taxes, enforce laws, and mobilize forces for war. Using this definition, most sources seem to be in rough agreement about the time when states arise in different countries. Accompanying this paper is an extensive online data appendix where we motivate the coding for each country-period observation.

Another issue concerns the unit of analysis, which is the territory delimited by modern-day country borders, for 159 contemporary countries in the sample. It is a well-known fact that the borders of current countries sometimes have very little resemblance with the geopolitical logic in ancient times.[2] Furthermore, African country borders were often drawn without consideration of indigenous state formations and several of the American countries have experienced an almost complete replacement of their indigenous populations since the colonial era (Putterman and Weil, 2010) while also having borders unrelated to pre-colonial realities.

However, to the extent that researchers are interested in tracking the histories of countries in order to understand contemporary levels of development, the modern configuration of countries is still a natural point of departure. A potential alternative to using country borders could have been to divide the world into equal-sized grid cells and then study the history of states in each such cell.[3] This would entail a very different type of analysis with its own methodological challenges. We leave this for future work.

When we study the determinants of state emergence in a formal regression analysis, we demonstrate that the time since the adoption of agriculture alone explains about 65 percent of the variation in state onset. The regression coefficient for the unconditional association between state age and time since the Neolithic transition indicates that a 1000 year earlier transition implies a 470 years earlier state emergence. When we include continental fixed effects and geographical controls, the equivalent calculation gives us 430 years. The point estimate is a lot higher in countries where states originated internally as opposed to by conquest.

The state history data that we extend here were initially compiled by Bockstette et al. (2002) with the aim of using presence and duration of experience with macro polities as one of several potential indicators of societal complexity and level of technological advancement. Anthropologists including Service (1971), Johnson and Earle (2000), and Richerson, Vila and Mulder (2001) have described a rough continuum of modes of social organization and economic adaptation which range from foragers in which small bands are the principal social and political unit, to horticul-

[2] Although this is a valid critique of the approach used here, there are also numerous instances of countries where states from their inception have evolved in close proximity to current borders. Examples of such countries include Egypt, Norway, Sweden, China, and Japan.

[3] State history has been coded at the grid-cell level for sub-Saharan Africa after 1000 CE by Depetris-Chauvin (2014).

tural and pastoral groups organized as tribes, to intensive agricultural and industrial societies marked by larger populations organized into macro political units that typically display greater economic specialization and social stratification than tribes or bands. Presence of a large, domestically based state (as opposed to band or tribal arrangements or an externally imposed empire) can thus be conjectured to serve as an indicator of "level of development," one having the advantage of relatively good coverage in historical sources.

Bockstette et al. were interested in investigating the effect of early social and technological development on post-Second World War economic growth rates, and they assumed that the impact of very early experience would decay over time, so they did not attempt to code information on state presence before 1 CE or after 1950. They coded all countries with substantial populations for which relevant economic growth and other indicators were available, resulting in a sample of 104 countries, of which their analysis focused especially on 70 non-OECD member countries. They found a significant and robust correlation between state history and recent growth rate, and a significant bivariate correlation between state history and income level that was not robust to addition of controls. Roughly the same data set was also used by Chanda and Putterman (2005), and Chanda and Putterman (2007).

Bockstette et al.'s (2002) data were subsequently expanded to include more ex-Communist countries (studied by Iliev and Putterman, 2007), more African countries (studied by Cinyabuguma and Putterman, 2011), and a few other countries for which complementary income or other required data had initially been viewed as unreliable. With the larger data set, Putterman and Weil (2010) demonstrated that ability of state history to predict current level of development is greatly strengthened by replacing the state history that transpired on a given country's territory by the weighted average state history of the places in which current residents' ancestors lived in the past, an adjustment motivated by the large movements of populations especially from "Old World" continents to the Americas, Australia and New Zealand after 1500. Chanda, Cook and Putterman (2014) apply the same procedure to demonstrate "persistence of fortune" of ancestral lines in former colonies that display a "reversal of fortune" (Acemoglu, Johnson and Robinson, 2002) in the absence of such ancestry and migration accounting.[4]

The paper is organized as follows: In section 2, we provide an overview of the literature on the definition of a state and present the principles behind our data collection. In section 3, we present and discuss a number of stylized patterns that

[4]The state history data have also been employed in a number of other studies, receiving focal attention in Ang (2013a, 2013b), playing important roles in Ahlerup and Olsson (2012), Hariri (2012), Ertan, Putterman and Fiszbein (2012), and Daniele (2013), and being included as a control in a number of other studies. None of the above studies attempts to extend the information on states to include the BCE years or fill in the last half of the 20th Century.

emerge from the new data. In section 4, we carry out an econometric analysis of the determinants of the timing of state emergence and the relationship between economic development and state history. Section 5 concludes.

2 Data

One of the contributions of this paper is the construction of a comprehensive index of state history tracing political organization within the territories of modern-day countries as far back as historical and archaeological evidence allow it. In doing so, we build on the State Antiquity Index previously developed by Bockstette, Chanda and Putterman (2002). This index and its subsequent versions were constructed for up to 159 modern-day countries, covering a period between 1 CE and 1950 CE. However, for as many as 58 modern-day countries in the dataset, states had emerged on their territories before the Common Era. For half of these, the state history before 1 CE goes back at least eight centuries (e.g. Italy), and for some even over three millenia (e.g. Iran, Egypt). Conceivably, this early state experience may also have a long-lasting impact on the economic development of the regions where it accumulated. In addition, no picture of the current distribution of wealth in the world would be complete without accounting for the most recent historical events, between 1950 and 2000. This period was marked by the mass decolonization of African countries, the incidence of civil wars, and the expansion and contraction of various spheres of political influence.

Therefore we have extended the state antiquity index in two directions. First, we have coded the index for the territories of the 58 present-day countries for which evidence suggests the emergence of statelike institutions before 1 CE; the added periods of state history range from 14 years (e.g. Hungary) up to 3500 years (e.g. Iraq) before 1 CE. Second, we have coded the index of state history for the 1950-2000 period for all 159 countries in the sample. The case of Iraq, for which we record the longest state history of 5500 years, illustrates how we depart from the previous versions of the index, which recorded 1950 years of state history only.

For these new additions we surveyed and summarized the events throughout the states' development and we mapped them into real numbers within 50-year periods, according to the existing methodology.[5] This generated a richness of longitudinal information which is a useful resource for undertaking minute inspections into any stage in the evolution of state institutions. Finally, appending the BCE to the CE *Statehist* scores, we also computed the aggregate score for state history, from state emergence to 2000 CE, as well as an aggregate score to 1500 CE and to 1 CE

[5]The previous State Antiquity Index, version 3, is presented at: www.econ.brown.edu/fac/louis_putterman/antiquity%20index.htm.

(virtually any aggregation is possible).

Naturally, tracing the state evolution back to the millennia BCE entails pinning down the dawn of states in history, which is a major undertaking, given the scarcity of written records. Therefore, in recording the approximate date of state formation in a consistent manner, we needed to resort to a set of conventions aligned with the historical, political and anthropological understanding of the concept of "state". After clarifying what we refer to as "states" below, we proceed to describe the actual data coding protocol.

2.1 Defining the "state"

How do we know when a state has emerged? The first challenge stems from the question of how to define the state, hardly a novel dilemma in social sciences. The classical understanding of the "state" comes from Weber (1919), who defined it as an entity which "upholds the claim to the monopoly of the legitimate use of physical force in the enforcement of its order" (Weber, 1978, p. 54). This implies that we should be looking for evidence of the initial monopolization of power within a certain territory. However, there is also the question of the extent of this original jurisdiction: how large is the population or the territory subject to the power monopoly? Is, for instance a village with 100 tribesmen, led by a chief, large enough to classify as "state"? It appears that we need to find an appropriate threshold to distinguish between small and large scale political organization. In some cases the distinction is unambiguous: there are written records attesting the date when large-scale centralized organization within the territory of certain modern-day countries was originally attained. For instance, the land of what is today Belgium came under large-scale political organization for the first time between 59 and 52 BCE, when it was integrated in the Roman Empire, having been inhabited by people with no more than tribal organization prior to 59 BCE. This is most often the case of states originated in colonies or expansion of pre-existing states (we call these *externally- or non-indigenously- originated states*). However, for territories in which the state was an indigenous development, i.e. *internally- or indigenously- originated states*, evidence of this transition is suggestive at best.[6]

Thus, the first task is to decide when to assign the first positive scores, marking the emergence of large scale political organization. We take the first documented manifestation of the presence of an overarching governing body, e.g. a local kingdom, or rule by a colonial power, to yield the first positive score for the ruled territory. Crucially, in order to qualify for a positive score, we adopt the convention that

[6]Internally originated states include both pristine states, where power centralization was a completely original development, as well as those with an indigenous but potentially externally-inspired origin.

the evidence should point to a type of political organization above the tribal level. This principle is not arbitrary, but is in fact consistent with established sociological and anthropological taxonomies of human societies throughout their evolution. For instance, Johnson and Earle (2000) proposed a division of societies into local group (further divided into family, village and the Big Man group) and regional polity, which can be a chiefdom or a state. This distinction was necessary to separate the small-scale organization of local groups from the next level of political organization - with the chiefdom at the lower bound. This distinction goes back even earlier, owing to Service's (1962) proposed typology of bands/tribes/chiefdoms/states. An additional indication of how to identify state institutions comes from Charles Tilly, who understood the state to be "a coercion-wielding organization that... exercises clear priority over all organizations within substantial territories. The term therefore includes city-states, empires, theocracies, and many other forms of government, but excludes tribes." (Tilly, 1990, p. 1).[7]

In practice, however, following this principle is not always straightforward. In some cases we could rely on written history to assign a date for state onset (e.g. in the case of Syria, the Ebla tablets dated 2600-2500 BCE document the existence of a flourishing Syrian kingdom). In other instances we had to rely on archaeological data, which compelled us to consider any evidence of emerging political or administrative cohesion above tribal level as an indication that a governing body came into existence. Accordingly, we sometimes followed a "diagnostic traits" approach, having to consider material manifestations, or consequences of monopolization of power, as an "archaeological confirmation of the process of state formation" (Jones and Kautz, 1981, pp. 16-17). These material manifestations can be monumental structures, such as palaces, temples or large urban settlements etc. In the case of Iraq, for instance, there is the transition from small to large urban centers with grand architectural structures such as Uruk in the middle of the 4th millennium BCE. Admittedly, the drawback of this "symptomatic" approach is that it blurs the boundary between state and civilization and it is susceptible to misclassifying an emerging or transient civilization into a state in the Weberian sense.

The second task is to recognize and mark the transition from chiefdom to fully-fledged state. Following the paradigm of the evolution of pristine states from chiefdoms (see e.g. Carneiro 1981, Earle 1987, Flannery 1995, Marcus 1992, Spencer 1990, Spencer and Redmond 2004), we mark this distinction in our data by assigning the following values: Band/tribe is marked by a rule score of 0, paramount chiefdom is assigned 0.75 and fully-fledged state receives the value 1. Robert Carneiro is a staunch proponent of the intermediary role of the paramount chiefdom as the evolutionary link between the stage of autonomous bands or tribes and the state. In

[7] We thank Jacob Gerner Hariri for useful references on the matter of state definition.

his definition, the paramount chiefdom is "an autonomous political unit comprising a number of villages or communities under the permanent control of a paramount chief" (Carneiro, 1981, p. 45), while the state is "an autonomous political unit, encompassing many communities within its territory and having a centralized government with the power to collect taxes, draft men for work or war, and decree and enforce laws" (Carneiro, 1970, p. 733). Although simple chiefdoms fall short of the notion of supra-tribal polity, paramount chiefdom which incorporates multiple individually substantial chiefdoms can be understood as a form of incipient state. Hence we decided to begin according partial weight when a polity reaches this level.

While it is difficult to know exactly where the chiefdom ends and where the state begins, we have made efforts to draw a sensible line where the evidence suggests a remarkable evolution in socio-political organization. Such is the case of Mexico, where we assign a score of 0.75 to the period 450 - 100 BCE for the early urban settlements at Chiapas and Oaxaca. We then raise this score to 1 in 100 BCE when large-scale urban growth at Teotihuacan and the development of previously missing institutions such as a standing army warrant the status of fully-fledged state. While this kind of judgement is not uncontroversial, it is the most feasible approach given limited documentary resources. We further detail the assignment of scores in the next section.

2.2 Constructing the index

The construction of the index for the BCE period follows the same principles developed by Bockstette et al (2002):

1. For every modern-day country in the sample, we survey the historical and archaeological evidence to identify the time of emergence of the first state institutions on the territory of the respective country (in accordance with the ground rules outlined above). We divide the time following that date into 50-year periods, or half centuries. The oldest state, established on the land of today's Iraq, is assigned 70 periods from 3500 BCE until 1 CE and 110 periods in total, including 1 - 2000 CE. Therefore 3500 BCE is the joint starting point of our analysis for all countries. In the case of Bulgaria, for instance, initial presence of supra-tribal rule is attested from 516 BCE. Hence the first period with positive scores is 550 – 501 BCE whereas all previous periods have a zero score.

2. For each modern-day country i and 50-year period, indexed by t, we classify the information regarding the state experience within that time frame into 3 major components, indicated by the superscript c. Hence, z^c_{it} is the score for

component c in country i for period t. The score is based on the following questions:

(a) Is there a government above the tribal level? This first score component, z^1_{it} is assigned 1 point if the answer is yes, 0.75 points if the organization of the state can at best be described as a paramount chiefdom, and 0 points if the answer is no.[8]

(b) Is this government foreign or locally based? The second component z^2_{it} is assigned 1 point if the rule is locally based, 0.5 points if foreign, i.e., the country is a colony, and 0.75 if the rule is exercised by a local government with substantial foreign oversight.[9]

(c) How much of the territory of the modern country was ruled by this government? The third component z^3_{it} is assigned a score as follows: 1 if more than 50 percent of the territory comprising the modern country is under some rule of a given state during the given 50-year period; 0.75 points if the ruled territory is between 25 percent and 50 percent; 0.5 points if the ruled territory is between 10 percent and 25 percent; 0.3 points if less than 10 percent of the territory is under some rule. In cases where substantial parts of the territory were under the rule of distinct states, we downgrade the z^3_{it} score to the next possible value (e.g. if more than 50 percent of the territory is under the rule of one state, then $z^3_{it} = 1$, but if the same proportion of the territory is divided between two states, $z^3_{it} = 0.75$).

3. We denote every 50-year period from 1951-2000 CE back to 3500 BCE by t where $t = 0$ is the most recent period and $t = 109$ for 3500-3451 BCE. For each t on the territory of country i, we compute a composite *State index* score by multiplying the three components by one another and by 50:

$$s_{it} = z^1_{it} \cdot z^2_{it} \cdot z^3_{it} \cdot 50 \tag{1}$$

If changes in the structure/origin/territory of the rule incurred within a 50-year window, the period t was subdivided into subperiods $\theta = 1, 2, 3...$ such that $z^1_{it\theta}$ would be the sub-period scores for component 1 in country i during

[8] In some special cases, we assign special values such as 0.5 for z^1_{it} due to of radical uncertainty with respect to the existence of rule on certain territory. The reader is referred to the online data appendix for a more detailed discussion of coding exceptions.

[9] In some cases where a given territory is divided into multiple powers with different rule origins, z^2_{it} is assigned a simple average of the basic scores corresponding to those origins. For instance, 0.875 is the average of the 1 and 0.75, for a territory with one part locally-based rule and one part locally-based with foreign oversight.

50-year period t and subperiod θ. The overall score s_{it} was then computed as the weighted average of the sub-period scores:

$$s_{it} = 50 \cdot \left[(z^1_{it1} \cdot z^2_{it1} \cdot z^3_{it1}) \cdot w_{it1} + (z^1_{it2} \cdot z^2_{it2} \cdot z^3_{it2}) \cdot w_{it2} + \cdots \right] \tag{2}$$

The weights $w_{it\theta}$ are obtained by dividing the number of years in each sub-period θ by 50. By applying these formulas we obtain a score s_{it} for every half century from 3500 BCE to 1950-2000 CE.

4. By joining the BCE- with the preexisting CE-era series, we obtain a complete description of the history of state presence for every modern-day country. In a small number of cases, harmonizing the scores around 1 CE required adjustments to the initial CE index. However, changes were minor and the correlation between the original and the new scores for the period 1 CE - 1950 CE is over 99 percent. The final aggregation of all 50-year scores s_{it} leads to one comprehensive index of the cumulative state history - *Statehist* - for country i, calculated using various rates $\rho \geq 0$ for discounting historical scores. The index is normalized by putting in the denominator the score of a hypothetical state with full discounted scores between 3500 BCE and the period of interest τ:

$$S_{i\tau} = \frac{\sum_{t=\tau}^{109} (1+\rho)^{\tau-t} \cdot s_{it}}{\sum_{t=\tau}^{109} (1+\rho)^{\tau-t} \cdot 50} \tag{3}$$

This cumulative *Statehist* index $S_{i\tau}$, which ranges between 0 and 1 and should be carefully distinguished from the "flow" *State index* observations s_{it} during each individual time period, can be calculated at virtually any point in history $\tau = \{0, 1, ...109\}$. Although the contemporary level of the *Statehist* index for 2000 CE (S_{i0}) is what we are primarily interested in, we calculate it also for 1500 CE (i.e. 10 periods back such that $\tau = 10$), and for 1 CE (i.e. when $\tau = 40$). The choice of discount factor ρ warrants some discussion. The previous literature has set the convention at $\rho = 0.05$, in light of the reasonable assumption that the more distant past matters less today than recent history. With the additional data, however, a 5 percent discount rate gives insufficient weight to the long stream of s_{it}-scores before 1 CE when the aggregation is done at 2000 CE or even at 1500 CE. In fact, applying this discount rate would lead to an extended *Statehist* score that has a correlation of up to 99.3 percent with the 5 percent discounted 1 - 1950 CE score. The implication is that all information before 1 CE would receive negligible weight. While it of course remains to be seen below just how useful placing weight on the distant past will be, our convention in what follows will be to employ the 1 percent discount factor of the normalized *Statehist* score in the forthcoming analyses.[10]

To answer the three questions (a-c) above in a manner that is consistent across periods, we relied mainly on information in the Encyclopedia Britannica Online. We detail on the data sources and illustrate the coding process and further data aggregations in the online Appendix B.

3 Patterns of state evolution

In this section, we will present some of the key stylized patterns that arise from the complete state history time series introduced in this paper. Our purpose is to get a feel for the data and potentially some perspective on the role of state history in economic development.

The first key pattern concerns the evolution of states in the world as a whole: *The evolution of state institutions in the world follows approximately an exponential upward trend with periods of rapid growth punctuated by periods of stagnation.*

Figure 1

This pattern is visualized in Figure 1 which shows the log of the aggregated percentage score for all contemporary countries in our sample at each 50-year period on the vertical axis and year on the horizontal axis. The percentage score in period t is calculated as *State index world* $(t) = 100 \cdot \sum_{i=1}^{N} s_{it} / (N \cdot 50)$ where $N = 159$ is the number of included countries and where $s_{it} \in [0, 50]$ is the state history score for country i during 50-year interval t, as described above.[11] A value close to 0 percent in this world index indicates that there is no sign of state presence in any of the included countries in period t whereas a score of 100 means that all 159 countries reach the maximum value $s_{it} = 50$ in our state measure during that period. Since many modern-day countries did not have full states in the spirit of our definition during the entirety of last time period 1950-2000, the aggregate percentage in the graph is about 88 percent ($\ln 88 = 4.48$) at the end of the time series.[12]

The logged percentage score for the world crawls around a fitted log-linear trend line. A simple regression of the aggregate world state index score on time shows that the fit is $R^2 = 0.90$. However, it is also clear that several periods are characterized by rapid state evolution whereas other periods are marked by a general decline. The

[10] The correlation between the 1 percent discounted Statehist index calculated for the year 2000 and the 1 - 1950 CE 1 percent discounted Statehist index is 0.93. The correlation between the former and the 1 - 1950 CE 5 percent discounted Statehist index is 0.86.

[11] Note that *State index world(t)* describes the "flow" level of state development in the world in period t and not the cumulative "stock" of state experience.

[12] Many states were de-colonized part way through the period, a number emerged from the Soviet Union and Yugoslavia, others experienced contending governments or state failure, etc.

first boom in state emergence appears already 3500-2300 BCE, which then ends with a long period of stagnation. The other major stagnations in the figure happened around 1750 BCE, 1200 BCE, and 400 CE. A second period of rapid growth was 850 BCE-1 CE during the Iron Age. From just after the collapse of the Roman empire around 450 CE, aggregate state emergence has shown a steady upward trend.

The aggregated graph summarizing the state history for the world as a whole in Figure 1 hides important differences among the major agricultural core regions. In Figure 2, we disaggregate the evolution of state history into the four main agricultural core areas: Western, Eastern Asia, Sub-Saharan Africa, and the Americas.[13] These four areas are created on the basis of how Neolithic agriculture and civilization spread during early historical times. We also show the trend for the world as a whole. The important Western area, for instance, comprises all modern-day countries in Europe, North Africa, Middle East, Central Asia, Iran, Pakistan and India, including the early civilizations of Mesopotamia and Egypt. People in this region adopted the agricultural production package from the Fertile Crescent, including domesticated crops like wheat and barley and animals like goats and sheep, and the states came into frequent contact at least from the 1st millennium BCE onwards.[14]

Figure 2

When we divide up the world in this way, some striking historical differences between the regions appear: *State evolution started earliest in the Western area, with Eastern Asia lagging behind until rough convergence (indeed, initially overtaking) around 500 CE, with the other regions gaining steam later and all converging only toward the end of the era of European colonialism.*

On the vertical axis in Figure 2 is the *State index* (in percent) for the countries included in the different categories, but unlike in Figure 1, we do not log the time series this time. As noted above, state emergence was earliest in Eastern Asia and in the Western region. Interestingly, both of these early civilizations took off on a more rapid path after 850 BCE. By the time of the Western Roman collapse after 450 CE, Asian state development overtakes the Western one for the first time.[15]

[13]The division into agricultural core areas follows the practice in Morris (2010) and Olsson and Paik (2013). Combining the two or three distinct agricultural cores of the Americas identified by some writers is a convenient simplification.

[14]We draw the boundary between the Western and core region Eastern Asia on the border between India and Bangladesh. India has clearly been influenced by both Western and Eastern traditions, although its earliest civilization in the Indus Valley was of Western origin. The Americas are generally regarded as having had three agricultural core areas in North, Central and South America. Agricultural practices in Sub-Saharan Africa spread in the Sahel and in the West African cradle of the Bantu expansion. See Bellwood (2005) for an exhaustive account of the Neolithic transition.

[15]See Morris (2010) for a detailed comparative analysis of Western and Eastern history since the Neolithic.

The other two regions, the Americas and Sub-Saharan Africa, clearly lag behind, in particular after the Eurasian turning point 850 BCE. From about 500 CE, the pace of state emergence starts to increase in Sub-Saharan Africa. When the colonial era starts in the late 15th century CE, the lagging regions experience a dramatic increase in the *State index*. This increase is of course to a great extent driven by the emergence of colonial states, created by European powers. By the final period of observation (1951-2000), the Americas has the highest score on state presence among all regions in the world.

In Figure C1 in the online appendix, we zoom in on the last 550 years of state history. This period of colonization witnessed some dramatic reversals in terms of economic and political development (Acemoglu et al, 2002 and Hariri, 2013).[16] One striking observation is that the territories that constitute today's Western offshoot countries displayed no signs of state emergence until the 1550-1600 period, placing them last among the regions in the initial centuries. After 1750, state development took off in these countries and reached 100 percent in the 1950-2000 interval. Latin America & Caribbean experienced a similar increase and also had a particularly quick development after 1750. Eastern Asia, on the other hand, had a long decline from the late 1600s which was not halted until the 20th century.[17] Africa had a state history index on par with West and Central Asia in the latter half of the 1800s but diverged as a result of the European scramble for African colonies in the 1880-1900 period. Decolonization after 1960 then brought convergence in levels of local state presence between Africa and the other regions.

4 Putting the data to work: Initial explorations

Having constructed and provided an initial description of our data, we are now ready to explore its usefulness in some initial empirical exercises. This section is organized around three major areas of analysis. In the first subsection, we study the determinants of the origin of states and duration of state history. In the second subsection, we analyze the relationship between state history and current economic development. In the third, we investigate the relationship between state history and indicators of historical economic development: population density, level of urbanization, general technological sophistication, and the emergence of writing.

[16] In the graph, we have divided the world up into Europe, Eastern Asia, West and Central Asia, Latin America & the Caribbean, Africa and Western offshoots (Australia, Canada, New Zealand, and the United States). For this later period, we argue it is reasonable to split up the Western area so that the largely Muslim West and Central Asia is a category of its own. Furthermore, countries such as United States and Canada had a very distinct history from the other parts of the Americas, which we refer to as Latin America & Caribbean.

[17] This was due to the fact that the modern territories of Indonesia and Sri Lanka were colonized by Europeans while for instance the government of Laos was increasingly dominated by foreign Asian powers.

4.1 State origins and persistence

In this section we explore the relationship between the timing of the transition to Neolithic agriculture and the date of state origins. It has been generally acknowledged in the literature that the adoption of an agricultural technology for food production, based on domesticated plants and animals, was also associated with a sedentary lifestyle, a dramatic increase in population density, and a socially stratified society with a dominant elite controlling a surplus from food production (Diamond, 1997; Johnson and Earle, 2000; Peregrine et al, 2007; Petersen and Skaaning, 2010). In such dense agricultural societies, chiefdoms eventually evolved into states with an ability to tax their population and to draft men for war or for the construction of extensive public works such as temples, irrigation systems, and city walls. Roughly 5,000 years after the first emergence of agriculture in the "Fertile Crescent" in the Middle East, the first known state appeared in Uruk around 3,500 BCE. China, Mesoamerica, and the Andes all likewise appear as cases in which a more-or-less independent flowering of agriculture was followed many centuries later by the pristine emergence of states. How is the emergence of agriculture and that of states linked statistically on a global scale? Is this link the same in states originated internally as opposed to by conquest?

As our baseline setup, we use a multiple linear regression model with a measure of state experience as the dependent variable:

$$State_i = \alpha_0 + \alpha_1 \cdot Agyears_i + \alpha'_j \cdot Z_i + \alpha'_k \cdot X_i + \epsilon_i \qquad (4)$$

In the equation above, the dependent variable is measured in two ways: 1) *State age* - the number of years elapsed in 2000 CE since a state/chiefdom first came into existence in the territory of modern-day country i. By using this variable we capture the timing of the actual state emergence in the regression. 2) *Statehist* ($S_{i\tau}$) – the cumulative state history index of the country evaluated at period τ, using a 1 percent time discount rate. This variable thus captures the experience with state institutions since the first emergence of a state on the country's territory (in terms of autonomy as well as territorial coverage and unity of the rule) until τ. We will mainly use the score for 2000 CE ($\tau = 0$), but we will also calculate the score for 1500 CE and 1 CE.

The main independent variable $Agyears_i$ measures the time before present since the Neolithic transition to agriculture in the country-area in question and is taken from Putterman with Trainor (2006). As discussed above, our key hypothesis is that $\alpha_1 > 0$, implying that country-areas where agriculture emerged earlier should have experienced both an earlier state formation and a longer state history. The magnitude of the estimate informs us about the exact relationship in years between the appearance of agriculture and states. Furthermore, when we use *State age* as the dependent variable, we expect to find a combination of estimates $\hat{\alpha}_1 \in (0, 1)$ and $\hat{\alpha}_0 < 0$, implying that agriculture generally precedes states.

The vector Z_i includes other historical control variables, for instance a variable *Origtime*, capturing the approximate time since the first settlement of a country territory by anatomically modern human beings. This variable was introduced by Ahlerup and Olsson (2012) as a determinant of the variation in levels of ethnic diversity across the world. Humans first appeared in East Africa about 160,000 years ago and then spread to other continents, reaching places like Scandinavia and Southern Argentina very late (the mean date for first settlement is 58,917 years ago). Ahlerup and Olsson (2012) showed that this measure was positively associated with current ethnic diversity. In the regressions below, we investigate whether this even deeper historical variable had any impact on the timing of state emergence.

The vector of geographical controls X_i includes variables such as the latitude of the centroid of the modern-day country i, whether the country is landlocked, its distance to coast or ocean-navigable river, average elevation, the land suitability for agriculture, climatic variables for temperature and precipitation, and the risk of malaria.[18] In most of the regressions, we also include continent dummies. These variables, as well as other variables employed in the empirical analysis, are summarized in Table 1 below.

Table 1

We present the estimation results for equation (4) in Table 2 for both *State age* and *Statehist*. The main estimate of interest is that associated with timing of agricultural transition α_1, and is the first coefficient in all specifications. In column (1), we present the unconditional estimate of the impact of timing of Neolithic transition on *State age*. The coefficient is 0.471, implying that a 1000-year earlier transition to agriculture is predicted to have been associated with a 471 years earlier first state formation.[19]

[18]These variables are taken from the Portland Physical Geograhy dataset and from the dataset compiled from various other sources by Ashraf and Galor (2013). We detail on the variable definitions in the appendix.

[19]Petersen and Skaaning (2010) provide the only other econometric estimate of which we are aware regarding the impact of timing of agricultural transition on emergence of states in a cross-section of countries, reporting a coefficient value of 0.406. Apart from this partially overlapping regression exercise, with its encouragingly similar coefficient, the scope of their paper is different in that we develop and analyze a full historical series for state presence in the BCE era, whereas they identify only the single data point, year of state emergence, their aim being to trace the impact of biogeographic conditions on state emergence through the channel of adoption and diffusion of agriculture. Their analysis thus lacks counterparts to our exercises on the impact of agricultural transition on *statehist* and to those distinguishing internally from externally originated states, as well as to all remaining parts of our paper. To conserve space, we detail differences in data and specification of the regression in question in our Appendix A6.

Inclusion of the control variables in column (2) and the continent dummies in (3) lead to a slightly lower estimate of α_1. In all specifications the coefficient remains strongly significant. All in all, these estimates confirm the catalytic impact of transition to agriculture on the emergence of states.

Table 2

The unconditional relationship is shown in Figure 3 where observations are distinguished by continent. We also show the line consistent with *State Age* = *Agyears*.[20] In a few African countries (like Sudan and Botswana), agriculture and states were introduced at the same time from outside. In all other countries, states evolved later. For countries close to the mean level of *Agyears* (4,717), the transition to agriculture is predicted to precede state emergence by approximately 3,050 years.[21]

Figure 3

In columns (4)-(6), we use the cumulative *Statehist* index for 2000 CE (S_{i0}) as the dependent variable. α_1 is significant also here but displays a lower magnitude. The main finding is that the timing of the Neolithic transition does not only affect the onset of state history, it also has a positive effect on state persistence throughout history. This result is robust to the inclusion of squared geographical controls (see Table D1 in the online appendix).

The control variables also reveal some interesting patterns. The coefficient for *Origtime* is never significant, suggesting that settlement events far back in prehistory did not have any direct impact on state formation. *Latitude* (of modern-country centroid) does not seem to influence state experience either. *Elevation*, however, has a positive impact. A plausible reason for this observed relationship seems to be the natural protection that a varied landscape (particularly with mountains) could provide, which would favor the better and/or earlier consolidation of large-scale politically-organized societies.

Among the other geographical variables, temperature has a positive and significant impact whereas the estimate for precipitation is negative, suggesting that hot and dry places (like Egypt and Iraq) were favorable for state emergence and persistence. An increase in temperature by one degree Celsius would, ceteris paribus, imply a 72-93 years earlier emergence of states according to the estimates in columns (2)-(3). Furthermore, being located on a landlocked territory or in a malaria-prone area also strongly delay state formation. Modern-day landlocked territories experienced the dawn of their first state almost 500 years later than non-landlocked ones.

[20] The line is also equivalent to regression parameters being $\alpha_0 = 0$ and $\alpha_1 = 1$.

[21] Figure C2 in the appendix displays the unconditional relationship between *Statehist* and the timing since the Neolithic transition.

Apart from their likely influence on the timing of the Neolithic transition, the variables capturing land suitability for agriculture and percent of arable land have no direct impact on state history or emergence.

As noted in section 2.1, the process of state formation is expected to differ between internally- and externally- originated states.[22] In Table 3, we subdivide the sample into 78 internally originated and 71 externally originated states with the purpose of understanding whether or not the agricultural transition timing and other factors influence domestic and foreign state formation heterogeneously.

Table 3

The estimate for *Agyears* is quite different between the two types of countries in Table 3; the coefficients imply that 1000 years earlier transition to agriculture is associated with 530 years earlier state emergence in internally-originated states, as opposed to just 300 years in externally-originated ones. The correlation is weaker in the latter states, likely due to other factors, such as geography or unobserved characteristics of their territories, driving both the conquest/colonization of their territories, as well as the introduction to agriculture. In internally-originated states the variable transition to agriculture is arguably more exogenous than in the case of externally-originated states. This is supported by the fact that in all cases of locally-based state emergence, agriculture preceded the emergence of large-scale political organization, whereas for some of the 71 externally-originated states, agriculture and the state arrived together. Moreover, for roughly half of the cases of locally-based state emergence, we documented a gradual transition from no to large-scale political organization. This indicates that, in certain territories, states gradually emerge long after the adoption of agriculture, as an intensification of economic and political activities, while in others agriculture might be introduced along with state institutions.[23]

A simple t-test of equality of coefficients in columns (1)-(2) reveals that the estimates are significantly different. The difference between the two estimates is

[22]We use the initial z_{it}^2 score to draw the distinction between the two. Specifically, we take the state to be internally-originated if the initial $z_{it}^2 = 1$, and externally-originated if the initial $z_{it}^2 < 1$. Another interesting distinction would be that between pristine states (with an entirely original development) and states created by local actors but in regions in which knowledge of the state concept had diffused by outside example (Mayan rulers could not know of the precedents of Mesopotamia, for example, but Hittite ones almost certainly did). Although we did not attempt to identify which cases could be considered strictly pristine in this sense, we suspect that their number would be too small to support statistical analysis.

[23]For externally-originates states, the endogeneity problem could occur due to the selection into the sample of colonially-based states, which may depend on their level of exposure to agriculture. One could expect the association with agriculture to overestimate the true impact of agriculture on state emergence in this sample (for instance due to geographical proximity to core agricultural areas, driving both earlier transition to agriculture and earlier political organization). However, the results suggest that selection (not controlling for the characteristics of these territories) leads to an underestimated impact of agriculture.

slightly reduced when control variables are added in columns (3)-(4). This difference is mitigated and becomes statistically insignificant when we control for continent fixed effects in columns (5) and (6).

4.2 State history and economic development

It is a well established empirical fact that history, recent and distant, has shaped the economic development of nations in ways that, to this day, still reverberate in their economies. Whether initial biogeographic endowment and transition to agriculture (e.g. Hibbs and Olsson, 2004; Olsson and Hibbs, 2005, Galor and Moav, 2007) or past technology adoption (Comin et al. 2006, 2009), early and productive starts have been typically shown to translate into better income and institutions in present times. The experience with state institutions has been put forth as one of the important correlates of the current wealth distribution in the world. Specifically, from its original development, the State Antiquity index has been shown to be positively associated with 1995 income and with the 1960-1995 GDP growth rate (Bockstette, Chanda and Putterman, 2002). The index of state history (along with the time from the transition to agriculture) was also shown to predict income levels today even better when adjusted by the post-1500 population flows, which accounts for the colonial era migrations (Putterman and Weil, 2010).

In short, previous work has largely agreed upon the fact that a linear positive association between long-run state history and current development exists. However, as scholars have acknowledged, the present shares complex links with the past. For instance, pre-1500 economic advantages seem to have become relative disadvantages among colonized countries during the colonial era (Acemoglu, Johnson and Robinson, 2001, 2002), although the effect seems attributable to large-scale migration (for instance, of Europeans to North America, Chile and Australia; see Chanda, Cook and Putterman, 2014). As of late, this idea of reversal has been revisited, pointing to a negative association between the time from Neolithic transition and current income levels in the Western agricultural core - Europe, North Africa and Southwestern Asia (Olsson and Paik, 2013). Moreover, the long-run persistence literature has begun to reveal nonlinearities in how events in the very distant past affect economic development. For instance, the migration out of Africa is argued to have generated a wide array of genetic diversity levels in human populations around the world. In turn, predicted genetic diversity displays an inverted-u shape relationship with indicators of economic development, including per capita income in 2000 (Ashraf and Galor, 2013).

With these developments in mind and with the new data on the extended state history index, we revisit the relationship between the degree of exposure to state institutions and current output. The questions we seek to answer are: 1) *Is there a* *relationship between state history, as measured from state emergence to the present, and current income per capita?;* 2) *Could a quadratic function describe the relationship between state history and GDP per capita in 2000 better than a linear function?*

The first question is motivated by the fact that in previous analysis the *Statehist* data was limited to the period 1 - 1950 CE. This effectively forced very old states such as Iraq or China to take similar values with intermediate states, such as England (the U.K.). The new data allows us to correct these shortcomings. Therefore, in the spirit of previous works, we have regressed log per capita GDP in 2000 against the extended index, and found that the coefficient is positive, significant, and slightly larger than if we used the 1 - 1950 CE index instead (see results below).

The second question is justified by the empirical observation that old states like Iraq, Turkey and China are poorer today than younger states like Britain, Denmark and Japan. The natural next step is to allow per capita output to vary non-linearly with state history.

Figure 4 illustrates the essence of our findings. On the Y-axis we have the logarithm of GDP per capita in 2000 and on the X-axis we have the extended *Statehist* (normalized with respect to 3500 B.C.E - 2000 CE and computed using a 1 percent discount rate per period). The figure displays a scatter plot of all countries in the sample, while also allowing for a *quadratic* fit of the relationship between output and *Statehist.* A hump-shaped relationship emerges when using the extended *Statehist.* The immediate implication is that states with extreme values of *Statehist* fare worse in terms of per capita GDP in 2000 than states with intermediate levels of *Statehist*, as measured by the extended index. However, this relationship is not observed when using the restricted *Statehist* 1-1950 CE, which only shows a relative disadvantage of very young states compared to all other states (see the strictly monotonic increase in log per capita GDP in 2000 when *Statehist* increases in figure C3 in the online appendix).

Figure 4

The figure above displays the unconditional relationship between income and *Statehist.*[24] The question arises whether or not this is a direct relationship or if it merely reflects other historical forces at play or natural conditions which may have shaped both the history of state institutions and current wealth. In order to investigate this issue, we set up the following model:

$$Log(GDPpc2000)_i = \beta_0 + \beta_1 \cdot Statehist_i + \beta_2 \cdot Statehist_i^2 + \beta_j' \cdot Z_i + \beta_k' \cdot X_i + \lambda_c + \epsilon_i \quad (5)$$

[24]This quadratic relationship is evident also when we divide the sample into internally- and externally- originated states and when we use the ancestry- adjusted *Statehist* index. See Figures C4-C6 in the appendix.

On the left hand-side of equation (5) we have per capita GDP in 2000 in logarithmic form. On the right-hand side we include our main independent variable, *Statehist,* both linear and squared, to account for the quadratic relationship. Z_i is a vector of historical controls including: $Agyears_i$, the time since Neolithic transition on the territory of country *i,* and $Origtime_i$ - the time since first human settlement on the territory of modern-day country *i.* In a more flexible specification, we include the square of $Origtime_i$ and a linear control *State age*$_i$. X_i is a vector containing all the geographic controls included in equation (5) above. λ_c is a vector of continent fixed effects. The results using the *Statehist* index are displayed in Table 4. In panel A, we use the new *Statehist* index, while in panel B, the *Statehist 1 -1950 CE* data. Columns (1)-(4) in Table 4 present the results without controlling for geographic characteristics. In columns (6)-(7) we present the results within continents.

Table 4

Our main coefficients of interest are β_1 and β_2, which estimate the relationship between current per capita income and state experience. In column (1) we display the simple association between per capita income and *Statehist,* which is positive and similar in magnitude across the two panels, but slightly less precisely estimated when the independent variable is (the new, extended) *Statehist*. In column (2) we add the squared *Statehist*, and the results mirror the pattern conveyed by Figure 4: In panel A, both coefficients are significant at 1 percent, β_1 is positive, while β_2 is negative, which confirms the concave relationship between log per capita GDP and state history. By contrast, in panel B, the counterpart of this specification using *Statehist 1 -1950 CE* displays coefficients with the same signs but much smaller and insignificant (the coefficient of the quadratic term is close to zero).[25]

We move directly to column (4) in panel A, where we introduce the first historical control - $Agyears$ (shown to be positively significantly correlated with the dependent variable in column 3, for comparison purposes). Its inclusion hardly changes the signs and the magnitude of the coefficients of the *Statehist* terms. Moreover, the effect of the time from transition to agriculture is insignificant. When we include $Origtime$ in column (5), the magnitude of the estimates changes, but the relationship remains concave. In columns (6) and (7), where we control for continent fixed effects, we learn that the quadratic relationship holds within continents as well.

The last column accounts for the age of states and also for recent developments in the literature postulating that the patterns of human settlement in prehistory may have complex effects on later economic development (Ashraf and Galor, 2013). By

[25]Note that we obtain similar estimates if we use the 1-2000 CE *Statehist* index instead, meaning that the 1950-2000 CE-period is not what is driving the quadratic relationship documented in panel A.

introducing the squared *Origtime* variable, we control for a nonlinear relationship in the time since first human settlement.[26] However, the coefficients of the terms containing *Origtime* are insignificant, while the *State Age* control has a negative and significant, albeit small effect. The introduction of state age diminishes the estimate on *Statehist* squared, indicating that the right extreme of Figure 4 is explained by the length of state existence (the extensive margin of state history), in addition to the overall degree of autonomy or territory considerations (the intensive margin). We note that in panel B, the main estimates when using the old *Statehist* are neither significant, nor similar in terms of signs with the estimates in panel A. This speaks to the added value of the extended *Statehist* data.

Lastly, from Table 4, based on the estimates of our coefficients of interest, we can infer that the optimal predicted level of *Statehist* is reached at 0.356, which is very close to that of the United Kingdom (0.357), and most countries in Western Europe. The effects' magnitudes are not straightforward to assess from the tables. However, some numerical examples may show more clearly how the impact of an increase in *Statehist* depends on the original level of state experience. Take for instance the case of Indonesia, which has 1350 years of state existence and a *Statehist* score of 0.254. If we could hypothetically increase the *Statehist* score by 0.1 (reaching the level of the UK score), the implied approximate effect on per capita GDP in 2000 would be roughly a 20 percent increase, from USD 773 to USD 944 in 2000.[27] The opposite would happen if we were to increase the value of the *Statehist* score by 0.1 for China, which starts off with a value of 0.582: the approximate effect would be a drop in per capita GDP in 2000 by 44.4 percent.

The findings so far are based on the raw *Statehist* data. This means that we only account for the history within the territories of modern-day countries. However, this ignores the state history of other territories from which people migrated in the past to settle in new territories. Population flows after 1500, when the era of colonization began, are instrumental in mapping the impact of historical events to today's economic performance. This is because the ancestors of today's population have evidently brought with them the history, the know-how and the experience with state institutions from their places of origin (Putterman and Weil, 2010; Comin et al, 2010; Ashraf and Galor, 2013).

We therefore use an alternative measure of state history which is obtained by adjusting the *Statehist* index with the migration matrix developed by Putterman and

[26] We explore alternative specifications in Table D2 in the appendix, where we include linear and squared variables such as the timing from transition to agriculture, state age, absolute latitude, migratory distance from Addis Ababa, and predicted genetic diversity (where the latter two are taken from Ashraf and Galor, 2013). Our main coefficients of interest are robust.

[27] The exact calculation based on estimates in column 2 of panel A is $[(7.010 - 2 \cdot 9.842 \cdot 0.254)/10] \cdot 100\% = 20.1\%$

Weil (2010). We then re-estimate equation (5) using this new measure - *the ancestry-adjusted Statehist* - which, for each country, represents the average *Statehist* of its year 2000 population's ancestors, with the weights for each source country being the share of then-living ancestors estimated to have lived on its present-day territory. The results, using two alternative adjustment methods, are displayed in Table 5. In panel A, we use the *Statehist* index in 1500, which we adjust by the migration matrix (as in previous work, but for the first time including full state history before 1 CE). In panel B, we use a composite index obtained by adding the raw 1500 - 2000 *Statehist* to the ancestry - adjusted *Statehist* index at 1500, which is then normalized by the full discounted score for 3500 BCE - 2000 CE. The 1500 - 2000 CE part is added in order to account for the places' histories in the past five centuries.[28]

Table 5

We find that the inverted-u shape relationship between per capita income and the ancestry-adjusted *Statehist* is robust to all specifications and that the coefficients of interest are significant at 1 percent level in all columns in panel A. Moreover, the explanatory power of the model when we introduce only the ancestry-adjusted *Statehist* terms (column 2) is now 20.9 percent vs 5.2 for unadjusted *Statehist.* The results using the measure used in panel B, look reassuringly similar to those in Table 4, panel A. The interpretation of these results is similar, but more nuanced than that where we use the raw data: territories which accumulated limited or extensive state experience, either locally or through an inflow of knowledge from migrant populations, have a lower per capita GDP in 2000 CE than those with an intermediate level of state experience.

How should this inverted-u shape relationship be understood? Although an extensive analysis of the causal mechanisms is beyond the scope of this paper, we can at least offer some reflections. First, our finding appears to be consistent with the fact that while there is indeed a great deal of persistence of early societal advantages, it is also the case that the technological and institutional know-how of societies can slowly diffuse to neighboring societies through migration or trade. These societies with younger states can then pick the best practices of the older societies and potentially avoid some of the pitfalls that might have become a drag for the old civilizations. Hence, while the capacity to organize states is a major asset in early stages of development, it is not necessarily the case that the oldest civilizations have the most efficient economies.

[28]Conceptually, the first part of the component index represents the history non-indigenous populations brought with them to their new homes in 1500 (or after), the second part the political experience they (and indigenous descendants, if any) experienced there since that time. Such a composite gives only a rough accounting for actual experience insofar as many migrants arrived long after 1500, and the timing of migration differs considerably both by receiving and by source country.

Our finding that economic development in countries with old civilizations typically lag behind the countries with an intermediate length of state history like the United Kingdom and the Scandinvian countries, has been discussed in recent work by Olsson and Paik (2013).[29] According to the authors, the main reasons for this “Western reversal of fortune“ since the onset of agriculture were institutional: the old civilizations developed autocratic, hierarchical societies that were not conducive to the emergence of democracy and innovation, which became critical factors for economic growth during the modern era.[30] The more peripheral regions, which were slower to develop state institutions, were furthermore less exposed to raids by roaming armies and to incursions by migrating peoples.

Other factors that have been proposed for explaining the reversal in the Western core include environmental degradation in the Fertile Crescent and in parts of the Mediterranean region. Once agriculture spread out of the Fertile Crescent, the more robust loess soils of northern Europe, combined with a reliance on rain rather than irrigation for cultivation, proved to be an advantage in the long run (Jones, 1981). It has also been suggested that the rise and fall of dominant empires of the Western core followed cycles of expansion, over-extension, and eventually decline, with a gradual shift of power towards the northwest (Kennedy, 1989). Acemoglu et al (2005) show that the emergence of Atlantic trade after 1500 CE had a major impact on the rise of for instance Spain and the United Kindgom.[31]

A similar process can potentially explain comparative development in East Asia. Japan’s less powerful central court and greater perceived vulnerability to potential Western colonizers led it to undertake decisive modernization measures almost a century before China. This development had spillover effects to Korea and Taiwan, all young states in comparison with China.

We leave it for future work to attempt to identify the exact causal mechanism behind the emergence of the inverted-u shaped relationship between state history and economic development.

[29]Olsson and Paik (2013) present preliminary evidence showing that similar reversals appear to have been in place also in East Asia and in Sub-Saharan Africa.

[30]Wittfogel’s (1957) ”hydraulic hypothesis” makes the related argument that the old riverine civilizations were autocratic due to the technological nature of large-scale irrigation. See also Acemoglu and Robinson (2012) for an analysis of how countries with inclusive, democratic institutions eventually tend to dominate countries with extractive, autocratic institutions.

[31]Acemoglu et al (2001) argue that there was also a reversal among former colonies such that relatively less advanced pre-colonial societies had an inflow of European migrants who installed strong institutions that still persist today. Hariri (2012) argues that non-European countries with older states that resisted European colonization had worse economic outcomes in the modern era due to the persistently autocratic nature of their states.

4.3 State history, population growth and technological progress

Having established that state history is related to current per capita income in a non-monotonic fashion, we also inquire whether there is a persistence of this pattern when it comes to historical economic development. Did state experience, measured through the index extended to the BCE period, make a difference for the distribution of wealth in the world in 1500 CE (or 1 CE)?

In order to answer these questions, we estimate equation (5), using various indicators of economic development in 1500 CE as the dependent variable: population density, urbanization and technology adoption (for a detailed description of the variables used see the appendix). In Table 6 we display the results from the model explaining population density in 1500 CE (panel A) and urbanization rate in 1500 CE (panel B). All specifications are analogous to those in tables in section 4.2.[32]

Table 6

In the first column in both panels we see that extended *Statehist* is positively and significantly correlated with past population density. Interestingly, in the second column, where we introduce *Statehist* squared, both coefficients are highly significant, displaying the same quadratic relationship with the left-hand side variable as uncovered in section 4.2. Unconditional on other characteristics of the territories of the modern-day countries in the sample, the positive impact of an increase in state experience on population density diminishes up to a point where it becomes null. Beyond this point, increased experience with state institutions impacts negatively on population density. For historical urbanization rates, we observe the same quadratic pattern displayed by the coefficients of *Statehist* and *Statehist* squared, which are significant at conventional levels for outcomes in 1500 CE. The non-linear pattern is robust to all changes in specification and it holds even within continents. The estimated state history effects are net of the contribution of early transition to agriculture, which was still influencing positively the population density at 1500 CE (note that the link between Neolithic transition and current per capita GDP disappears when the two state history terms were introduced in Tables 4 and 5). This indicates that the same dynamics leading towards or away from an optimal intermediate level of state history were at work even before the colonial era.

Remarkably, the same concave relationship emerges when the dependent variable is the average index of technology adoption (constructed by Comin et al., 2010), particularly in 1500 and 2000 CE (panels A and B in Table 7). Furthermore, using the ancestry-adjusted *Statehist* to explain the differences in average technology adoption

[32] We also fitted multiple regression models where the economic outcomes depend linearly on *Statehist.* The results displayed in Tables D7-D10 in the appendix reveal that these models have slightly lower explanatory power than the models allowing for quadratic *Statehist* effects.

in 2000 yields significant estimates both across and within continents (see online appendix Table D4, panel B). This result is perfectly consistent with the explorations of current income in relation to *Statehist*.[33]We note that the inverted-u relationship is also observed for all outcomes in year 1 CE, but it becomes insignificant when we gradually introduce controls in the regressions (see online appendix Table D3 - panels A and B, and Table D4, panel A).

Table 7

Note that our main estimates of interest remain very similar when we control for *State age* (in column 7 of all tables). This indicates that the intensive margin of the index, given by the autonomy, coherence and territorial extent of the centralized rule, is the main driver of the results. We conclude that places with a long, solid and unchallenged experience of state institutions incurred economic disadvantages relative to ones with intermediary *Statehist* scores. Equally, places at the opposite extreme, with little, often interrupted, or territorially very limited state experience, are also worse-off than those in the middle of the *Statehist* range. Moreover, our evidence suggests that these comparative disadvantages may have already afflicted the territories corresponding to today's countries two thousand years ago. They appear to have already shaped their economic trajectories by 1500, setting the stage for the marked wealth disparities in the world today.

The new pattern uncovered by the extended *Statehist* shows that, beyond a certain point, a longer enduring state history is associated with economic disadvantages. However, caution is recommended against the interpretation of these disadvantages as embedded in long histories, and hence insurmountable. Our view is not that a long uninterrupted state history is "bad" for economic developent per se and as such undesirable. We believe this is a story of adaptability of the state institutions to the ever-changing economic realities, a trait which *Statehist* is merely a proxy for. Moreover, apart from adjusting our state history scores downwards when the territory of what is now a country was incorporated into an externally-based empire, we have not explored systematically what impact (if any) the interactions between states has had. Technology diffusion, for instance, is a manifestation of such complex interactions creating positive externalities (Comin and Hobijn, 2010), where the risk associated with the initial development of a new technology is reduced for states that adopt the already tested technology from earlier experimenters.

Although the analysis of this section has demonstrated a clear association between a history of state presence and the overall level of technological development,

[33]As a robustness check, we have redone the estimations using the overall technology adoption index excluding the agriculture components; the results are very similar. See Table D6 in the appendix.

we have thus far presented no evidence that states contributed directly to any specific area of technical progress. Speculation about the roles of states in fostering specific technological developments is common, however. In our final empirical exercise, we explore the statistical relationship of states to a fundamental technology often surmised to be related to their emergence: writing. It proved possible to assemble estimates of the year in which writing first emerged in 148 of our sample countries, and we make use of this data for the exercises shown in Table 8.[34]

Table 8

Before turning to the regressions, some descriptive comments are of interest. Although development of writing is associated in the conventional wisdom with the record-keeping needs of early states, it would be oversimplifying, we think, to suppose that state emergence was the only or crucial impetus for writing. Of the 143 countries in our sample, emergence of writing follows that of a state in 80 cases, leaving 63 cases in which writing preceded the state. Many of the latter may reflect instances in which writing happened to diffuse to a given peripheral location somewhat faster than state organization, with both originating in a common core area. So the order of appearance for cases in which both the state and writing emerged with little external influence is of special interest. Of the six most likely cases of this kind—Mesopotamia (Iraq), Egypt, China, the Indus Valley (India or Pakistan), Mesoamerica (Mexico) and the Andes (Peru)—our data show state formation preceding writing's emergence in four cases, the exceptions being Egypt and Mexico.[35] In the four cases in which the state came first, writing appeared at least a half millennium and on average of 1.21 millennia after the state, while in both cases in which writing emerged first, the timing difference is only 0.2 millennia. The pattern of shorter time lag when writing precedes the state also holds for the full sample, although with less pronounced asymmetry: a state precedes writing by an unweighted average of 517 years in the 80 countries in which the state appears first, whereas writing precedes the state by an average of 397 years in the 63 countries having the reverse order.

There are also differences when we divide the sample into internally-originated and externally-originated states. The former are significantly more likely to have writing appear after the state does (which occurs in 69.4 percent of cases) than are

[34]Comin et al.'s data include presence of writing in 1000 BCE, 1 CE and 1500 CE, and the main source on which they rely, Peregrine (2001) provide a few additional observations by millennium. However, we found an online compilation of estimates for first emergence of writing, checked its sources, and assembled those that appeared to reflect consensus views. Details can be found in the appendix.

[35]Another three cases are also relatively early and possibly but less definitively independent: Iran, Guatemala, and Bolivia. In all three, states emerged first, then writing.

the former (which have states before writing in only 42.2 percent of cases).[36]

In panel A of Table 8 we report the results using the entire sample for which we have data on extended state history and the emergence of writing. The estimate of the relationship between state onset and writing is significant at the 1 percent level and its value, 0.701, implies that states that emerged 1000 years earlier, adopted writing 701 years earlier. By comparison, one millennium of earlier transition to agriculture, also quite significantly correlated with writing's emergence, is associated with only 387 years of advance in adopting writing (column 2) although both appearance of agriculture and emergence of a state are significant predictors of writing when entered jointly in column (3).

Moreover, when geographical controls are added to the model, the relationship between state onset and the emergence of writing remains significant, while the impact of early transition to agriculture becomes insignificant. This suggests that it may be something about the presence of a state per se that encouraged the development or adoption of writing. Although we do not attempt to disentangle possible channels here, the explanations may lie in the administrative requirements of the states, those of closely associated early priesthoods, or simply the greater division of labor aided by macro-level political order.

Since writing in some states may have been diffused into adjacent territories not yet politically organized, these results may also reflect a catalytic effect of writing for the emergence of centralized of power. Therefore, in panel B, we also present the result from the subsample of states where the emergence of political institutions preceded the emergence of writing. All results are robust to the inclusion of geographical controls and continent indicators, reinforcing the idea that political institutions played a part in the invention and adoption of writing. This is consistent with the general view that writing was a means of facilitating taxation and record-keeping, features that are central to the state apparatus.

5 Conclusions

We coded and assembled a comprehensive index of state history from state emergence (which often occurred before the Common Era) to 2000 CE for a sample of 159 countries, building on the previously constructed *State antiquity index* of Bockstette et al (2002). Grounding our definition of state in the anthropological and political literatures, we outlined the guiding principles followed in mapping historical information into the three components that make up the state history index: 1) Existence of a state, 2) whether the state is home-based or imposed from without;

[36] A t-test finds the difference between the two types of state in probability that states precede writing to be significant at the 0.1% level.

and 3) territorial coverage of the state relative to the boundaries of the present-day country. Tracing these characteristics of the states back to the emergence of the state in the territories defined by modern country borders, we obtained three overarching measures of state presence and evolution: 1) a cumulative *Statehist* index (as in Bockstette et al.), 2) *State age* (time since state emergence) and 3) a contemporaneous *State index* capturing the level of state presence at different points in time. A particular advantage with our data stems from the availability of state history information at various levels of spatial and temporal aggregation. The inherent features of the process of compiling state scores render our data particularly versatile for a large variety of comparative analyses.

After presenting the comprehensive state history data, we revealed a series of patterns of state evolution, reversal and catching-up. We employed the data to explore the determinants of state formation. In particular, we demonstrated empirically the strong catalytic relationship between the Neolithic transition to agriculture and state emergence, as well as a strong association between the timing of that transition and cumulative state history.

In our regression analysis, we explored the relationships between state history and income level, early historical proxies for income (population density and urbanization), and technology. Whereas previous estimates using data for the period 1 to 1950 CE only had suggested a linear relationship, we showed that cumulative state history from earliest emergence to 2000 CE has a concave, rather than linear relationship with current income and with our demographic proxies for income, especially in 1500, and we confirmed that inclusion of the BCE period is crucial to this result. Contrary to what was previously thought about the positive linear relationship beween the two, we showed that they are related through an inverted-u function. Countries with scores at the extremes of the *Statehist* range are worse off in terms of both current and historical economic development than countries with intermediate values of *Statehist*. The optimal level of state history as defined here is estimated to be that of modern-day United Kingdom. This points to the role of adaptable institutions and learning. These features may be lacking in countries with the longest histories of state presence, but may prevail in the younger yet mature states, such as those in Western Europe.

Finally, after finding a similar relationship between states and a broad index of technologies as between states and income, we carried out what may be the first global-scale investigation of the temporal connection between state formation and emergence or adoption of writing. While writing was found to precede state presence in almost half of the 143 countries for which we could compile data, we found that writing followed state formation in the large majority of cases in which states emerged from an internal process rather than being imposed from without.

We also found a more statistically robust and significant relationship of writing with state emergence than with agricultural transition, supporting the notion of a special role of states in the adoption and transmission of writing systems.

References

[1] Acemoglu, D., Johnson, S., and Robinson, J. (2001). "The colonial origins of comparative development: An empirical investigation", American Economic Review 91 (5), pp. 1369–1401.

[2] Acemoglu, D., Johnson, S., and Robinson, J. (2002). "Reversal of fortunes: Geography and institutions in the makingof the modern world income distribution", Quarterly Journal of Economics 117 (4), pp. 1231–1294.

[3] Acemoglu, D., Johnson, S. and Robinson, J. (2005). "Institutions As the Fundamental Cause of Long Run Economic Growth", in Handbook of Economic Growth, Aghion, P. and Durlauf, S. eds., North-Holland, Amsterdam.

[4] Acemoglu, D. and Robinson, J. (2012)",Why Nations Fail: The Origins of Power, Prosperity and Poverty", Crown Business, New York.

[5] Ahlerup, P. & Olsson, O. (2012). "The roots of ethnic diversity", Journal of Economic Growth, Springer, vol. 17(2), pp. 71-102, June.

[6] Ang, J. (2012). "What Drives the Formation and Persistent Development of Territorial States since 1 AD?", MPRA Paper 42357, University Library of Munich, Germany.

[7] Ang, J. (2013a). "Are Modern Financial Systems Shaped by State Antiquity?", Journal of Banking and Finance 37 (11), pp. 4038 – 4058.

[8] Ang, J. (2013b). "Institutions and the Long-run Impact of Early Development", Journal of Development Economics 108m, pp. 1 – 18.

[9] Ashraf, Q. and Galor, O. (2013). "The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development", American Economic Review, 103(1), pp. 1-46.

[10] Bellwood P. S. (2005). "The first farmers: The origins of agricultural societies". Malden, MA: Blackwell.

[11] Besley, T. and T. Persson (2009). "The Origins of State Capacity: Property Rights, Taxation and Politics“ American Economic Review 99(4), pp. 1218-1244.

[12] Bockstette, V. & Chanda, A. and Putterman, L. (2002). "States and Markets: The Advantage of an Early Start", Journal of Economic Growth, Springer, vol. 7(4), pp. 347-69, December.

[13] Carneiro, R.L. (1970). A theory of the origin of the state. Science 169, pp. 733-38.

[14] Carneiro, R.L. (1981). "The chiefdom: precursor of the state". In "The Transition to Statehood in the New World", ed. GD Jones, RR Kautz, Cambridge: Cambridge Univ. Press, pp. 37-79.

[15] Chanda, A. and Putterman, L. (2005). "State Effectiveness, Economic Growth, and the Age of States", in Matthew Lange and Dietrich Rueschemeyer, eds., "States and Development: Historical Antecedents of Stagnation and Advance", Basingstoke, England, Palgrave MacMillan, pp. 69-91.

[16] Chanda, A. and Putterman, L. (2007). “Early Starts, Reversals and Catch-up in the Process of Economic Development”, Scandinavian Journal of Economics, Wiley Blackwell, vol. 109(2), pp. 387-413.

[17] Chanda, A., Cook, A. and Putterman, L. (2014). “Persistence of Fortune: Accounting for Population Movements, There was No Post-Columbian Reversal”, American Economic Journal – Macroeconomics (in press, August issue).

[18] Cinyabuguma, M. and Putterman, L. (2011). “Sub-Saharan Growth Surprises: Being Heterogeneous, Inland and Close to the Equator Does not Slow Growth within Africa”, Journal of African Economies 20(2), pp. 217 – 262.

[19] Comin, D., Easterly, W. and Gong, E. (2010). “Was the Wealth of Nations Determined in 1000 BC?”, American Economic Journal: Macroeconomics, American Economic Association, vol. 2(3), pp. 65-97, July.

[20] Comin, D., and Hobijn, B. (2010). “An Exploration of Technology Diffusion?”, American Economic Review, American Economic Association, vol. 100(5), pp. 2031-59, December.

[21] Daniele, V. (2013). “Does the Intelligence of Populations Determine the Wealth of Nations?”, Journal of Socio-Economics 46, pp. 27 – 37.

[22] Depetris-Chauvin, E. (2014) “State history and contemporary conflict: Evidence from Sub-Saharan Africa“ unpublished working paper, Brown University.

[23] Diamond, J.M. (1997). “Guns Germs and Steel: The Fate of Human Societies”, W.W. Norton & Co., New York NY.

[24] Earle, T.K. (1987). “Chiefdoms in archaeological and ethnohistorical perspective”. Annu. Rev. Anthropol. 16, pp. 279-308.

[25] Ertan, A., Putterman, L. and Fiszbein, M. (2012). “Determinants and Economic Consequences of Colonization: A Global Analysis”, Brown University Department of Economics Working Paper, August, 2012.

[26] Fagan, B. (2002). The Little Ice Age, Basic Books.

[27] Flannery, K.V. (1995). “Prehistoric social evolution”. In “Research Frontiers in Anthropology”, ed. M Ember, Englewood Cliffs, NJ: Prentice-Hall, pp. 1-26.

[28] Galor, O. and Moav, O. (2007). “The Neolithic Revolution and Contemporary Variations in Life Expectancy”, Working Papers 2007-14, Brown University, Department of Economics.

[29] Hariri, J.G. (2012). “The Autocratic Legacy of Early Statehood ’ American Political Science Review“ , vol. 106(3), pp. 471-494.

[30] Hibbs, D. A., and Olsson, O. (2004). “Geography, biogeography, and why some countries are rich and others are poor”, Proceedings of the National Academy of Sciences, 101, pp. 3715-3720.

[31] Iliev, P. and Putterman, L. (2007). “Social Capability, History and the Economies of Communist and Post-Communist States”, Studies in Comparative International Development 42 (1), pp. 36-66.

[32] Johnson, A. and Earle, T. (2000). “The evolution of human societies: from foraging group to agrarian state”, Stanford Unviersity Press, 2nd edition, December.

[33] Jones, E.L. (1981). “The European Miracle: Environments, Economies and Geopolitics in the History of Europe and Asia”, Cambridge: Cambridge University Press.

[34] Jones, G. D., and Kautz, R.R. (1981). “Issues in the Study of New World State Formation”. In “The Transition to Statehood in the New World”. Grant Jones and Robert Kautz, eds.,Cambridge: Cambridge University Press, Pp. 3-34.

[35] Kennedy, P. (1989). “The Rise and Fall of the Great Powers”, Vintage Books, New York.

[36] Marcus, J. (1992). “Dynamic cycles of Mesoamerican states: political fluctuations in Mesoamerica”. Nati. Geogr. Res. Explor. 8:, pp. 392-411.

[37] Morris, I. (2010). “Why the West Rules—For Now: The Patterns of History, and What They Reveal About the Future”,London: Profile Books.

[38] North, D. (1990) “Institutions, Institutional Change and Economic Performance”, Cambridge: Cambridge University Press.

[39] Olsson, O. and Hibbs, D. (2005). “Biogeography and Long -run Economic Development”, European Economic Review, Vol. 49, No. 4, pp. 909-938.

[40] Olsson, O. and C. Paik (2013). “A Western Reversal Since the Neolithic: The Long-Run Impact of Early Agriculture“ University of Gothenburg, mimeo.

[41] Peregrine, P.N., Ember, C.R. and Ember M. (2007). “Modeling State Origins Using Cross-Cultural Data”, Cross-Cultural Research, Vol. 41, No. 1, pp. 75-86.

[42] Peregrine, P.N. and Ember, M., eds. (2001). “Encyclopedia of Prehistory”, New York: Kluwer Academic/Plenum Publishers.

[43] Petersen, M.B., and Skaaning. S.E. (2010). “Ultimate Causes of State Formation: The Significance of Biogeography, Diffusion, and Neolithic Revolution.” Historical Social Research, vol. 55 (3), pp. 200–26.

[44] Putterman, L. (2008). “Agriculture, Diffusion and Development: Ripple Effects of the Neolithic Revolution”, Economica, vol. 75(300), pp. 729-748, November.

[45] Putterman, L. with C.A Trainor (2006) ”Agricultural Transition Year - Country Data Set” http://www.econ.brown.edu/fac/louis_putterman/.

[46] Putterman, L. and Weil, D. N. (2010). “Post-1500 Population Flows and the Long-Run Determinants of Economic Growth and Inequality”, The Quarterly Journal of Economics, MIT Press, vol. 125(4), pp. 1627-1682, November.

[47] Richerson,P.J., Mulder, M.B. and Vila, B.J (1996). “Principles of Human Ecology”. Pearson Custom Publishing.

[48] Service, E. R. (1962). “Primitive social organization; an evolutionary perspective”, Random House, New York.

[49] Spencer, C.S. (1990). “On the tempo and mode of state formation: Neoevolutionism reconsidered”. J. Anthropol. Archaeol. 9, pp. 1-30.

[50] Spencer, C.S and Redmond, E.M. (2004). “Primary state formation in Mesoamerica”. Annu. Rev. Anthropol. 33, pp. 173–99.

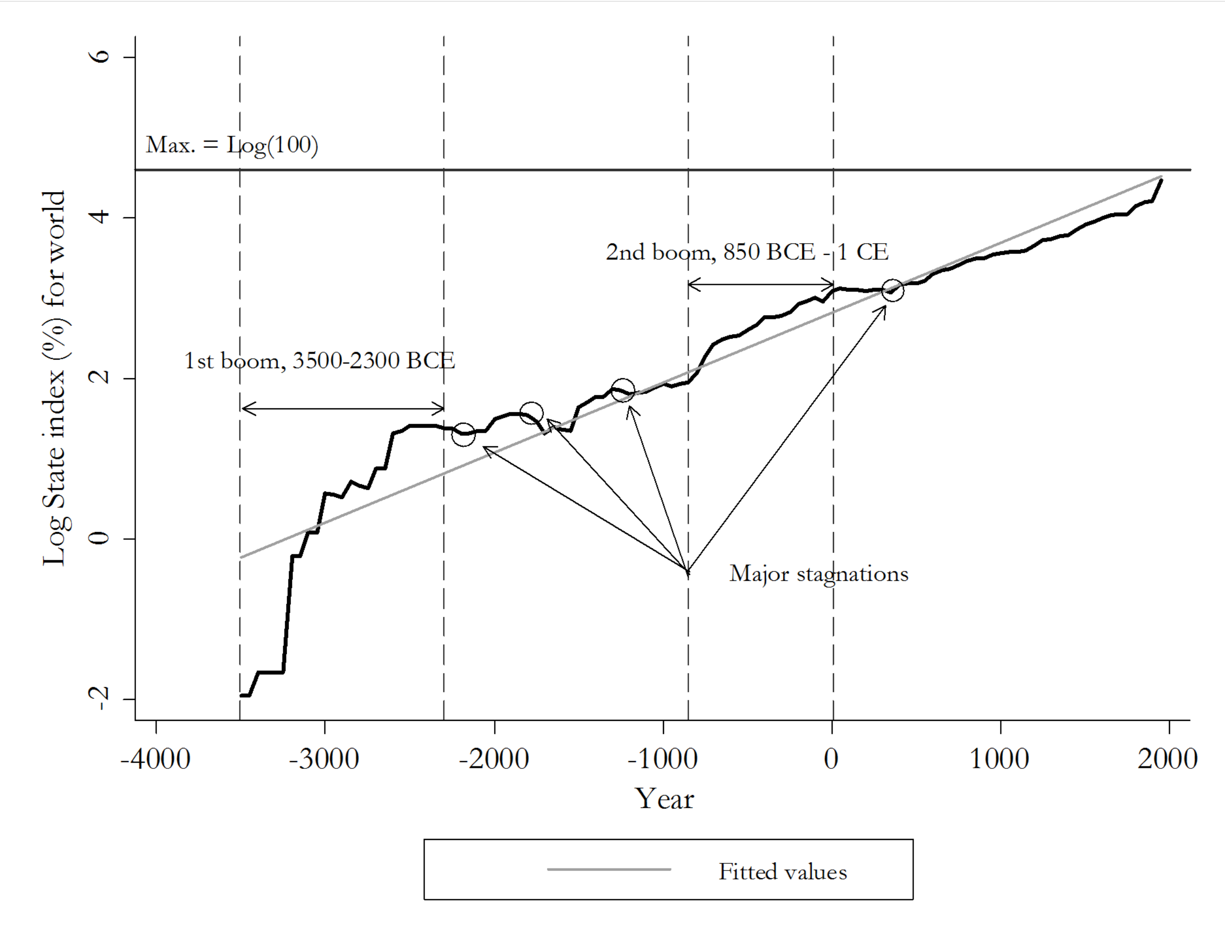
[51] Tilly, C. (1990). “Coercion, Capital and European States”, AD 990-1990, Oxford.

[52] Weber, M. (1919). “Politics as a vocation”.

[53] Wittfogel, K.A. (1957). "Oriental Despotism; a Comparative Study of Total Power", Yale University Press.

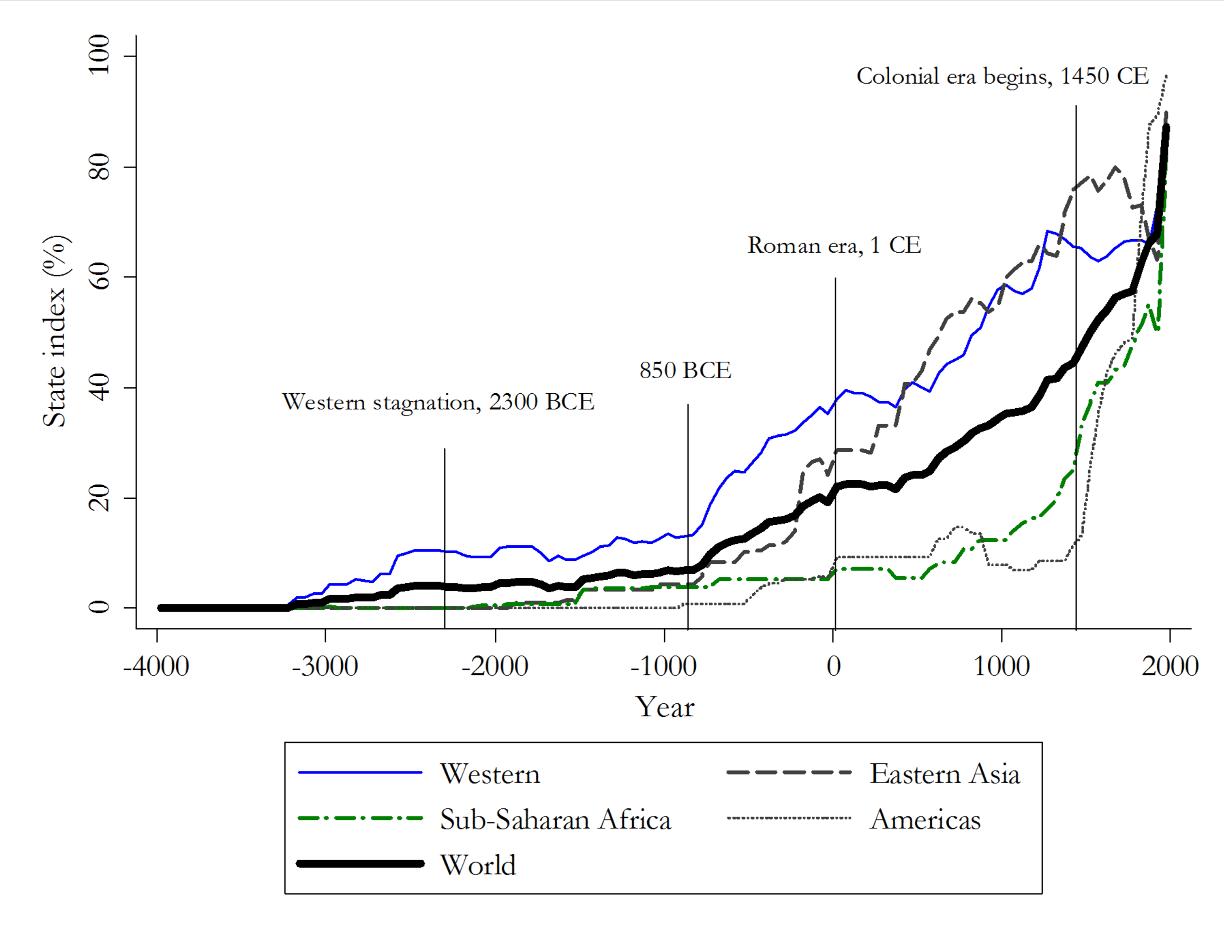
FIGURES AND TABLES

Figure 1: Emergence of states in the world 3500 BCE-2000 CE



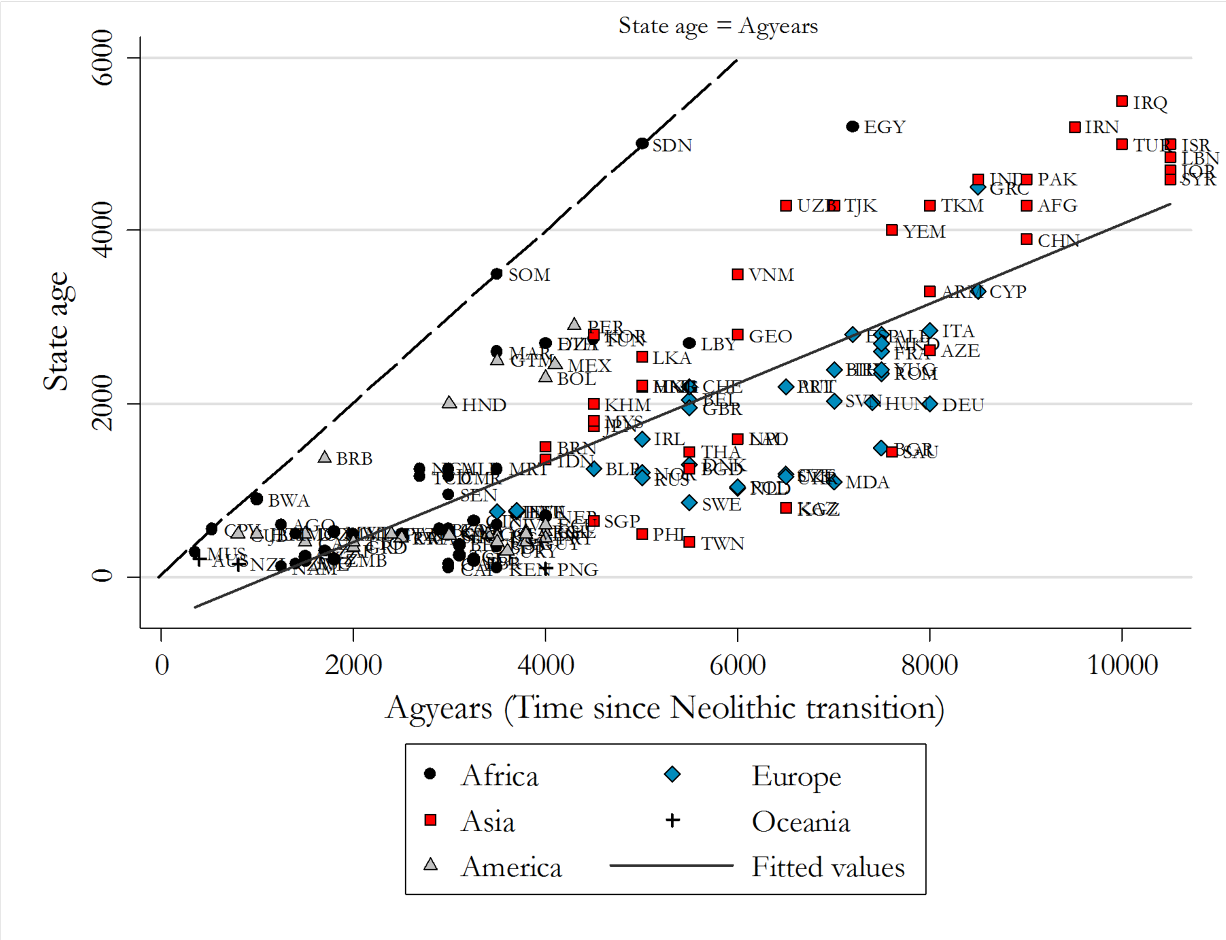
Note: The graph shows the logged value of the aggregate State index for 159 countries identified during 110 50-year intervals between 3500 BCE 2000 CE. The value 100 is equivalent to all 159 countries in our sample being full states, as defined in the text. On the horizontal axis, negative values imply years BCE whereas positive values show the CE-period. A linear fitted regression line has been included. The State index is calculated as described in text.

Figure 2: Emergence of states in four agricultural core areas and in the world as a whole 3500 BCE- 2000 CE



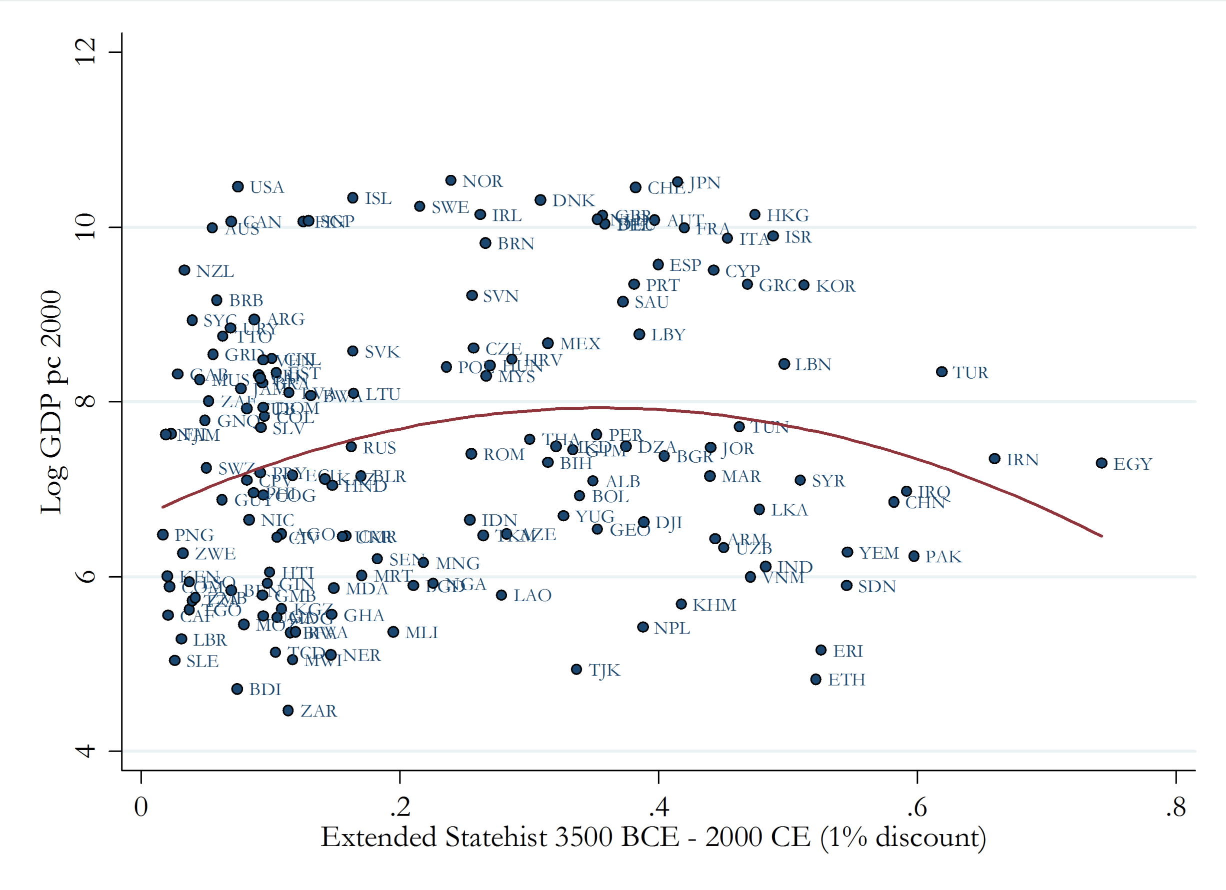
Note: The figure shows the development of the aggregated State index in the Western agricultural zone (including 62 current countries in Europe, North Africa, the Middle East, as well as Afghanistan, Armenia, Azerbaidjan, Georgia, India, Iran, Kazachstan, Kyrgyzstan, Pakistan, Russia, Tajikistan, Turkmenistan, Uzbekistan), Eastern Asia (20 countries), Americas (including 27 countries in North and South America and in the Caribbean), and Sub-Saharan Africa (47 countries). Oceania (including 3 countries) is omitted. It also shows the aggregate index for the 159 countries in the world as a whole (solid black line). On the horizontal axis, negative values imply years BCE whereas positive values show the CE-period. Particular years with trend breaks are marked.

Figure 3: Relationship between State age and Time since Neolithic Transition



Note: The figure shows the simple relationship between State age and Agyears including a fitted regression line and a 45 degree-line where State age = Agyears. Observations are distinguished on the basis of the continent to which they belong and their 3-letter country isocodes.

Figure 4: Non-linear relationship between Log GDP per capita in 2000 and Statehist index



Note: The figure shows a fitted quadratic regression line corresponding to the estimates in Table 4, Panel A, column2, with 154 country observations distinguished by 3-letter country isocodes.

Table 1: Summary Statistics

Panel A State history indicators	N	Mean	SD	Min	Max
Statehist	159	0.234	0.172	0.017	0.743
Statehist 1 - 1950 CE	159	0.386	0.261	0.012	0.978
Ancestry - Adjusted Statehist in 2000 CE	154	0.252	0.189	0.017	0.811
Ancestry - Adjusted Statehist in 1500 CE	154	0.218	0.167	0.000	0.747
State Age (millenia)	159	1.639	1.430	0.100	5.500
Internally- originated	159	0.490	0.501	0	1
Panel B Outcome Variables					
Time since Writing Emerged (millenia)	149	1.649	1.186	0.100	5.400
(Log) Population Density in 1 CE	135	-0.112	1.530	-4.510	3.170
(Log) Population Density in 1500 CE	154	0.905	1.461	-3.817	3.842
Urbanization Rate in 1 CE	128	2.641	0.624	1.000	3.000
Urbanization Rate in 1500 CE	83	7.278	5.134	0.000	28.000
Average Technology Adoption in 1 CE	128	0.739	0.274	0.000	1.000
Average Technology Adoption in 1500 CE	112	0.487	0.317	0.000	1.000
Average Technology Adoption in 2000 CE	130	0.451	0.198	0.174	1.012
(Log) GDP pc in 2000	154	7.488	1.606	4.463	10.531
Panel C Covariates					
Agyears (millenia)	151	4.717	2.442	0.362	10.500
Origtime (millenia)	158	58.917	49.958	0.200	160.000
Absolute centroid latitude	159	26.368	17.704	0.422	67.469
Landlocked	134	0.224	0.418	0.000	1.000
Distance to coast and rivers	149	374.333	457.408	7.952	2385.58
Mean Elevation	149	637.715	551.281	9.167	3185.920
Land Suitability	145	0.378	0.248	0.000	0.960
Percentage Arable Land	156	15.852	14.001	0.040	62.100
Temperature	158	18.226	8.350	-7.929	28.639
Precipitation	158	92.959	61.700	2.911	259.952
Malaria (percentage population at risk)	151	0.316	0.426	0.000	1.000

Note: The table summarizes all variables used in the analysis, as follows: 1) Panel A describes the State history variables created by us; 2) Panel B outlines some historical and economic variables which are used as dependent variables in the regression analysis. The time since writing emerged is a novel variable that we compiled, measured in thousands of years. The data for historical population density is based on population data from McEvedy and Jones(1978) and land data from World Banks World Development Indicators. The data for urbanization rate in 1 CE is taken from Comin, Easterly and Gong (2010) and is based on Peregrine (2003). The data for urbanization rate in 1500 CE is that reported by Acemoglu, Johnson and Robinson (2005). The Average Technology Adoption indices in 1 CE, 1500 CE and 2000 CE are constructed by Comin, Easterly and Gong (2010). Per capita GDP is expressed in current US dollars, as provided by the World Ban; 3) Panel C details on the covariates included in the regressions. Agyears was assembled by Putterman with Trainor (2006) and it records the number of millennia elapsed in 2000 C.E. since the Neolithic transition took place. Origtime was coded by Ahlerup and Olsson (2012) and it represents the time since initial uninterrupted settlement by modern humans (before 2000 CE). The geographic and climatic controls are retrieved from various sources. The variables' construction is detailed in Appendix A.

Table 2: Time since the Neolithic transition, state age and state history; overall sample

	State Age			Statehist		
	(1)	(2)	(3)	(4)	(5)	(6)
Agyears	0.471***	0.469***	0.430***	0.053***	0.050***	0.040***
	(0.025)	(0.027)	(0.047)	(0.003)	(0.004)	(0.008)
Origtime		0.001	-0.002		0.000	-0.001
		(0.002)	(0.003)		(0.000)	(0.001)
Absolute latitude		0.023	0.013		0.004*	0.002
		(0.017)	(0.019)		(0.002)	(0.002)
Landlocked		-0.486***	-0.464**		-0.052**	-0.053**
		(0.173)	(0.183)		(0.025)	(0.027)
Distance		0.000	0.000		-0.000	-0.000
to coast and river		(0.000)	(0.000)		(0.000)	(0.000)
Mean Elevation		0.001***	0.000**		0.000***	0.000*
		(0.000)	(0.000)		(0.000)	(0.000)
Land suitability		-0.065	0.193		-0.032	0.018
		(0.445)	(0.501)		(0.059)	(0.065)
Percentage		-0.009	-0.010		-0.000	-0.001
arable land		(0.007)	(0.007)		(0.001)	(0.001)
Temperature		0.093***	0.072**		0.011***	0.008*
		(0.031)	(0.033)		(0.004)	(0.005)
Precipitation		-0.004**	-0.005**		-0.000	-0.000*
		(0.002)	(0.002)		(0.000)	(0.000)
Malaria (percentage		-0.480*	-0.559*		-0.033	-0.063
population at risk)		(0.266)	(0.299)		(0.043)	(0.047)
Constant	-0.555***	-2.504*	-1.528	-0.013	-0.295*	-0.131
	(0.116)	(1.302)	(1.480)	(0.014)	(0.166)	(0.190)
Continent FE	No	No	Yes	No	No	Yes
Observations	151	128	128	151	128	128
R-squared	0.647	0.781	0.791	0.578	0.672	0.697

Note: The dependent variable is *State age* in columns (1)-(3) and *Statehist* in columns (4)(6). Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Time since the Neolithic transition and State age in internally vs. externally originated states

	State Age					
	Internally originated	Externally originated	Internally originated	Externally originated	Internally originated	Externally originated
	(1)	(2)	(3)	(4)	(5)	(6)
Agyears	0.530***	0.307***	0.472***	0.347***	0.444***	0.397***
	(0.026)	(0.029)	(0.036)	(0.057)	(0.061)	(0.084)
Origtime			-0.000	-0.006*	-0.000	-0.011**
			(0.003)	(0.003)	(0.005)	(0.004)
Absolute latitude			0.014	0.048*	0.015	0.054
			(0.022)	(0.025)	(0.023)	(0.038)
Landlocked			-0.438**	-1.107***	-0.369	-1.034**
			(0.200)	(0.380)	(0.223)	(0.396)
Distance			0.001***	0.000	0.001*	0.001
to coast and river			(0.000)	(0.000)	(0.000)	(0.000)
Mean Elevation			0.000*	0.000	0.000	0.001
			(0.000)	(0.000)	(0.000)	(0.000)
Land suitability			0.386	-0.331	0.422	-0.265
			(0.495)	(0.442)	(0.588)	(0.463)
Percentage			-0.013	0.006	-0.012	0.010
arable land			(0.009)	(0.009)	(0.009)	(0.012)
Temperature			0.102***	0.128**	0.082**	0.140**
			(0.037)	(0.048)	(0.039)	(0.067)
Precipitation			-0.002	-0.006**	-0.003	-0.005
			(0.003)	(0.003)	(0.004)	(0.003)
Malaria (percentage			-1.259***	1.009*	-1.137**	1.207
population at risk)			(0.381)	(0.591)	(0.497)	(0.730)
Constant	-0.667***	-0.115	-2.225	-3.375*	-1.960	-3.914
	(0.153)	(0.117)	(1.777)	(1.931)	(1.809)	(2.879)
P-value	0.000		0.039		0.600	
Continent FE	No	No	No	No	Yes	Yes
Observations	78	73	71	57	71	57
R-squared	0.761	0.392	0.870	0.709	0.876	0.742

Note: The dependent variable is *State age.* p-values reported at the bottom of the table represent the p-values from the test of the null hypothesis of equality of coefficients of *agyears* within the three pairs of specifications. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Statehist vs. Statehist 1-1950 CE and (Log) GDP pc 2000. Nonlinear relationship

Panel A	Log GDP pc 2000						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Statehist	1.326*	7.010***		7.337***	3.869**	4.530**	6.790***
	(0.723)	(2.291)		(2.658)	(1.921)	(2.057)	(2.496)
Statehist squared		-9.842***		-9.832***	-4.718	-4.970*	-4.657*
		(3.529)		(3.549)	(2.854)	(2.793)	(2.776)
Agyears			0.105**	0.004	-0.071	-0.087	0.010
			(0.048)	(0.079)	(0.063)	(0.079)	(0.081)
Origtime					0.002	0.008**	0.010
					(0.003)	(0.004)	(0.013)
Origtime squared							-0.000
							(0.000)
State Age							-0.460**
							(0.183)
Observations	154	154	147	147	125	125	125
R-squared	0.020	0.052	0.026	0.064	0.702	0.719	0.734
Panel B	Log GDP pc 2000						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Statehist 1-1950 CE	1.277**	1.940		2.200	0.066	0.251	1.267
	(0.531)	(2.049)		(2.278)	(1.441)	(1.597)	(1.667)
Statehist 1-1950 CE squared		-0.783		-0.748	0.942	0.962	0.453
		(2.518)		(2.625)	(1.608)	(1.811)	(1.776)
Agyears			0.105**	-0.011	-0.069	-0.080	0.012
			(0.048)	(0.068)	(0.055)	(0.072)	(0.081)
Origtime					0.001	0.007*	0.011
					(0.003)	(0.004)	(0.013)
Origtime squared							-0.000
							(0.000)
State Age							-0.267**
							(0.127)
Observations	154	154	147	147	125	125	125
R-squared	0.043	0.044	0.026	0.058	0.704	0.722	0.730
Controls	No	No	No	No	Yes	Yes	Yes
Continent FE	No	No	No	No	No	Yes	Yes

Note: The dependent variable is Log per capita GDP in 2000. In panel A the main independent variables are extended *Statehist index* linear and squared. In panel B the main independent variables are the *Statehist index 1-1950 CE*, linear and squared. The list of controls includes: absolute latitude, an indicator of whether the modern-day country is landlocked, distance to coast and rivers, mean elevation, land suitability, percentage arable land, temperature, precipitation, percentage population at risk of contracting malaria. Robust standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1

Table 5: Ancestry-Adjusted Statehist and (Log) GDP pc 2000. Nonlinear relationship

Panel A	Log GDP pc 2000						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ancestry Adjusted Statehist in 1500	2.778***	12.213***		13.110***	6.068***	5.347***	6.041***
	(0.794)	(2.263)		(2.100)	(1.574)	(1.647)	(1.761)
Ancestry Adjusted Statehist in 1500 squared		-18.218***		-18.636***	-8.984***	-7.519***	-6.794***
		(4.326)		(4.076)	(2.176)	(2.324)	(2.317)
Agyears			0.105**	-0.025	-0.056	-0.075	-0.027
			(0.048)	(0.059)	(0.056)	(0.078)	(0.082)
Origtime					0.003	0.006*	0.008
					(0.003)	(0.004)	(0.013)
Origtime squared							-0.000
							(0.000)
State Age							-0.233
							(0.146)
Observations	149	149	147	144	125	125	125
R-squared	0.083	0.209	0.026	0.243	0.722	0.727	0.733

Panel B	Log GDP pc 2000						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ancestry Adjusted Statehist in 2000	1.389**	7.074***		6.661***	3.514**	4.123**	6.268***
	(0.670)	(2.113)		(2.426)	(1.727)	(1.849)	(2.286)
Ancestry Adjusted Statehist in 2000 squared		-9.085***		-8.378***	-4.034*	-4.250*	-4.033*
		(3.021)		(3.000)	(2.364)	(2.308)	(2.289)
Agyears			0.105**	0.021	-0.069	-0.085	0.010
			(0.048)	(0.080)	(0.063)	(0.079)	(0.081)
Origtime					0.002	0.008**	0.010
					(0.003)	(0.004)	(0.013)
Origtime squared							-0.000
							(0.000)
State Age							-0.463**
							(0.186)
Observations	149	149	147	144	125	125	125
R-squared	0.027	0.066	0.026	0.071	0.702	0.719	0.734
Controls	No	No	No	No	Yes	Yes	Yes
Continent FE	No	No	No	No	No	Yes	Yes

Note: The dependent variable is Log per capita GDP in 2000. In panel A the main independent variables are the *ancestry-adjusted extended statehist index* between 3500 BCE and 1500 CE, linear and squared. In panel B the main independent variables are the *composite ancestry-adjusted Statehist index* (where the discounted ancestry - adjusted scores between 3500 BCE and 1500 CE are added to the raw discounted scores between 1500 2000 CE, and the final score is normalized by the sum of discounted full scores between 3500 BCE and 2000 CE), linear and squared. The list of controls includes: absolute latitude, an indicator of whether the present-day country is landlocked, distance to coast and rivers, mean elevation, land suitability, percentage arable land, temperature, precipitation, percentage population at risk of contracting malaria. Robust standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1

Table 6: State history, Log Population Density and Urbanization in 1500 CE

Panel A	Log Population Density in 1500 CE						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Statehist in 1500 CE	3.883***	9.559***		6.184***	7.473***	6.129***	11.077***
	(0.670)	(1.666)		(2.119)	(1.802)	(1.709)	(3.426)
Statehist in 1500 CE squared		-12.324***		-9.893***	-7.339**	-4.894**	-7.326**
		(3.098)		(3.498)	(3.169)	(2.253)	(2.905)
Agyears in 1500 CE			0.315***	0.211***	0.157**	0.131*	0.217***
			(0.042)	(0.067)	(0.065)	(0.067)	(0.068)
Origtime in 1500 CE					0.005**	-0.003	-0.020
					(0.003)	(0.004)	(0.014)
Origtime in 1500 CE squared							0.000
							(0.000)
State Age in 1500 CE							-0.509**
							(0.214)
Observations	154	154	147	147	128	128	128
R-squared	0.184	0.254	0.269	0.314	0.709	0.767	0.786
Panel B	Urbanization in 1500 CE						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Statehist in 1500 CE	16.678***	29.429***		35.364***	48.134***	41.542***	69.670***
	(2.384)	(7.323)		(8.662)	(12.193)	(12.707)	(22.835)
Statehist in 1500 CE squared		-25.531**		-29.862**	-43.924***	-35.621**	-49.359***
		(12.514)		(12.662)	(15.010)	(14.076)	(17.606)
Agyears in 1500 CE			0.761***	-0.382	-0.206	-0.323	0.152
			(0.177)	(0.244)	(0.325)	(0.417)	(0.481)
Origtime in 1500 CE					-0.076**	-0.082*	-0.216
					(0.037)	(0.042)	(0.146)
Origtime in 1500 CE squared							0.002
							(0.002)
State Age in 1500 CE							-2.831
							(1.750)
Observations	83	83	83	83	76	76	76
R-squared	0.278	0.311	0.111	0.324	0.459	0.498	0.532
Controls	No	No	No	No	Yes	Yes	Yes
Continent FE	No	No	No	No	No	Yes	Yes

Note: In panel A, the dependent variable is log population density in 1500 CE and the main independent variables are the *extended statehist index* between 3500 BCE and 1500 CE, linear and squared. In panel B, the dependent variable is the urbanization rate in 1500 CE and the main independent variables are the *extended statehist index* between 3500 BCE and 1500 CE, linear and squared. The data for historical population density is based on population data from McEvedy and Jones(1978) and land data from *World Banks World Development Indicators*. The data for urbanization rate at 1500 CE is that reported by Acemoglu, Johnson and Robinson (2005), defined as the percentage of a countrys total population residing in urban areas (each with a city population size of at least 5,000). The list of controls includes: absolute latitude, an indicator of whether the present-day country is landlocked, distance to coast and rivers, mean elevation, land suitability, percentage arable land, temperature, precipitation, percentage population at risk of contracting malaria. Robust standard errors in parentheses.*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 7: State history and average technology adoption in 1500 CE and 2000 CE

Panel A	Technology Adoption in 1500 CE						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Statehist in 1500 CE	1.227***	2.841***		1.695***	1.727***	1.068***	1.782***
	(0.157)	(0.350)		(0.444)	(0.338)	(0.229)	(0.416)
Statehist in 1500 CE squared		-3.359***		-2.587***	-1.855***	-0.743**	-0.943***
		(0.738)		(0.887)	(0.624)	(0.346)	(0.319)
Agyears in 1500 CE			0.104***	0.073***	0.038***	0.004	0.012
			(0.008)	(0.014)	(0.013)	(0.010)	(0.011)
Origtime in 1500 CE					0.001	-0.001**	0.000
					(0.001)	(0.001)	(0.001)
Origtime in 1500 CE squared							-0.000
							(0.000)
State Age in 1500 CE							-0.092*
							(0.049)
Observations	112	112	110	110	107	107	107
R-squared	0.389	0.521	0.532	0.616	0.809	0.902	0.911
Panel B	Technology Adoption in 2000 CE						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Statehist	0.086	0.842***		0.664**	0.302	0.461*	0.604**
	(0.095)	(0.318)		(0.332)	(0.239)	(0.244)	(0.272)
Statehist squared		-1.285***		-1.192**	-0.405	-0.554*	-0.531*
		(0.452)		(0.459)	(0.347)	(0.322)	(0.319)
Agyears			0.011	0.011	-0.007	-0.004	0.002
			(0.007)	(0.010)	(0.008)	(0.012)	(0.012)
Origtime					0.000	0.001**	0.001
					(0.000)	(0.001)	(0.002)
Origtime squared							0.000
							(0.000)
State Age							-0.030
							(0.021)
Observations	130	130	129	129	125	125	125
R-squared	0.006	0.044	0.016	0.050	0.643	0.683	0.688
Controls	No	No	No	No	Yes	Yes	Yes
Continent FE	No	No	No	No	No	Yes	Yes

Note: In panel A, the dependent variable is the technology adoption index in 1500 CE and the main independent variables are the *extended statehist index* between 3500 BCE and 1500 CE, linear and squared. In panel B, the dependent variable is the technology adoption index in 2000 CE and the main independent variables are the *extended statehist index* between 3500 BCE and 2000 CE, linear and squared. The list of controls includes: absolute latitude, an indicator of whether the present-day country is landlocked, distance to coast and rivers, mean elevation, land suitability, percentage arable land, temperature, precipitation, percentage population at risk of contracting malaria. Robust standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1

Table 8: State age and the emergence of writing

Panel A	Time since Writing emerged. All states					
	(1)	(2)	(3)	(4)	(5)	(6)
State Age in 2000 CE	0.701***		0.506***	0.566***	0.566***	0.561***
	(0.040)		(0.091)	(0.111)	(0.114)	(0.116)
Agyears in 2000 CE		0.387***	0.150***	0.074	0.076	0.071
		(0.022)	(0.045)	(0.056)	(0.062)	(0.063)
Origtime in 2000 CE				0.001	-0.002	0.001
				(0.002)	(0.002)	(0.006)
Origtime in 2000 CE squared						-0.000
						(0.000)
Observations	143	138	138	116	116	116
R-squared	0.729	0.647	0.780	0.816	0.830	0.830
Panel B	Time since writing emerged. States that precede writing					
	(1)	(2)	(3)	(4)	(5)	(6)
State Age in 2000 CE	0.724***		0.607***	0.594***	0.642***	0.652***
	(0.037)		(0.113)	(0.143)	(0.158)	(0.160)
Agyears in 2000 CE		0.396***	0.085	0.056	0.066	0.067
		(0.026)	(0.062)	(0.083)	(0.087)	(0.087)
Origtime in 2000 CE				0.005**	0.002	0.021
				(0.002)	(0.004)	(0.020)
Origtime in 2000 CE squared						-0.000
						(0.000)
Observations	80	76	76	64	64	64
R-squared	0.840	0.736	0.870	0.896	0.903	0.905
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Continent FE	No	No	No	No	Yes	Yes

Note: The dependent variable is the *time since writing emerged*, measured in thousands of years. In both panels the main independent variable is *State age*, in thousands of years. In panel A we present estimates from the entire sample, while in Panel B we present estimates from the subsample of countries for which state emergence preceded the emergence of writing. The list of controls includes: absolute latitude, an indicator of whether the present-day country is landlocked, distance to coast and rivers, mean elevation, land suitability, percentage arable land, temperature, precipitation, percentage population at risk of contracting malaria. Robust standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1

APPENDIX – For Online Publication

Appendix A. Variables Description and Data Sources

A1. State History Variables:

Statehist. The *extended statehist* is the normalized aggregate index of state history. This index is defined as the sum of all 50-year period state history scores, adjusted by a discount factor, divided by the maximum value of a discounted index, corresponding to a state with a score of 50 in every half century between 3500 B.C.E. and 2000 C.E.. The index can be calculated using various discount rates to put more weight on recent history than on the distant past. Throughout the paper we use the 1% discount rate.

Statehist 1-1950 CE. This is the *statehist* computed according to the initial version of the index in Bockstette et al. (2002), considering only the period 1 – 1950 C.E. This is also a normalized index (with respect to a virtual state that would have full scores for every half century between 1 and 1950 C.E.). In this paper we use a discount factor of 1% for this index.

Ancestry – Adjusted Statehist in 1500 C.E. This is the extended *statehist* index at 1500 C.E., adjusted by the migration matrix of Putterman and Weil (2010), as follows: for each country *i* in the sample, we use the matrix to identify the share of the current population that has ancestry that can be traced to the territory of country *j;* for each country *j* we multiply its statehist score accumulated at 1500 C.E. (discounted by 1% and normalized with respect to a state with full scores from 3500 B.C.E. until 1500 C.E.) by a weight which is the share of population in country *i* identified to have roots in country *j;* the sum of all weighted *statehist* scores thus obtained across all *j* is the ancestry-adjusted *statehist* index at 1500 C.E. Using this adjustment, we account for the state experience prior to 1500 C.E. of other territories, brought by post-1500 migrants into a certain country, in addition to the state history of the country's own territory.

Ancestry – Adjusted Statehist in 2000 C.E. This is the extended *statehist* index at 2000 C.E., adjusted by the migration matrix of Putterman and Weil (2010), as follows: we compute the *statehist* index between 1500 and 2000 C.E. (discounted by 1 % and normalized with respect to the same period); we add this score to the ancestry – adjusted statehist in 1500 C.E. defined as above.

State Age. This variables, scaled in millennia, represents the total amount of time elapsed from the first date (exact or approximate) when state experience is assigned a positive scores (the first date when the component S1 pertaining to the existence of a rule above tribal level is positive) until 2000 C.E. State age does not account for periods of state collapse (scores revert to 0) incurred after the original state emergence date.

A2. The emergence of writing

Time since writing emerged. This novel variable, which we scale in millennia throughout the analysis in this paper, represents the time elapsed in 2000 C.E. since the date of the first evidence of existence of a writing system on the territory of a given modern-day country. Only forms of writing beyond proto-writing are considered, the earliest if which are considered to be the cuneiform script in Mesopotamia (Iraq, dated around 2900 B.C.E) and the hieroglyphic script of ancient Egypt (which new evidence indicates to have been already invented in 3400 B.C.E.) In order to document the first evidence of writing on the territory of each country in the sample, Encyclopaedia Britannica served as main source, but various other sources have been consulted. The variable is coded for a total of 155 countries, 146 of which have the writing information coded with relative certainty. The values of the variable for each country along with the corresponding reference are found in the online appendix.

A3. Historical controls

Agyears. This variable assembled by Putterman with Trainor (2006) records for each present-day country in a sample of 170 countries the number of millennia elapsed in 2000 C.E. since the Neolithic transition of populations that lived on the territory of that country. The year of transition is assigned by cross-referencing expert opinions about the time when the population is a particular region covered more than half of their calorie intake from agriculture.

Origtime. This variable coded by Ahlerup and Olsson (2012) represents the time since initial uninterrupted human settlement (before 2000 C.E.) on the territories that now belong to modern-day countries. The variable was coded for 191 countries and the coding was based on Oppenheimer (2003) and Bradshaw Foundation (2007), as well as Encyclopedia Britannica (2007) for the island cases. Since the original settlements follow the paths of the migration routes out of Africa, the variable is correlated with the migration distance, and can therefore also be employed as a proxy for the latter. [1]

The Matrix of Migration since 1500 C.E.. This matrix was developed by Putterman and Weil (2010) to describe the composition of the populations of modern-day countries in terms of ancestry at 1500 C.E., before the migration flows of the colonial era. The matrix contains 165 rows (each row corresponding to a present-day country) and 165 columns (representing the same countries), where every cell records the percentage of current population in country on row *i* that traces its ancestry to the population in the source country on column *j*, such that the sum of all cells on each row is 1. In their paper, Putterman and Weil (2010) obtained ancestry-adjusted measures of *statehist* (1-1950 C.E.) and *agyears*, by multiplying each row to each one of the vectors containing the values of their variables of interest at 1500 C.E. for each country in their sample (which amounted to a sum over the values of the variables of interest of each source country by the corresponding share of the population with ancestry in other countries). We follow the same procedure in order to obtain the extended ancestry-adjusted *statehist.*

A4. Outcomes variables

GDP per capita in 2000. Data in current US dollars, as reported by the World Bank. In a robustness check in appendix C, we use the real GDP per capita in constant 2000 international dollars from World Penn Table, version 6.2.

[1] In our sample the correlation coefficient between the time since original human settlement and the migration distance of Ashraf and Galor (2013) is -0.51, which indicates that the shorter the migration distance to a particular territory, the earlier the first human settlement.

Population density in 1 C.E. and 1500 C.E.. This variable is measured in number of individuals per square km. The variable is retrieved from Ashraf and Galor (2013), who employ the population size data from McEvedy and Jones (1978), and the land area from World Development Indicators. Since the territorial unit employed in McEvedy and Jones (1978) is based on 1975 country borders, in some cases, the same value of the population density is assigned to contiguous present-day countries (that may have been part of the same constituency in 1975, such as Yugoslavia).

Urbanization in 1 C.E. This measure of urbanization was computed by Peregrine (2003) in the Atlas of Cultural Evolution (ACE). The variable takes three values: 1 if the largest settlement on the territory defined by the borders of a given modern-day country was under 100 persons, 2 if the largest settlement was between 100 and 399 persons and 3 for settlements larger than 400 persons. The variable had been previously used by Comin, Easterly and Gong (2010), from whose dataset we retrieved it.

Urbanization in 1500 C.E. The urbanization rate for 1500 A.D. comes from Acemoglu, Johnson and Robinson (2005) and it is calculated as the percentage of a country's urban area population (for cities with at least 5,000 inhabitants).

Technology Adoption in 1 C.E.. These variables are three indices created by Comin, Easterly and Gong (2010), henceforth CEG. The index in 1 C.E. is based on data from Peregrine's (2003) "ACE" in which various cultural traits of 289 prehistoric cultures are evaluated: writing, agriculture, transportation, urbanization. CEG used this and additional data to code country – level data on technology adoption in five sectors: agriculture, transportation, communications, writing, and military. The authors structured the information in "ACE" into indicators that denoted the presence or absence of a technology within a certain sector and territory, which they then averaged over to create the sector specific technology adoption index between 0 and 1 (e.g. where "ACE" codes "technological specialization" by 1 for none, 2 for pottery and 3 for metalworks, CEG marked pottery and metalwork as the two potential technologies within the "industry sector" at the time, which they coded using a binary convention 1 – if technology is present and 0 if not; the average over all these dummies within every sector is the value of the technology adoption index for that sector; this average for the industry sector in this case would be 0 if neither technology was present, 0.5 if only one was present and 1 if both were present). Then, the overall adoption level, the variable that we use in this paper, for each country was calculated as the average of the adoption levels across sectors.

Technology Adoption in 1500 C.E.. For the average technology adoption measure in 1500 C.E., CEG (2010) used many different sources to summarize information on 20 technologies across 4 sectors excluding agriculture (for instance, for "Industry", the two possible technologies are "presence of iron" and "presence of steel"). For the latter they used a proxy based not on technology presence, but rather on which type of agriculture was the primary source on a particular territory – e.g. pastoralism, hand or plough cultivation, or none). As with overall technology adoption in 1 C.E., the overall measure in 1500 C.E. is obtained by averaging over the scores for each sector.

Technology in 2000 C.E. The technology measure in 2000 C.E. is constructed in CEG(2010), based on Comin, Hobijn and Rovito(2008) and it captures the gap in the intensity of technology adoption for every country with respect to the US (in terms of years of usage of each technology relative to the number of years since the invention of that technology) for ten technologies: electricity, internet, PC's, cell phones, telephones, cargo and passenger aviation, trucks, cars and tractors, in per capita terms. The average across the technologies' scores is subtracted from 1 (the level of US, by construction) to obtain the country-level technology adoption gap measure. This measure is different from the measures for 1 and 1500 C.E., since it also measures adoption along the intensive margin.

A5. Geographical variables

Absolute latitude. This is the absolute value of the country's centroid latitude. The variable was retrieved from the Portland Physical Geography dataset.

Distance to coast and river. This variable represents the mean distance to the nearest coastline or sea-navigable river, measured in km. The variable was retrieved from the Portland Physical Geography dataset.

Mean elevation. The mean elevation above sea level is measured in meters. The variable was retrieved from the Portland Physical Geography dataset. The original source is NOAA's National Geophysical Data Center.

Land suitability. This is a measure of land suitability for agriculture, computed at country level by Michalopoulos (2012), based on grid-cell data reported by Ramankutty et al. (2002). For details on the construction of the original index, the reader is referred to Ramankutty et al (2002). The index includes information on ecological indicators of climate and soil suitability for agriculture (such as drowing degree days, evapotranspiration, soil carbon density and soil pH).

Percentage arable land. This measures the percentage of a modern-day country's area that is arable. The source is World Bank's World Development Indicators.

Temperature. This a mean across the average monthly temperature over time (1961-1990) in 1-degree resolution grids within a country. This variable was retrieved from Ashraf and Galor (2013), whose source is the G-ECON project (Nordhaus 2006).

Precipitation. This a mean across the average monthly precipitation over time (1961-1990) in 1-degree resolution grids within a country. This variable was retrieved from Ashraf and Galor (2013), whose source is the G-ECON project (Nordhaus 2006).

Malaria (percentage population at risk). This variable represents the level of risk of contracting malaria (measured by the percentage population in 1994 in areas of high risk of contracting malaria, times the share of cases in the country involving fatal species of P. Falciparum). The original data was constructed by Gallup and Sachs (2001).

Landlocked. This is a dummy variable equal to 1 if the country is landlocked.

A6. Remarks on data and regressions of Petersen and Skaaning (2010)

In Section 4.1 of our paper, especially tables 2-3, we investigate the impact of timing of transition to agriculture on timing of emergence of states. An exercise in Table 3 of Petersen and Skaaning (hereafter, P&S) bears similarity to our initial regressions in Table 2, but differences are sufficient to merit placing comparative remarks in this appendix. The first remark concerns the data used. P&S use the same data for year of transition to agriculture as does our paper, and they also take year of state emergence from the same source, for states listed by it (Putterman, 2007) as starting in 1 CE or later. However, they independently determine years of state emergence for those countries having a state before that year. Because their data were not published, and because unlike our project they did not code state presence, degree to which locally based, or geographic extent and unity on a continuous basis for the half centuries of the BCE period, we did not access their data while developing ours. However, we requested their data as our paper was being completed, and found that our estimates of early states' emergence times are 431 years earlier, on average, mainly due to our inclusion of paramount chiefdoms and of corresponding archeological evidence of incipient states when assessing state emergence (i.e., *State age* begins for us at the paramount chiefdom or corresponding level). The correlation of estimated year of state emergence in our own and P&S's data is 0.82. Regarding the estimation of partial correlation coefficient on year of agricultural transition in a regression for year of state emergence, the most similar specifications are P&S's model 2 and our Table 2 column 1. Those two models differ in that P&S always include the biological conditions measure from Hibbs and Olsson (2005) in addition to *Agyears*, which we do not do. In our full series of related regressions, in tables 2 and 3, we also include numerous geographic controls absent from the corresponding P&S work, complete with continent fixed effects. And we exam separately the impact of *Agyears* on *Stateage* among internally versus externally originated states, as well as the impact of *Agyears* on *Statehist*, which reflects the stock of state experience at various points in time.

Appendix B – Additional *Statehist* coding information and illustrations

To code all components of the index in a manner that is consistent across periods, we relied mainly on information in the Encyclopedia Britannica Online. We surveyed the main articles on the history of the modern-day country (E.g."History of Azerbaijan"), but also articles connected to events in its history (e.g. "Azerbaijan -historical region", "Ancient Iran: The Sāsānian period"). There were a number of instances where information in Britannica was sparse, in which cases we surveyed alternative sources, such as books or journal articles treating individual cases.

Table B1 illustrates an example of coding based on information from Encyclopedia Britannica, covering the period 450 BCE - 1 CE for the territory of modern-day Bosnia and Herzegovina.

Table B1. Coding example - the case of Bosnia and Herzegovina, 400 BCE – 1 CE.

Year BCE	Government is above tribe level? (z^1_{it})	Government domestic? (z^2_{it})	What percentage of the territory is ruled? (z^3_{it})	Weight* ($w_{it\theta}$)	s_{it}
400-201	0.75	1	0.3	1	11.25
200-151	0.75	1	0.3	0.9	12.625
(cutoff at 155)	1	0.5	1	0.1	
150-1	1	0.5	1	0.72	25

CODING INFORMATION

400 – 200: (0.75, 1, .3). From the 4th century BCE, along with the coming of Celtic tribes in the area, the Illyrian tribes became gradually more politically cohesive. Sources recall the existence of early indigenous petty kingdoms in Illyria on the territory of present-day Albania only. We therefore mark the occasional Illyrian tribe alliances by z^1_{it}=.75.

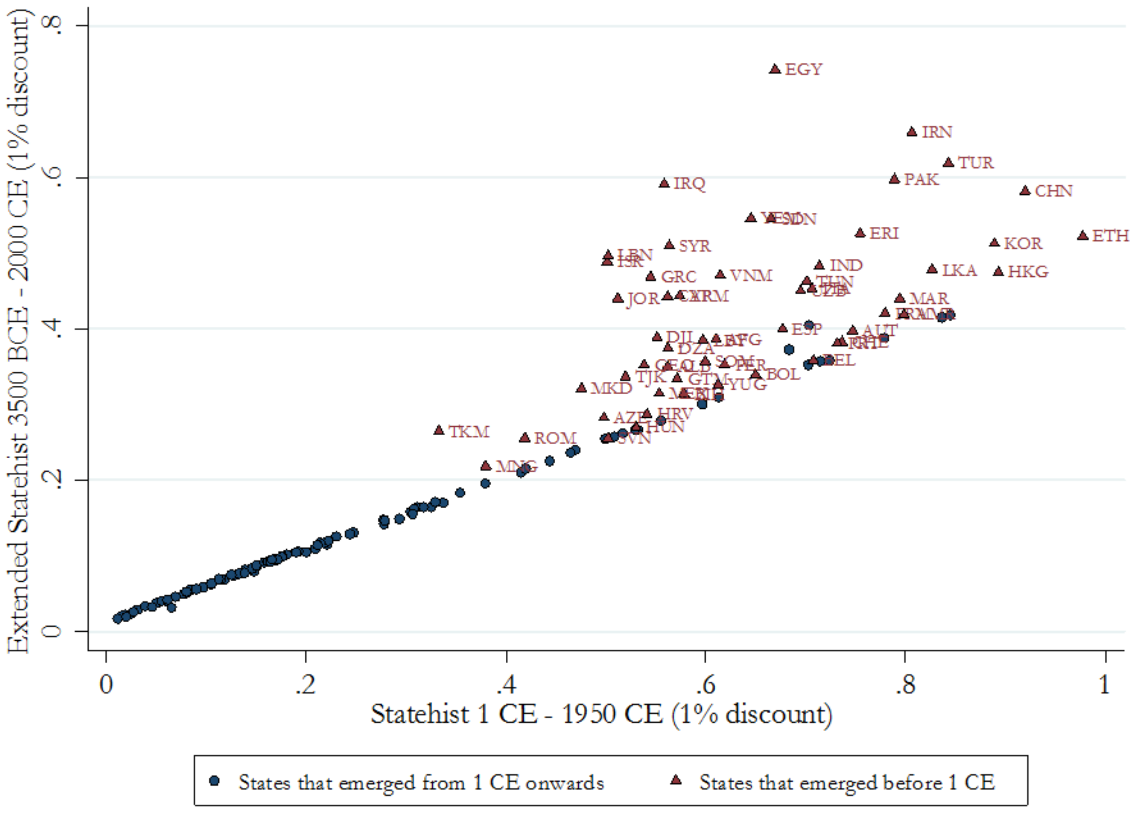
200-151: (0.75, 1, .3) until 155 and (1, .5, 1) from 155 onwards. Delminium (on modern-day Bosnian territory) was taken by the Romans in 155 BCE, hence z^2_{it}=.5. Most of the area of Bosnia was integrated in the Roman province Dalmatia, hence z^3_{it}). The score is [0.9(0.75, 1, .3) + 0.1(1, .5, 1)]*50.

150-1: (1, .5, 1). Bosnia was under Roman occupation.

Note: $w_{it\theta}$ = number of years between period ends and cutoff, or between two cutoffs, divided by 50.

Figure B1 shows the comparison between the original old *Statehist* index for the CE-period and the new one presented in this paper. A notable feature is that the previous index failed to reflect the state history of several ancient civilizations like Egypt and Iran, countries that now receive a substantially higher score. Ethiopia, which had a full state by 1 CE but was a relative newcomer compared to those just mentioned, has the highest state history value going by the old measure, but now loses in relative terms to the older civilizations.

Figure B1. Extended Statehist (3500 BCE – 2000 CE) vs Statehist 1-1950 CE



Note: The red, triangular observations with 3-letter isocodes show country observations where states emerged before 1 CE and whose index score changes considerably with the extended coding. Both variables use a 1 percent discount rate.

Due to discounting, the *Statehist*-index does not provide a strong sense of how long a state has existed on a country's territory. A place with thousands of years of early state history can have a relatively low *Statehist* if incorporated into empires during recent centuries, for example. An intuitive non-normalized (and non-discounted) measure of state history is the simple aggregation of a country's half-century scores

$$\sum_{t=0}^{109} s_{it} = \tilde{S}_i$$

which we can refer to as *Full state equivalent years of state history*. Table B2 shows the countries with the highest and lowest full state-equivalent years. Egypt is by far the country with the longest state history thus defined, having accumulated the equivalent of 4116 years as a home-based state with at least 50 percent of its current territorial extent. An interpretation could be that since the first state in Egypt emerged 5200 years ago, 1084 full state-equivalent years were lost through history due to state collapse, foreign invasions and dominance, etc. Also Iran, Iraq, Turkey and Pakistan have had very long state histories. The *State age* (total number of years since state emergence until 2000 CE) and *Statehist* index are also displayed for these countries. It should be noted that the ranking of countries by full state-equivalent years is not the same as that by *State age* or *Statehist*, owing to the variability of the s_{it}-score across time.

Table B2. Full state equivalent years, State age and Statehist: The 10 largest and 10 smallest country levels

Country	**Full state equivalent years** ($\tilde{S}_i$)	**State age**	**Statehist** (Non-normalized)	**Statehist** (Normalized)
10 largest $\tilde{S}_i$				
Egypt, Arab Rep.	4116.34	5200	2495.010	0.742
Iran, Islamic Rep.	3280.94	5200	2217.940	0.660
Iraq	3232.94	5500	1988.640	0.591
Turkey	2962.31	5000	2079.380	0.618
Pakistan	2922.81	4600	2007.710	0.597
Yemen	2677	4000	1835.450	0.546
Syrian Arab Republic	2617.88	4600	1713.200	0.509
Lebanon	2607.56	4850	1669.840	0.497
Sudan	2603.69	5000	1832.890	0.545
China	2566.2	3900	1955.590	0.582
10 smallest $\tilde{S}_i$				
New Zealand	113.25	150	112.379	0.033
Zimbabwe	110	150	108.767	0.032
Gabon	95	150	94.259	0.028
Sierra Leone	87.18	213	86.399	0.025
Fiji	78	126	77.496	0.023
Comoros	73	114	72.614	0.021
Central African Republic	70	100	69.752	0.020
Kenya	68.5	100	68.252	0.020
Namibia	63	116	62.594	0.018
Papua New Guinea	56.25	100	56.064	0.016

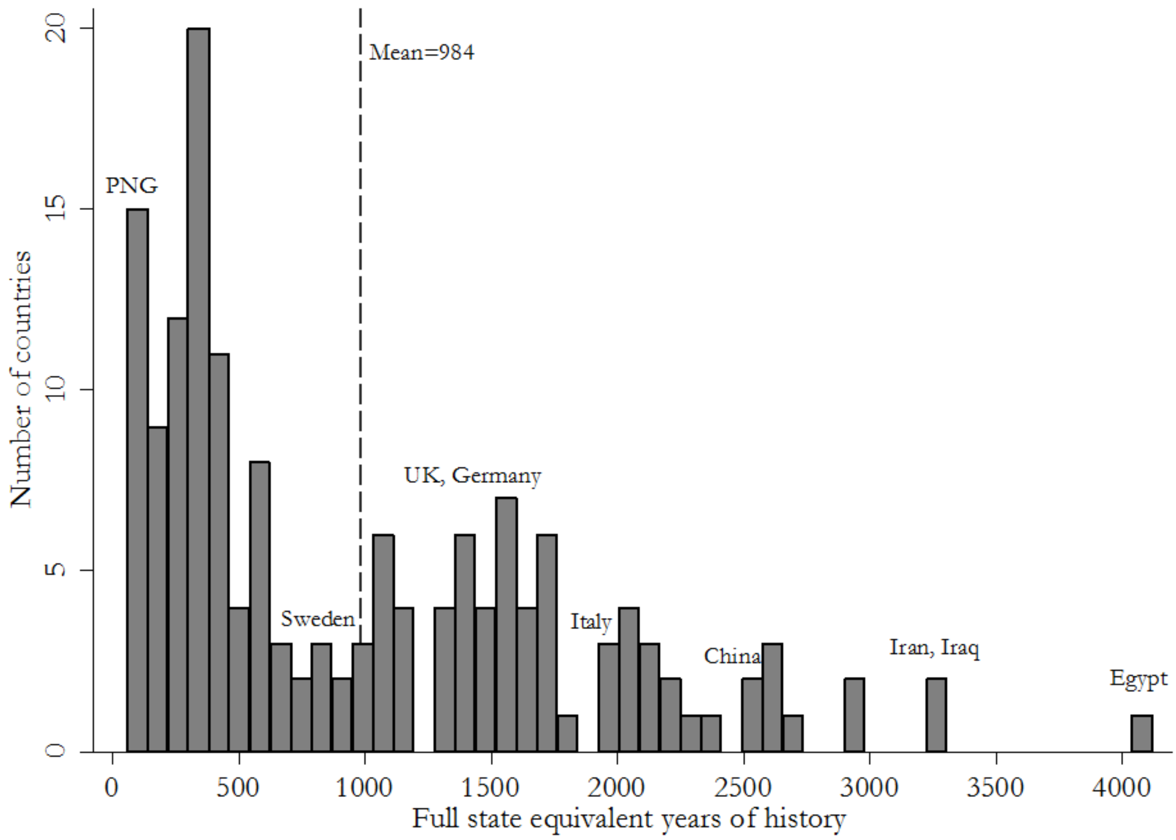
Notes: As discussed in the text, *Full state equivalent years* ($\tilde{S}_i$) is simply the undiscounted sum of all half-century scores s_{it} from the first state emergence until 2000 CE. *State age* shows the time since the appearance of the first state (including paramount chiefdom) on the country's territory. We also report the non-normalized and normalized *Statehist* (S_{i0}, the index calculated for 2000 CE, discounted 1%). Note that the ordering by *Full state equivalent years* is not perfectly consistent with the ordering by *State age* or *Statehist*.

At the other end of the scale, we find Papua New Guinea with the shortest full state equivalent history of all - only 56 years. Among the ten countries with the shortest state history, seven are African. In general, the distribution of $\tilde{S}_i$ is strongly skewed to the left, as shown in the histogram in Figure B2. The mean is 984 years, whereas the median is 639. Some of the countries in the middle of the distribution like UK (1415 years) and Germany (1420 years) are today among the world's richest countries.

These aggregated measures of state history hide many interesting patterns for individual countries. For instance, several countries with a long state history have experienced repeated periods of fundamental regime changes and state failures.

Figure B3 shows the *State index* of only two individual countries - Iraq and Sweden - representing one of the oldest and one of the youngest state formations in the Western core area. Their *Full state equivalent years* of state history are 3233 and 781 years respectively

Figure B2. Distribution of full state-equivalent years of state history



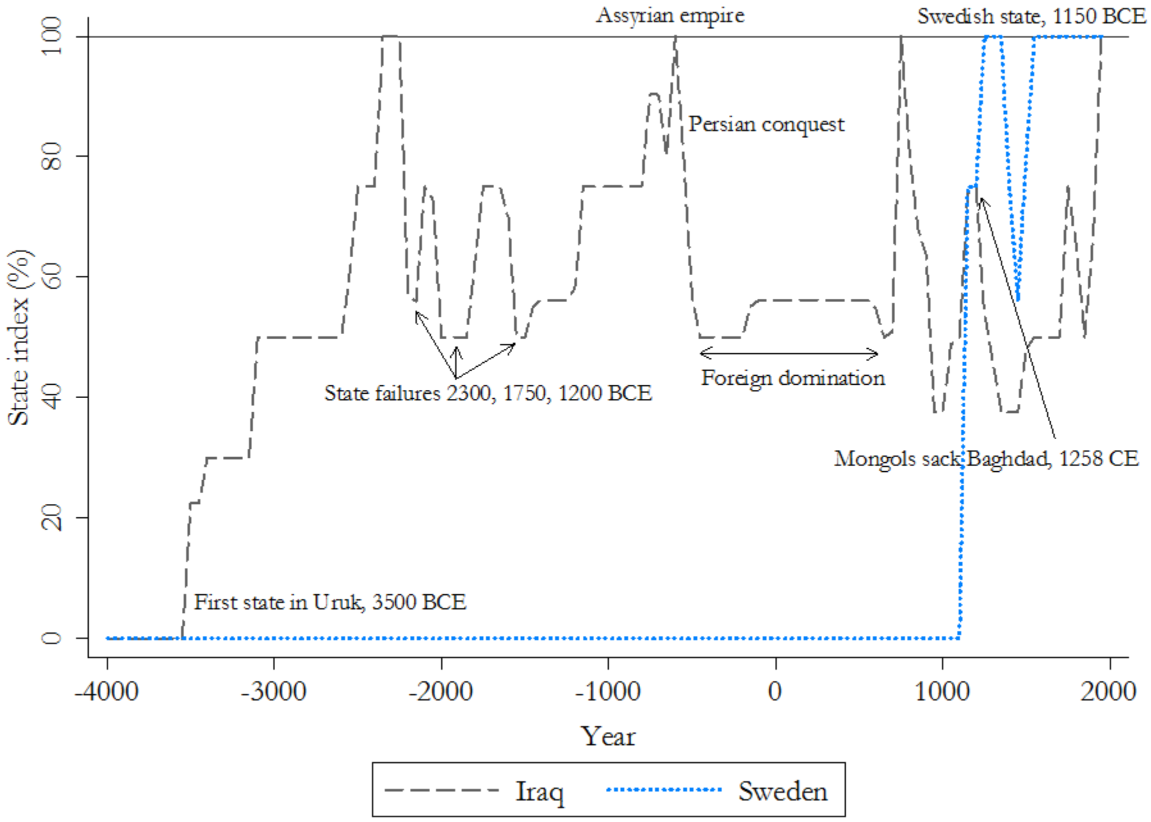
Note: The graph shows the distribution of our measure of *Full state equivalent years* of state history ($\tilde{S}_i$). See the text for variable definition. The frequency of countries is shown on the vertical axis. Selected countries are identified in the graph.

The first known state in the world according to our index arose in Uruk in current Iraq around 3500 BCE and then developed into a full-blown civilization that heavily influenced large parts of historical Mesopotamia and neighboring regions for several centuries. Prominent cities such as Ur, Kish, Nippur, and Lagash inherited the early civilization of Uruk and developed it even further.

A period of stagnation then set in from 2300 BCE with repeated state failures in 1750 and 1200 BCE. A gradual recovery of state strength then took place up until the heyday of the Assyrian empire, whereupon the area started to be dominated by Persians from Iran from about 550 BCE. The Persian conquest is the most long-lasting crisis in Iraq's state history, according to the index used. Baghdad then revived as a culturally and economically important center around 1000 CE, but this period came to an abrupt end with the Mongols' sacking of Baghdad in 1258.

Sweden's history is very different. A state above tribal level did not arise here until 1150 CE - more than four millennia after Uruk's start - but immediately emerged as a full-blown state in terms of our index. The only setback (a period of Danish domination) did not last long. Since 1527 CE, Sweden has had a maximum score on the State index.

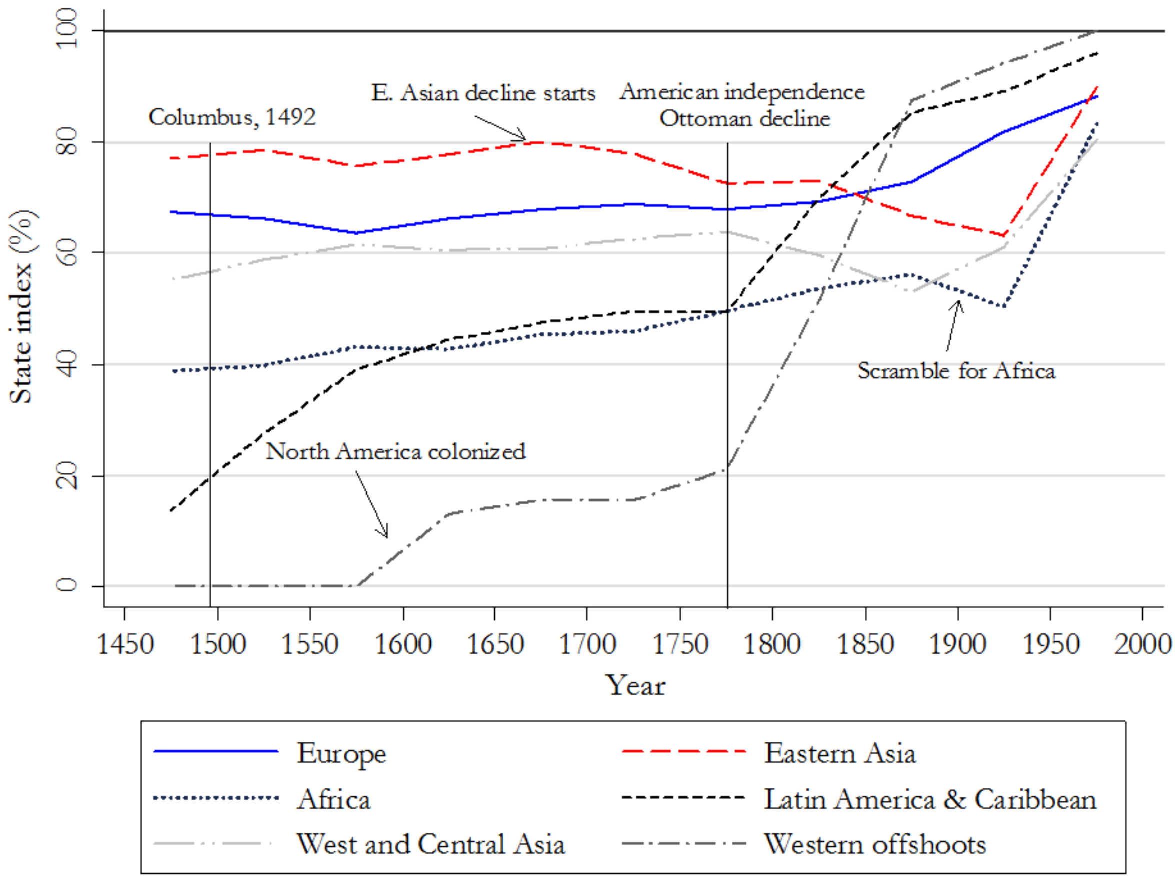
Figure B3. State history in Iraq and Sweden, 3500 BCE – 2000 CE



Note: The lines show the development of the *State index*-score in Iraq and Sweden. The *State index* scores for each period are simply equal to $2s_{it}$ so that a score of 100 percent implies a s_{it}=50. Particularly important historical episodes are marked.

Appendix C – Supplementary Figures

Figure C1. Emergence of states in six world regions during the colonial era, 1450-2000 CE



Note: The figure shows the development of the aggregated *State index* in Europe, Eastern Asia, West and Central Asia (including Turkey and India and the located countries in between), Latin America & the Caribbean (all countries in the Americas except Canada and USA), Africa (including North Africa), and the Western offshoots (USA, Canada, Australia, and New Zealand). Oceania is omitted. On the horizontal axis, negative values imply years BCE whereas positive values show the CE-period. Particular years with trend breaks are marked.

Figure C2. Transition to agriculture and state history

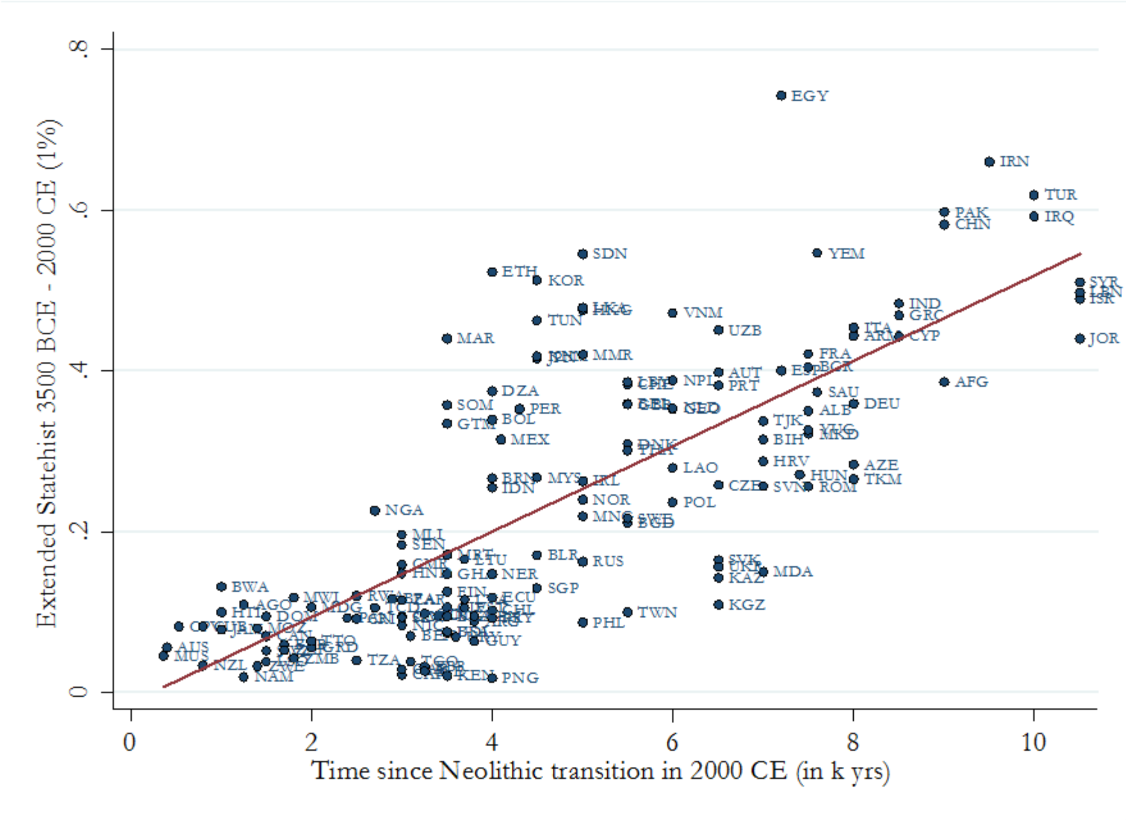
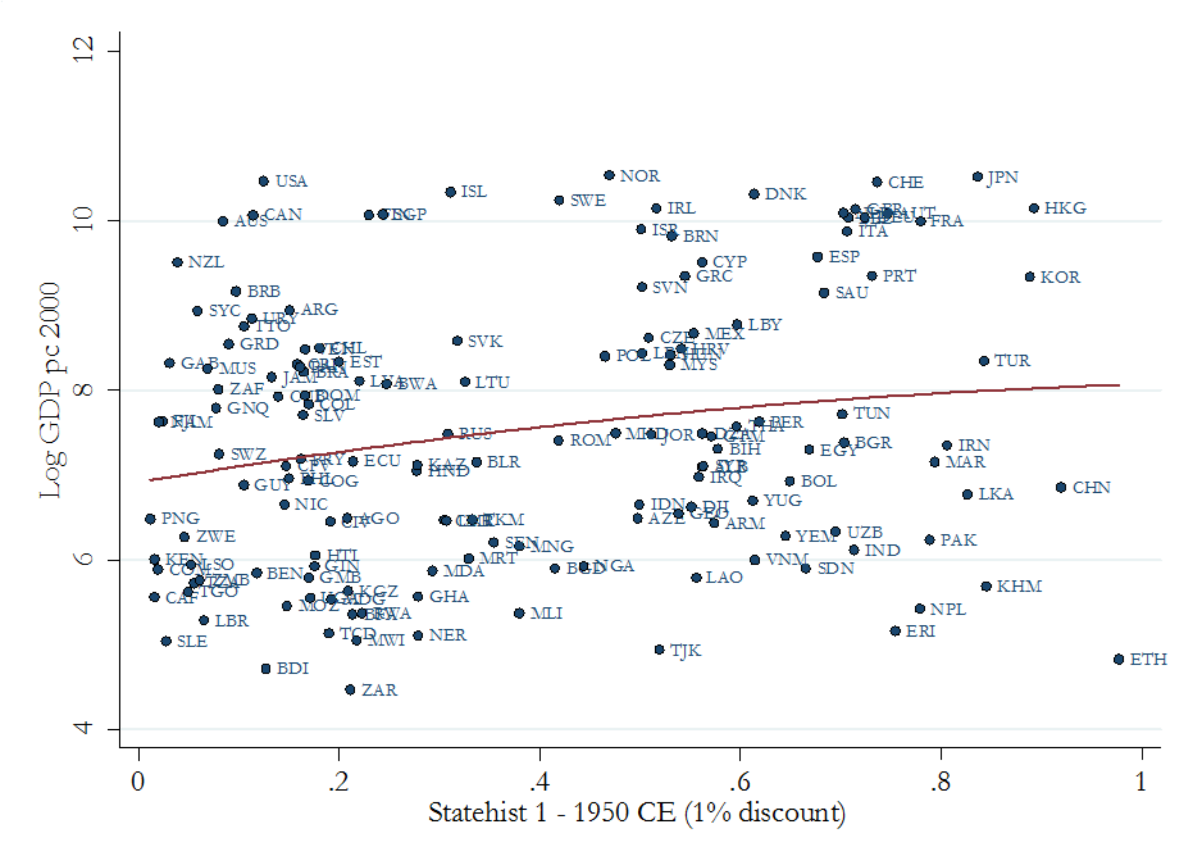


Figure C3. Non-linear relationship between Log GDP per capita in 2000 and Statehist 1 – 1950 CE



Note: The figure shows a fitted quadratic regression line corresponding to the estimates in Table 4, Panel B, column 2, with 154 country observations distinguished by 3-letter country isocodes.

Figure C4. GDP pc in 2000 C.E. and state history in internally-originated states

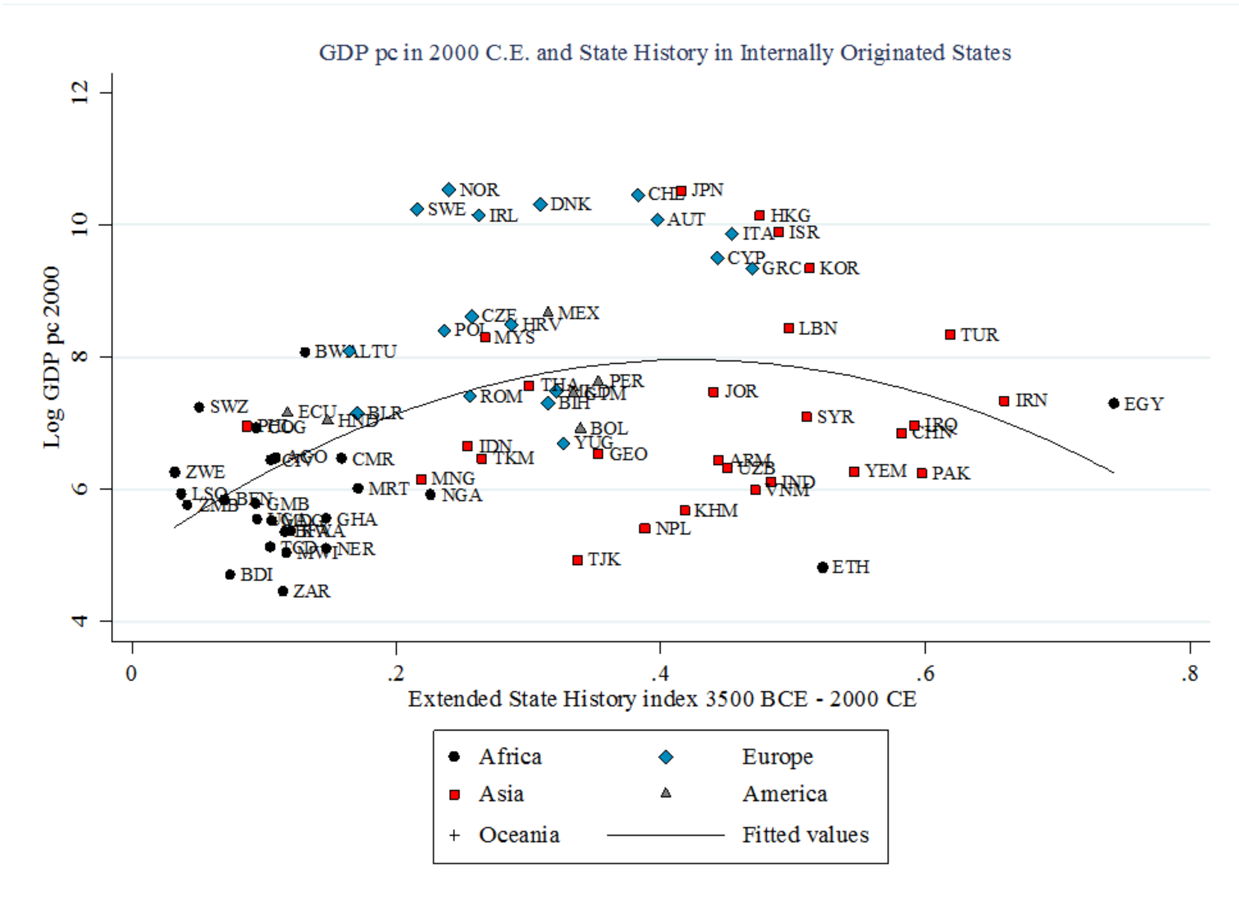


Figure C5. GDP pc in 2000 C.E. and state history in externally-originated states

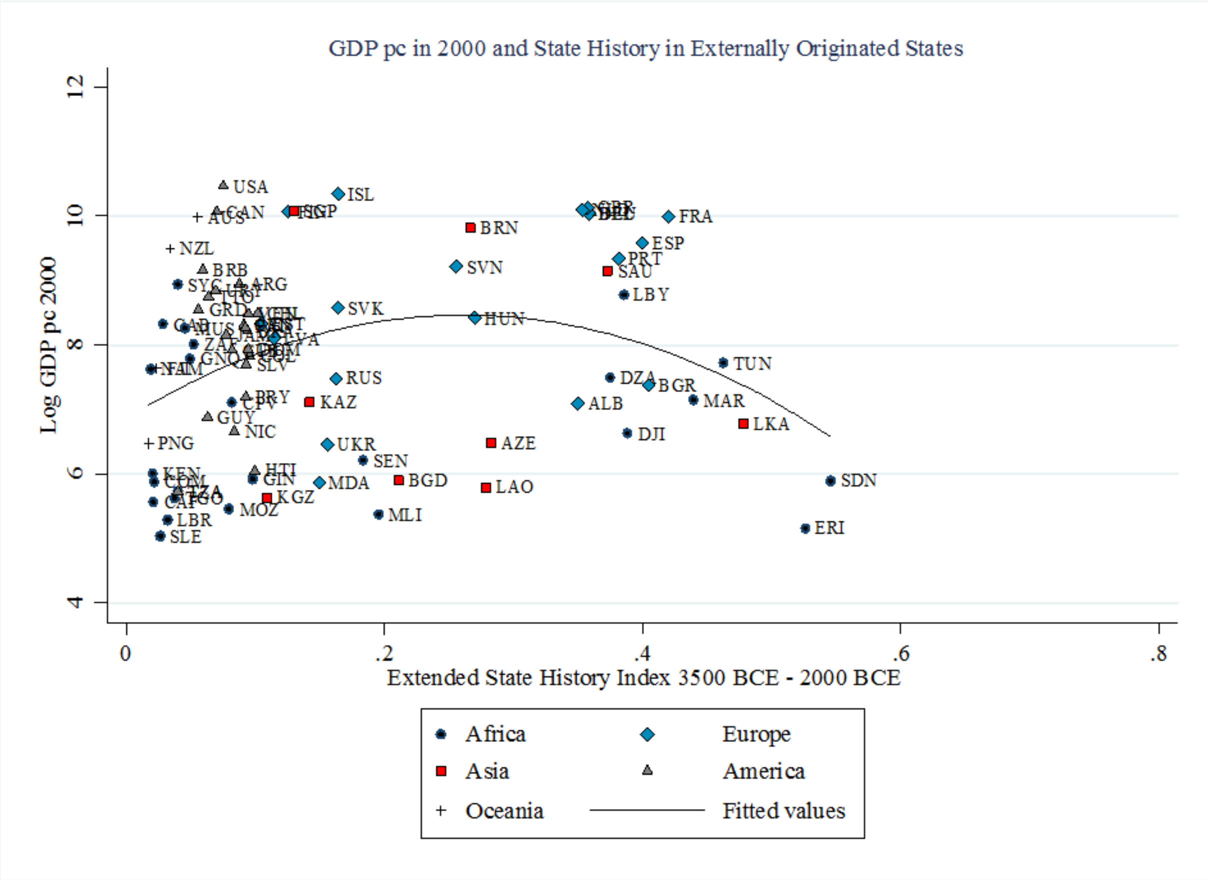
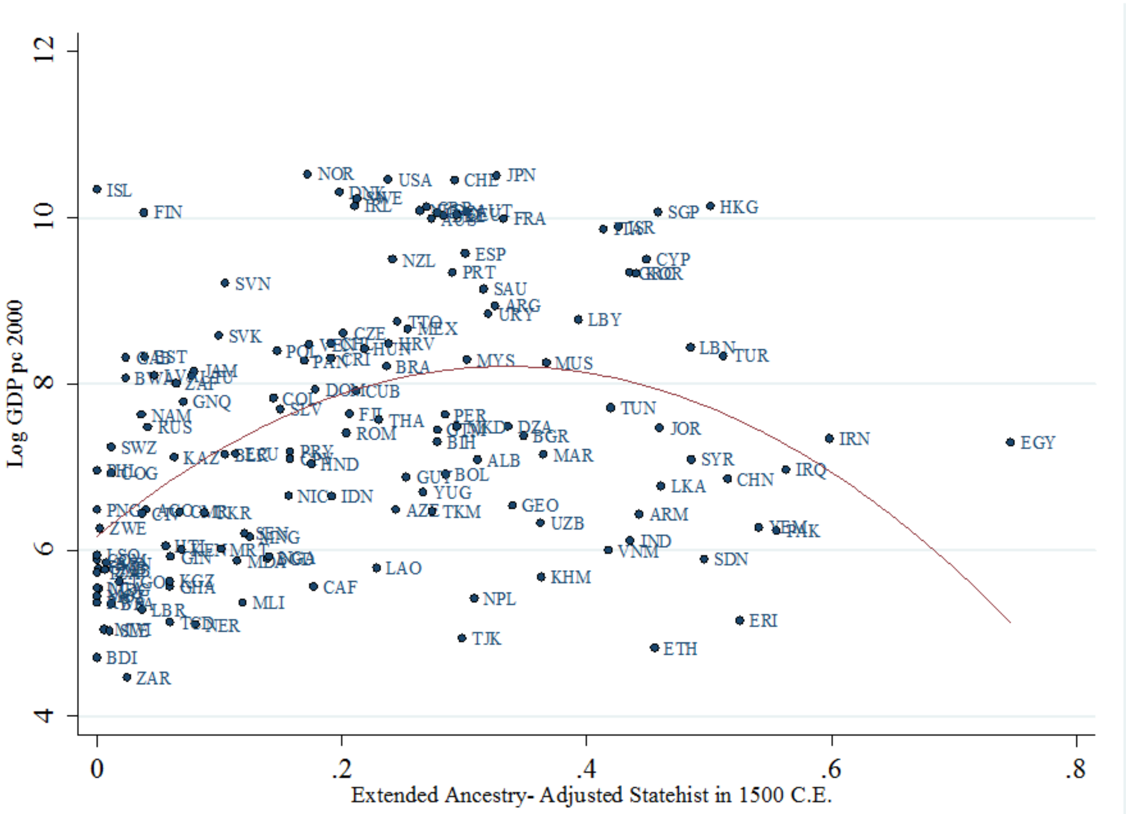


Figure C6. GDP pc in 2000 C.E. and ancestry- adjusted state history



Appendix D – Supplementary Tables

Table D1. Robustness checks – The Neolithic transition, state onset and state history.

	State Age				Statehist			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Agyears	0.468***	0.465***	0.468***	0.469***	0.050***	0.049***	0.049***	0.050***
	(0.027)	(0.028)	(0.028)	(0.027)	(0.004)	(0.004)	(0.005)	(0.005)
Origtime	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
Absolute centroid latitude	0.025	0.043*	0.024	0.023	0.004*	0.006*	0.005**	0.004*
	(0.017)	(0.022)	(0.018)	(0.017)	(0.002)	(0.003)	(0.002)	(0.002)
Landlocked	-0.535***	-0.498***	-0.485***	-0.484***	-0.057**	-0.053**	-0.051**	-0.049*
	(0.178)	(0.176)	(0.173)	(0.173)	(0.027)	(0.026)	(0.025)	(0.026)
Distance to coast and rivers	0.000	0.000	0.000	0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mean Elevation		0.000*	0.001*	0.001***		0.000*	0.000**	0.000***
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
Land suitability	-0.022	-0.276	-0.073	-0.061	-0.027	-0.054	-0.037	-0.028
	(0.446)	(0.482)	(0.447)	(0.433)	(0.059)	(0.062)	(0.059)	(0.058)
Percentage arable land	-0.010	-0.008	-0.008	-0.009	-0.001	-0.000	-0.000	-0.000
	(0.007)	(0.007)	(0.007)	(0.007)	(0.001)	(0.001)	(0.001)	(0.001)
Temperature	0.094***	0.062	0.093***	0.091*	0.011***	0.008	0.012***	0.009
	(0.031)	(0.038)	(0.031)	(0.052)	(0.004)	(0.006)	(0.004)	(0.006)
Precipitation	-0.004**	-0.004*	-0.004**	-0.004**	-0.000	-0.000	-0.000	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
Malaria (percentage population at risk)	-0.447	-0.367	-0.476*	-0.484*	-0.030	-0.021	-0.031	-0.038
	(0.271)	(0.280)	(0.267)	(0.267)	(0.045)	(0.047)	(0.044)	(0.044)
Mean elevation G-ECON	0.710***				0.087***			
	(0.252)				(0.032)			
Absolute centroid latitude squared		-0.001				-0.000		
		(0.000)				(0.000)		
Mean Elevation squared			-0.000				-0.000	
			(0.000)				(0.000)	
Temperature squared				0.000				0.000
				(0.002)				(0.000)
Constant	-2.573*	-1.841	-2.568*	-2.502*	-0.297*	-0.226	-0.335*	-0.292*
	(1.328)	(1.376)	(1.360)	(1.315)	(0.171)	(0.189)	(0.174)	(0.166)
Continent FE	no	no	no	no	no	no	no	no
Observations	128	128	128	128	128	128	128	128
R-squared	0.780	0.784	0.781	0.781	0.670	0.675	0.675	0.672

Notes: Columns (1) and (5) use a measure of mean elevation from G- ECON project (Nordhaus 2006). Columns (2) and (6) include the squared absolute latitude. Columns (3) and (7) include the squared mean elevation. Columns (4) and (8) include squared average temperature. Robust standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1

Table D2. Robustness checks – Log (GDP) per capita in 2000 and *Statehist.*

	Log (GDP) per capita in 2000				
	(1)	(2)	(3)	(4)	(5)
Statehist	9.565***	16.357***	5.985**	7.306**	7.306**
	(2.547)	(4.402)	(2.382)	(2.867)	(2.867)
Statehist squared	-6.609**	-15.975***	-3.803	-4.791	-4.791
	(2.746)	(5.219)	(2.926)	(3.083)	(3.083)
State Age	-0.663***	-1.892***	-0.434**	-0.557***	-0.557***
	(0.195)	(0.546)	(0.186)	(0.212)	(0.212)
Agyears	-0.480***	-0.040	0.009	-0.001	-0.001
	(0.171)	(0.078)	(0.083)	(0.091)	(0.091)
Agyears squared	0.047***				
	(0.016)				
State Age squared		0.242***			
		(0.087)			
Absolute centroid latitude			-0.022		
			(0.031)		
Absolute centroid latitude squared			0.001		
			(0.000)		
Distance from Addis Ababa				-0.064	
				(0.110)	
Distance from Addis Ababa squared				-0.001	
				(0.003)	
Predicted genetic diversity					22.403
					(73.470)
Predicted genetic Diversity squared					-9.011
					(55.974)
Constant	11.299***	11.312***	10.765***	12.264***	0.320
	(1.459)	(1.492)	(1.545)	(1.804)	(24.057)
Controls	yes	yes	yes	yes	yes
Continent FE	yes	yes	yes	yes	yes
Observations	125	125	125	125	125
R-squared	0.743	0.741	0.725	0.730	0.730

Notes: The variables “Distance from Addis Ababa” and “Predicted Genetic Diversity” are those constructed by Ashraf and Galor (2013). Robust standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1

Table D3. State history, Log Population Density and Urbanization in 1 CE.

Panel A	Log Population Density in 1 CE						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Statehist in 1 CE	4.350***	8.417***		-1.237	1.270	2.754	4.553
	(0.810)	(1.837)		(2.175)	(2.645)	(1.874)	(3.222)
Statehist in 1 CE squared		-8.254***		0.880	-1.063	-3.034	-4.265
		(2.838)		(3.251)	(4.258)	(2.512)	(3.322)
Agyears in 1 CE			0.455***	0.490***	0.456***	0.457***	0.457***
			(0.040)	(0.063)	(0.079)	(0.081)	(0.084)
Origtime in 1 CE					0.001	-0.016***	-0.010
					(0.004)	(0.006)	(0.010)
Origtime in 1 CE squared							-0.000
							(0.000)
State Age in 1 CE							-0.213
							(0.266)
Constant	-0.425***	-0.518***	-1.469***	-1.510***	-6.309***	-6.073***	-6.215***
	(0.140)	(0.144)	(0.159)	(0.168)	(1.267)	(1.296)	(1.270)
Observations	135	135	130	130	115	115	115
R-squared	0.154	0.182	0.455	0.458	0.717	0.800	0.803

Panel B	Urbanization in 1 CE						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Statehist in 1 CE	1.128***	2.778***		0.599	0.235	1.312	1.500
	(0.237)	(0.624)		(0.723)	(1.066)	(0.958)	(1.761)
Statehist in 1 CE squared		-3.260***		-1.013	-0.576	-1.312	-1.452
		(1.045)		(1.117)	(1.474)	(1.164)	(1.404)
Agyears in 1 CE			0.101***	0.093***	0.077**	-0.040	-0.039
			(0.020)	(0.027)	(0.034)	(0.039)	(0.039)
Origtime in 1 CE					0.004*	-0.000	-0.001
					(0.002)	(0.002)	(0.005)
Origtime in 1 CE squared							0.000
							(0.000)
State Age in 1 CE							-0.020
							(0.191)
Constant	2.566***	2.533***	2.348***	2.354***	1.592**	1.304	1.300
	(0.064)	(0.069)	(0.100)	(0.105)	(0.759)	(0.990)	(0.989)
Observations	128	128	128	128	125	125	125
R-squared	0.063	0.087	0.139	0.141	0.371	0.526	0.526
Controls	no	No	no	no	yes	yes	yes
Continent FE	no	No	no	no	no	yes	yes

Notes: In panel A, the dependent variable is log population density in 1 CE and the main independent variables are the *extended statehist index* between 3500 BCE and 1 CE, linear and squared. In panel B, the dependent variable is the urbanization rate in 1 CE and the main independent variables are the *Statehist index* between 3500 BCE and 1 CE, linear and squared. The data for historical population density is based on population data from McEvedy and Jones(1978) and land data from World Bank's *World Development Indicators*. The data for urbanization rate in 1 CE is taken from Comin, Easterly and Gong (2010) and is based on Peregrine (2003) and takes three values: 1 if the largest settlement is smaller than 100 persons; 2 if it is between 100 and 399 persons; and 3 if it is larger than 400 persons. The list of controls includes: absolute latitude, an indicator of whether the present-day country is landlocked, distance to coast and rivers, mean elevation, land suitability, percentage arable land, temperature, precipitation, percentage population at risk of contracting malaria. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table D4. State history and average technology adoption in 1 CE. Ancestry-Adjusted State history and technology adoption in 2000 CE

Panel A	Technology Adoption in 1 CE						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Statehist in 1 CE	0.762***	1.695***		0.415	0.016	0.534	0.788
	(0.122)	(0.279)		(0.390)	(0.519)	(0.351)	(0.601)
Statehist in 1 CE squared		-1.842***		-0.524	-0.150	-0.691*	-0.862*
		(0.479)		(0.572)	(0.746)	(0.412)	(0.508)
Agyears in 1 CE			0.064***	0.055***	0.064***	0.023	0.031*
			(0.007)	(0.012)	(0.016)	(0.015)	(0.016)
Origtime in 1 CE					0.001	-0.002***	-0.005***
					(0.001)	(0.001)	(0.002)
Origtime in 1 CE squared							0.000**
							(0.000)
State Age in 1 CE							-0.026
							(0.060)
Constant	0.688***	0.669***	0.554***	0.562***	-0.062	-0.123	-0.130
	(0.027)	(0.028)	(0.037)	(0.039)	(0.337)	(0.332)	(0.322)
Observations	128	128	128	128	124	124	124
R-squared	0.149	0.189	0.281	0.285	0.541	0.763	0.772
Panel B	Technology Adoption in 2000 CE						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ancestry- Adjusted Statehist in 1500 CE	0.233**	1.332***		1.329***	0.794***	0.672***	0.787***
	(0.092)	(0.275)		(0.309)	(0.215)	(0.219)	(0.223)
Ancestry-Adjusted Statehist in 1500 CE squared		-2.088***		-2.085***	-1.125***	-0.938***	-0.887***
		(0.513)		(0.514)	(0.287)	(0.317)	(0.303)
Agyears			0.011	-0.000	-0.010	-0.005	0.001
			(0.007)	(0.011)	(0.009)	(0.011)	(0.012)
Origtime					0.001*	0.001**	0.001
					(0.000)	(0.001)	(0.002)
Origtime squared							0.000
							(0.000)
State Age							-0.029
							(0.018)
Constant	0.399***	0.314***	0.405***	0.315***	0.708***	0.769***	0.711***
	(0.023)	(0.025)	(0.038)	(0.033)	(0.164)	(0.207)	(0.208)
Observations	130	130	129	129	125	125	125
R-squared	0.037	0.151	0.016	0.150	0.675	0.698	0.704
Controls	no	no	no	no	yes	yes	yes
Continent FE	no	no	no	no	no	yes	yes

Notes: In panel A, the dependent variable is the technology adoption index in 1 CE and the main independent variables are the *extended statehist index* between 3500 BCE and 1 CE, linear and squared. In panel B , the dependent variable is the technology adoption index in 2000 CE and the main independent variables are the ancestry-adjusted *extended statehist index*, between 3500 BCE and 1500 CE, linear and squared. The list of controls includes: absolute latitude, an indicator of whether the present-day country is landlocked, distance to coast and rivers, mean elevation, land suitability, percentage arable land, temperature, precipitation, percentage population at risk of contracting malaria. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table D5. Robustness checks – Log (GDP) per capita in 2000 and ancestry-adjusted *statehist.*

	Log (GDP) per capita in 2000				
	(1)	(2)	(3)	(4)	(5)
Ancestry- Adjusted Statehist	7.326***	7.875***	6.657***	6.919***	6.919***
	(1.724)	(1.911)	(1.890)	(1.918)	(1.918)
Ancestry- Adjusted Statehist squared	-8.103***	-9.780***	-7.532***	-7.476***	-7.476***
	(2.286)	(2.975)	(2.365)	(2.473)	(2.473)
State Age	-0.334**	-0.052	-0.017	-0.013	-0.013
	(0.154)	(0.079)	(0.083)	(0.088)	(0.088)
Agyears	-0.300**	-0.583**	-0.246*	-0.322**	-0.322**
	(0.138)	(0.247)	(0.147)	(0.154)	(0.154)
Agyears squared	0.029**				
	(0.014)				
State Age squared		0.071			
		(0.049)			
Absolute Centroid Latitude			-0.032		
			(0.032)		
Absolute Centroid Latitude squared			0.001		
			(0.001)		
Distance from Addis Ababa				0.007	
				(0.089)	
Distance from Addis Ababa squared				-0.003	
				(0.003)	
Predicted Genetic diversity					70.494
					(60.738)
Predicted Genetic Diversity squared					-46.137
					(46.088)
Constant	10.416***	10.380***	9.861***	11.160***	-15.763
	(1.527)	(1.578)	(1.573)	(1.789)	(20.284)
Controls	yes	yes	yes	yes	yes
Continent FE	yes	yes	yes	yes	yes
Observations	125	125	125	125	125
R-squared	0.737	0.732	0.733	0.734	0.734

Notes: The variables "Distance from Addis Ababa" and "Predicted Genetic Diversity" are those constructed by Ashraf and Galor (2013). Robust standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1

Table D6. Robustness checks – Technology adoption (excluding agriculture) in 1 CE and 1500 CE and *statehist.*

Panel A	Technology adoption in 1 C.E. excluding agriculture					
	(1)	(2)	(3)	(4)	(5)	(6)
Statehist in 1 CE	2.012***		0.526	0.113	0.600	1.025
	(0.328)		(0.464)	(0.564)	(0.391)	(0.691)
Statehist in 1 CE squared	-2.179***		-0.650	-0.271	-0.799*	-1.095*
	(0.567)		(0.682)	(0.821)	(0.462)	(0.587)
Agyears in 1 CE		0.075***	0.064***	0.071***	0.028	0.036*
		(0.008)	(0.014)	(0.017)	(0.017)	(0.019)
Origtime in 1 CE				0.001	-0.002***	-0.006***
				(0.001)	(0.001)	(0.002)
Origtime in 1 CE squared						0.000*
						(0.000)
State Age in 1 CE						-0.044
						(0.069)
Constant	0.604***	0.469***	0.480***	-0.152	-0.105	-0.115
	(0.032)	(0.041)	(0.043)	(0.391)	(0.403)	(0.397)
Observations	128	128	128	124	124	124
R-squared	0.201	0.293	0.298	0.577	0.759	0.768
	Technology adoption in 1500 C.E. excluding agriculture					
	(1)	(2)	(3)	(4)	(5)	(6)
Statehist in 1500 CE	2.969***		1.747***	1.783***	1.019***	1.549***
	(0.386)		(0.484)	(0.339)	(0.197)	(0.447)
Statehist in 1500 CE squared	-3.524***		-2.684***	-2.015***	-0.728**	-0.857***
	(0.797)		(0.948)	(0.643)	(0.322)	(0.295)
Agyears in 1500 CE		0.107***	0.076***	0.041***	-0.000	0.004
		(0.009)	(0.015)	(0.014)	(0.010)	(0.011)
Origtime in 1500 CE				0.001**	-0.001*	0.001
				(0.001)	(0.001)	(0.001)
Origtime in 1500 CE squared						-0.000*
						(0.000)
State Age in 1500 CE						-0.071
						(0.055)
Constant	0.169***	-0.013	-0.007	-0.042	0.031	-0.053
	(0.027)	(0.039)	(0.039)	(0.266)	(0.149)	(0.161)
Observations	111	109	109	107	107	107
R-squared	0.498	0.510	0.589	0.813	0.915	0.922
Controls	no	no	no	yes	yes	yes
Continent FE	no	no	no	no	yes	yes

Notes: The list of controls includes: absolute latitude, an indicator whether the modern-day country is landlocked, distance to coast and rivers, mean elevation, land suitability, percentage arable land, temperature, precipitation, percentage population at risk of contracting malaria. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table D7. State history and Log Population Density in 1 CE and 1500 CE – linear relationship.

Panel A	Log Population Density in 1 CE					
	(1)	(2)	(3)	(4)	(5)	(6)
Statehist in 1 CE	4.350***		-0.734	0.583	0.793	0.614
	(0.810)		(0.861)	(1.097)	(0.811)	(1.025)
Agyears in 1 CE		0.455***	0.484***	0.466***	0.481***	0.463***
		(0.040)	(0.057)	(0.068)	(0.078)	(0.083)
Origtime in 1 CE				0.001	-0.016***	-0.008
				(0.004)	(0.005)	(0.010)
Origtime in 1 CE squared						-0.000
						(0.000)
State Age in 1 CE						0.031
						(0.184)
Constant	-0.425***	-1.469***	-1.507***	-6.369***	-6.193***	-6.262***
	(0.140)	(0.159)	(0.167)	(1.281)	(1.297)	(1.282)
Controls	no	no	no	yes	yes	yes
Continent FE	no	no	no	no	yes	yes
Observations	135	130	130	115	115	115
R-squared	0.154	0.455	0.458	0.717	0.798	0.799

Panel B	Log Population Density in 1500 CE					
	(1)	(2)	(3)	(4)	(5)	(6)
Statehist in 1500 CE	3.883***		0.902	3.653***	3.605***	6.072***
	(0.670)		(0.957)	(0.943)	(0.915)	(1.803)
Agyears in 1500 CE		0.315***	0.269***	0.180***	0.115*	0.169**
		(0.042)	(0.064)	(0.064)	(0.069)	(0.069)
Origtime in 1500 CE				0.006**	-0.003	-0.017
				(0.003)	(0.004)	(0.012)
Origtime in 1500 CE squared						0.000
						(0.000)
State Age in 1500 CE						-0.334*
						(0.180)
Constant	0.359**	-0.411*	-0.345	-2.761*	-3.271**	-3.382***
	(0.144)	(0.212)	(0.223)	(1.487)	(1.343)	(1.258)
Controls	no	no	no	yes	yes	yes
Continent FE	no	no	no	no	yes	yes
Observations	154	147	147	128	128	128
R-squared	0.184	0.269	0.273	0.687	0.759	0.770

Notes: The data for historical population density is based on population data from McEvedy and Jones(1978) and land data from World Bank's *World Development Indicators*. The list of controls includes: absolute latitude, an indicator whether the modern-day country is landlocked, distance to coast and rivers, mean elevation, land suitability, percentage arable land, temperature, precipitation, percentage population at risk of contracting malaria. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table D8. State history and urbanization in 1 CE and 1500 CE – linear relationship

Panel A	Urbanization in 1 CE					
	(1)	(2)	(3)	(4)	(5)	(6)
Statehist in 1 CE	1.128***		0.003	-0.135	0.475	0.170
	(0.237)		(0.232)	(0.360)	(0.313)	(0.728)
Agyears in 1 CE		0.101***	0.101***	0.082***	-0.030	-0.037
		(0.020)	(0.025)	(0.028)	(0.039)	(0.039)
Origtime in 1 CE				0.004*	-0.000	0.000
				(0.002)	(0.002)	(0.004)
Origtime in 1 CE squared						-0.000
						(0.000)
State Age in 1 CE						0.063
						(0.154)
Constant	2.566***	2.348***	2.348***	1.560**	1.260	1.285
	(0.064)	(0.100)	(0.104)	(0.755)	(0.995)	(0.984)
Controls	no	no	no	yes	yes	yes
Continent FE	no	no	no	no	yes	yes
Observations	128	128	128	125	125	125
R-squared	0.063	0.139	0.139	0.371	0.523	0.524

Panel B	Urbanization in 1500 CE					
	(1)	(2)	(3)	(4)	(5)	(6)
Statehist in 1500 CE	16.678***		18.666***	22.878***	21.837***	30.860*
	(2.384)		(3.452)	(6.078)	(7.446)	(16.414)
Agyears in 1500 CE		0.761***	-0.201	-0.135	-0.598	-0.411
		(0.177)	(0.239)	(0.356)	(0.474)	(0.472)
Origtime in 1500 CE				-0.052	-0.070*	-0.141
				(0.036)	(0.041)	(0.131)
Origtime in 1500 CE squared						0.001
						(0.001)
State Age in 1500 CE						-1.256
						(1.693)
Constant	4.487***	3.633***	5.119***	-7.853	-7.314	-8.246
	(0.569)	(1.019)	(1.015)	(8.277)	(8.019)	(8.291)
Controls	no	no	no	yes	yes	yes
Continent FE	no	no	no	no	yes	yes
Observations	83	83	83	76	76	76
R-squared	0.278	0.111	0.282	0.386	0.458	0.467

Notes: The data for urbanization rate in 1 CE is taken from Comin, Easterly and Gong (2010) and is based on Peregrine (2003) and takes three values: 1 if the largest settlement is smaller than 100 persons; 2 if it is between 100 and 399 persons; and 3 if it is larger than 400 persons. The data for urbanization rate at 1500 CE is that reported by Acemoglu, Johnson and Robinson (2005), defined as the percentage of a country's total population residing in urban areas (each with a city population size of at least 5,000). The list of controls includes: absolute latitude, an indicator whether the modern-day country is landlocked, distance to coast and rivers, mean elevation, land suitability, percentage arable land, temperature, precipitation, percentage population at risk of contracting malaria. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table D9. State history and average technology adoption in 1 CE and 1500 CE - linear relationship

Panel A	Technology Adoption in 1 CE					
	(1)	(2)	(3)	(4)	(5)	(6)
Statehist in 1 CE	0.762***		0.106	-0.080	0.093	-0.001
	(0.122)		(0.122)	(0.196)	(0.134)	(0.246)
Agyears in 1 CE		0.064***	0.059***	0.065***	0.028**	0.032*
		(0.007)	(0.010)	(0.013)	(0.014)	(0.016)
Origtime in 1 CE				0.001	-0.002***	-0.005***
				(0.001)	(0.001)	(0.002)
Origtime in 1 CE squared						0.000*
						(0.000)
State Age in 1 CE						0.023
						(0.048)
Constant	0.688***	0.554***	0.559***	-0.071	-0.147	-0.139
	(0.027)	(0.037)	(0.039)	(0.332)	(0.329)	(0.320)
Controls	no	no	no	yes	yes	yes
Continent FE	no	no	no	no	yes	yes
Observations	128	128	128	124	124	124
R-squared	0.149	0.281	0.283	0.541	0.759	0.769

Panel B	Technology Adoption in 1500 CE					
	(1)	(2)	(3)	(4)	(5)	(6)
Statehist in 1500 CE	1.227***		0.288	0.754***	0.684***	1.140***
	(0.157)		(0.193)	(0.172)	(0.135)	(0.353)
Agyears in 1500 CE		0.104***	0.088***	0.042***	-0.000	0.004
		(0.008)	(0.013)	(0.013)	(0.011)	(0.011)
Origtime in 1500 CE				0.001	-0.001**	0.001
				(0.001)	(0.001)	(0.001)
Origtime in 1500 CE squared						-0.000
						(0.000)
State Age in 1500 CE						-0.070
						(0.047)
Constant	0.315***	0.065*	0.090**	-0.143	0.026	-0.062
	(0.030)	(0.036)	(0.037)	(0.258)	(0.170)	(0.180)
Controls	no	no	no	yes	yes	yes
Continent FE	no	no	no	no	yes	yes
Observations	112	110	110	107	107	107
R-squared	0.389	0.532	0.541	0.776	0.897	0.904

Notes: The list of controls includes: absolute latitude, an indicator whether the modern-day country is landlocked, distance to coast and rivers, mean elevation, land suitability, percentage arable land, temperature, precipitation, percentage population at risk of contracting malaria. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table D10. State history, ancestry-adjusted state history and technology adoption in 2000 CE – linear relationship

Panel A	Technology Adoption in 2000 CE					
	(1)	(2)	(3)	(4)	(5)	(6)
Statehist	0.086		-0.074	0.063	0.128	0.289*
	(0.095)		(0.132)	(0.106)	(0.106)	(0.174)
Agyears		0.011	0.015	-0.008	-0.008	-0.000
		(0.007)	(0.010)	(0.008)	(0.011)	(0.012)
Origtime				0.000	0.001**	0.001
				(0.000)	(0.001)	(0.002)
Origtime squared						-0.000
						(0.000)
State Age						-0.032
						(0.021)
Constant	0.430***	0.405***	0.406***	0.761***	0.818***	0.791***
	(0.029)	(0.038)	(0.038)	(0.162)	(0.206)	(0.206)
Controls	no	no	no	yes	yes	yes
Continent FE	no	no	no	no	yes	yes
Observations	130	129	129	125	125	125
R-squared	0.006	0.016	0.018	0.639	0.679	0.684

Panel B	Technology Adoption in 2000 CE					
	(1)	(2)	(3)	(4)	(5)	(6)
Ancestry-Adjusted	0.233**		0.231	0.215*	0.194*	0.354**
Statehist in 1500 CE	(0.092)		(0.156)	(0.121)	(0.115)	(0.161)
Agyears		0.011	-0.000	-0.013	-0.011	-0.003
		(0.007)	(0.012)	(0.009)	(0.012)	(0.012)
Origtime				0.000	0.001**	0.001
				(0.000)	(0.001)	(0.002)
Origtime squared						0.000
						(0.000)
State Age						-0.034*
						(0.018)
Constant	0.399***	0.405***	0.401***	0.754***	0.790***	0.721***
	(0.023)	(0.038)	(0.035)	(0.162)	(0.203)	(0.204)
Controls	no	no	no	yes	yes	yes
Continent FE	no	no	no	no	yes	yes
Observations	130	129	129	125	125	125
R-squared	0.037	0.016	0.036	0.649	0.683	0.690

Notes: The list of controls includes: absolute latitude, an indicator whether the modern-day country is landlocked, distance to coast and rivers, mean elevation, land suitability, percentage arable land, temperature, precipitation, percentage population at risk of contracting malaria. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1