The Geographical Origins of Early State Formation*

Anastasia Litina†

April 30, 2014

Abstract

This research theoretically and empirically advances the hypothesis that in early stages of development land and climatic variability had a persistent beneficial effect on the advent of early statehood. A high degree of diversity, and its association with potential gains from trade accentuated the incentives to develop social, political and physical infrastructure that could facilitate interregional interaction. Hence, the emergence of states, was expedited in more diverse geographical environments. Exploiting exogenous sources of variation in variability in land suitability for agriculture across countries as well as variation in climatic variability within countries over the period 500-1500 CE, the research establishes that: i) the advent of statehood was expedited in regions characterized by a higher degree of variability in agricultural suitability and climatic conditions, ii) the effect of variability on statehood operates through the advancement of medium of exchange and transportation, suggesting that it is the pivotal role of states in facilitating trade that ultimately contributed to their emergence and consolidation, and iii) the effect of land variability on statehood dissipates over time.

Keywords: State formation, Climatic Variability, Land suitability

JEL Classification Numbers: O50, O57, N70

---

*I am grateful to Jean Francois Carpantier, Antonio Cosma, Carl-Johan Daalgard, James Fenske, Oded Galor, Masashiye Hamano, Andreas Irmen, Marc Klemp, Gautham Tripathi, Skerdi Zanaj and participants at the Copenhagen and CREA seminar series for their helpful comments and suggestions. I would also like to thank James Fenske, Stelios Michalopoulos, Omer Ozak and Luis Putterman for generously sharing data. Finally I would like to thank Antonia Margherita for excellent research assistance.

†University of Luxembourg, Faculty of Law, Economics and Finance, 148, Avenue de la Faiencerie, L - 1511 Luxembourg (e-mail: anastasia.litina@uni.lu)

1 Introduction

The origins of statehood and the transition of mankind from hunting-gathering societal structures to that of organized governments has been intensely debated, yet the origins of early states remain largely unexplored in the empirical literature. As the imprint of early states on the economic, cultural and institutional evolution of modern states has been empirically established, tracing the deep rooted factors that have affected the emergence of early states, sheds light on the sources of comparative development across the globe.

This research advances the hypothesis that in early stages of development variability in land suitability for agriculture and climatic variability had a persistent effect on the advent of early statehood. A high degree of variability, being associated with increased benefits from trade, amplified the incentives to develop infrastructure that could facilitate exchanges across regions. Consequently, the emergence of states was expedited in areas where land suitability for agriculture and climatic conditions reflected larger variability.

Differences in comparative advantage in agriculture across regions, driven primarily by variation in land suitability for agriculture and climatic variability, generated incentives for a government to emerge. In particular, the role of such a government was to develop the necessary physical infrastructure (e.g. trade routes) and institutional infrastructure (e.g. protection along trade routes) that would facilitate trade.

The analysis employs a notion of statehood that captures three distinct characteristics: i) whether a state existed or not, ii) whether the government is foreign or locally based and iii) what fraction of the modern territory, was ruled by this government.[1] For instance, one of the oldest states according to the statehood index, is the Kingdom of Ethiopia, which has been ruled by a single domestic kingdom since the year 1 CE and whose great variability of terrain "determines wide variations in climate, soils, natural vegetation, and settlement patterns".[2] Similar examples are that of China which has a long history of statehood and a variable landscape, whereas countries with less variability, such as Iceland, delayed significantly to develop statehood.

[1] The notion of state has been extensively defined in the related literature. While multiple variants exist, central features in all definitions are i) the presence of a territorial base, and, ii) the tiers of decision making. Claessen and Skalník (1978) set several criteria to define an early state, relating state to the notions of territorial unit, citizenship by birth or residence, centralization of power, the maintenance of law and order, and political independence. Ames (2007) gives the characterization of state to regions where the society is stratified, with at least a three-tier decision-making hierarchy, where leadership has both tactical and strategic power and the population usually has a territorial base. According to Peregrine et al. (2007) any centralized polity with three or more levels of decision making above the community is defined as state. Statehood, as defined in the current paper, entails both concepts of the tiers of decision making and of the territorial unit (Chanda and Putterman, 2007).

[2] Wikipedia lemma on Ethiopia.

This research contributes primarily to the literature that explores the origins of statehood. A part of the literature explores the forces behind the centralization of states and the type of institutions that emerged. Fenske (2013) has explored the factors behind the centralization of states in pre-colonial Africa and argued that trade supported class stratification between rulers and ruled, whereas Giuliano and Nunn (2013); Hariri (2012); Lagerlöf (2012) have identified the factors behind the emergence of autocratic versus democratic regimes.

As to the emergence of early states, four main hypotheses about the origins of the state have been advanced thus far. The first suggests that the vast majority of early civilizations developed along rivers. The second highlights the role of food surplus. The third emphasizes the role of population density and the fourth underlies the role of transparency in the production technology and its effect on the ability to extract resources from the population.

The first hypothesis is widely know as sedentism and it argues that the notions of stratification and social complexity became relevant only after populations settled in particular locations (Mann, 1986; Cohen and Yosef, 2002). Different factors can contribute to the emergence of sedentism, such as population density, property rights and scalar stress. Upon the emergence of sedentism and the attachment of the population to a particular location, states gradually emerged as the outcome of surplus, stratification etc.

According to the surplus hypothesis, the availability of surplus expedited the emergence of an elite class sufficiently powerful to extract rents (e.g. via control of labor or debt). The resulting stratification ultimately led to the formation of the state as a means for the elite to maintain and amplify its power (Childe, 1954; Gosden, 1989; Halstead, 1989; Allen, 1997; Arnold, 1993). As to the factors that led to the creation of surplus, a number of theories have been advanced with the prevailing ones being: i) the transition to the Neolithic, that allowed the domestication of grains and other storable crops[3], ii) trade, that allows for the materialization of welfare gains, accruing primarily to the elites via extraction (Bates, 1983; Fenske, 2013), iii) improvements in technology that allowed for increases in productivity, and iv) an exogenous shock (e.g. a climatic shock such as the Middle Ages warm period). The main criticism, suggests that the surplus theory does not take into account the Malthusian mechanism, according to which increases in surplus, particularly the ones driven by increases in productivity, were sufficiently slow thereby allowing the population to catch up (Mayshar and Moav, 2011).[4]

[3]Diamond (1997) argued that the transition to the Neolithic gave rise to the domestication of different crops and to the creation of surplus. In particular, the domestication of crops that can be stored contributed to increased sedentism, higher population densities and have often been used as a prestige good (Testart et al., 1982).

[4]It has been empirically established that in the Malthusian era, increases in productivity were reflected in a larger but not richer population (Ashraf and Galor, 2011; Galor, 2011). Nevertheless, this literature does not provide a link to the emergence of the state. In addition, it could be argued that in the presence of a shock, that generated surplus, population could not adjust instantaneously, thereby allowing the emergence of a state.

According to the hypothesis that features the role of population density, differences in productivity and geographic and climatic conditions, were associated with differences in population density. Higher population density operated via two distinct channels; first it required an increased degree of cooperation and centralized decision making for the society to be functional, and second, it generated population pressure and social unrest, thereby accentuating the need for the emergence of an authority that could mitigate the adverse effect of population pressure via organizing the society and generating surplus (e.g. storage or trade).[5] As already argued, differences in population density can be traced to differences in the natural environment. Alternative determinants of a higher population density could be the transition to the Neolithic[6], improvements in technology (Ashraf and Galor, 2011; Vollrath, 2011) as well as all the channels associated with the creation of surplus.[7]

According to the transparency hypothesis, advanced by Mayshar and Moav (2011), changes in the production technology that are affecting the ability to appropriate a fraction of the output, are associated with the emergence of a ruling class, that generates surplus and ultimately leads to the emergence of the state. Therefore, it is not the existence of surplus that is leading to stratified societies, instead it is the emerging elite that generates the surplus necessary for the emergence of states.

The suggested theory contributes an alternative theory to the emergence of states by advancing the hypothesis that it is the willingness to trade that leads to the emergence of statehood and it highlights the role of regional coordination in constructing the infrastructure necessary to facilitate trade. Moreover, it attributes the incentives for trade to ample geographic and climatic variability. The underlying assumption, is that as long as strong incentives and/or preference for trade exist, a government will emerge with the aim to construct the necessary infrastructure to facilitate and enhance trade. Therefore, in the context of this analysis, the notion of state entails the notion of public good provision (e.g. trade routes). As to the advent of trade, different determinants can be put forward such as preference for variety or as a precautionary mechanism against bad times. It is thus argued that in regions with a larger geographic and climatic variability, and thus with a larger variability in produced goods, the gains from regional trade are higher.

[5] See Carneiro (1994); Kosse (1994).

[6] It has been argued however that the transition to the Neolithic was not reflected in gains in health, thereby suggested a limited effect on population density (Steckel and Rose, 2002).

[7] The fact that in the Malthusian era, any surplus is channelled towards population density, suggests that population density could be categorized as a special case of the surplus theory (Ashraf and Galor, 2011). Nevertheless, it can operate as an independent channel, reflecting differences in productivity across the globe as driven by initial geographical conditions, and is thus reported as a distinct strand of the related literature.

The theory also sheds light on the origins of the contemporary differences in income per capita across the globe.[8] Recently, a number of researchers have traced the imprint of early states on the economic, cultural and institutional evolution of modern states (Chanda and Putterman, 2007). The hypothesis advanced by this research on the hypothesized effect of geography both on the incentives for trade and on its persistent effect on the advent of statehood, can account for differences in comparative development across the globe.

The theoretical model considers two neighboring regions that are identical in all respects except for the type of productivity of land they are endowed with. Driven partly by differences in land productivity, they may produce different goods and they the option either to produce locally and live in autarky or to trade with the neighboring region. Autarky is a viable equilibrium,yet consuming a variety of goods increases individuals' utility. The social planner of each region chooses the optimal volume of trade that maximizes collective utility. If in equilibrium trade emerges, then developing infrastructure that facilitates trade (e.g. trade routes, bridges, etc.) is always cost efficient. Moreover, if the two regions cooperate in the development of infrastructure, the costs can be further reduced for each community whereas they reap the total benefit of the infrastructure. In line with the theory advanced by the paper, the incentive to trade can trigger the development of trade infrastructure that fosters the political cooperation and unification of the two regions and can thus be associated with the emergence of early states.

Exploiting exogenous sources of variation in variability in land suitability for agriculture across countries as well as variation in climatic variability within countries over the period 500-1500 CE, the research establishes that: i) the advent of statehood was expedited in regions characterized by a higher degree of variability in agricultural suitability and climatic conditions, ii) the effect of variability on statehood operates through the advancement of medium of exchange and transportation, suggesting that it is the pivotal role of states in facilitating trade that ultimately contributed to their emergence and consolidation, and iii) the effect of land variability on statehood dissipates over time.

The empirical strategy of this paper is developed in two parts. First, all three hypotheses are tested exploiting variations in land suitability for agriculture in a cross-section of countries. The analysis exploits variations in the variability of land suitability for agriculture, to account for cross-country variations on the emergence of statehood. In examining the hypothesized effect of variability in land suitability, all confounding factors are controlled for, e.g. geographical factors, historical factors (e.g. distance from the nearest technological frontier in the year 1 CE, Roman and Persian occupation), as well as the channels suggested

[8] Important contributions in the literature are Ashraf and Galor (2013); Özak (2012); Andersen, Dalgaard and Selaya (2011); Andersen, Bentzen, Dalgaard and Sharp (2011); Ashraf and Galor (2011b); Galor et al. (2009); Olsson and Hibbs (2005).

in the historical literature.[9]

Second, the reduced form hypothesis (the effect of variability in land suitability on the emergence of states) is further established exploiting variations in climatic conditions using a newly constructed panel dataset and exploiting within country variations. The empirical results are verified in a panel of countries for the period 500-1950 CE. The analysis exploits within country variations, thereby mitigating the concern of omitted variable bias. Additional time varying controls such as the timing of the Neolithic, past state history, and regional trends are taken into account.

The results are robust to a number of robustness tests, related to the validity of the statehood index and the land variability index, to controlling for all other channels through which states could emerge (e.g. the presence of large empires, the creation of surplus, higher population density, earlier transition to the Neolithic, climatic shocks etc.) as well as to tests related to the validity of the estimation (outliers) and omitted variable bias (spatial autocorrelation, regional fixed effects). Moreover, the robustness section uses the measure of fractal dimension of countries, constructed by Alesina et al. (2011) to show that higher variability in land suitability is associated with a higher index of fractal dimension and thus with less "artificial" borders.

A simple model illustrating how trade fosters statehood is developed in Section 2. Section 3 presents the empirical strategy and the data. Section 4 illustrates the empirical results. Section 5 establishes the robustness of the results and Section 6 concludes.

2 The Basic Structure of the Model

Consider two neighboring regions. Both of them operate in the agricultural sector and they are identical in all respects except for the type of productivity of land they are endowed with. Driven partly by differences in land productivity, they may produce different goods and they have two options: i) to produce locally and live in autarky; and ii) to trade with the neighboring region. Whereas autarky is a viable equilibrium, nevertheless consuming a larger variety of goods increases individuals' utility and therefore there is always an incentive to trade provided that the transportation costs are not forbidding.[10] The costs assumed can be related to distance as well as to climatic and geographical conditions.

[9] Appendix C explores analytically some alternative channels associated with the emergence of states suggested by historians. In particular it tests four main hypotheses as to the forces behind the emergence of states: i) sedentism, ii) the surplus hypothesis, iii) the hypothesis that links the emergence of states to the timing of the Neolithic Revolution, and iv) the population density hypothesis.

[10] An alternative option explored below is that of conflict. It is assumed that instead of trade between the two communities, conflict emerges in an attempt to usurp the produced good.

Each region has a social planner who aims at maximizing collective utility via choosing the optimal amount of producing locally and the optimal volume of trade (if any). In the equilibrium where trade is viable, the social planner of each region can reduce the associated costs via developing infrastructure that facilitates trade (e.g. trade routes, bridges, etc.). Moreover, if the two regions cooperate in the development of infrastructure, the costs can be further reduced for each community whereas they reap the total benefit of the infrastructure.[11] In line with the theory advanced by the paper, the incentive to trade can trigger the development of trade infrastructure that fosters the political cooperation and unification of the two regions and can thus be associated with the emergence of early states.

Throughout the paper the analysis will be undertaken under the simplifying assumption that only one of the two regions is an active decision maker (Region i). The second region (Region j) will be assumed to "passively" respond to the actions of the first region.[12] Solving a model where both regions respond simultaneously yields qualitatively similar insights, yet significantly complicates the analysis.

2.1 Economy I: No Political Unification

2.1.1 Social Planner

Region i is run by a social planner who aims at maximizing the collective utility of its inhabitants. Whereas collective action is a strong assumption, nevertheless it allows to focus on the mechanism suggested by the paper, i.e. the role of trade and of trade-related infrastructure. Therefore the model deliberately abstracts from addressing coordination mechanisms.[13]

The aim of the social planner is to maximize the collective utility of the region.[14] The utility function of the region i, U_i is described by

$$U_i = x_i(\varepsilon + x_{ij}) \tag{1}$$

[11] For instance think of a trade route that connects two regions. The two communities can share the cost of building the route, however they can both use to full length of it.

[12] This is not equivalent to assuming that the two regions always cooperate. As a sub-case it will be assumed that when region j is invaded, it will resist the invasion of region i thereby causing destruction, in line with the conflict theory (SKAPERDAS).

[13] The complexity hypothesis (CITE) advances the role of population in fostering the formation of early states. As already mentioned in the literature review, the formation of states was driven by a multitude of alternative channels. Shutting off the complexity channel does not ignore this determinant, it simply allows the analysis to focus on the alternative channel of trade. Moreover it can be assumed that these channels can operate in a complementary fashion, e.g. the complexity hypothesis leads to the first formation of states within the community and once a leader is elected within the community, intra-regional trade can take place and further reinforce states.

[14] Assuming that the social planner is not benevolent and is interested in maximizing his own utility would yield similar qualitative results, provided that he also benefits from variability of consumption goods. The structure would be somewhat different, to capture his rent-seeking attitude, nevertheless his economic incentives would still be in accordance with the incentives to trade and to facilitate trade.

where x_i is a good that is the output of agricultural production in region i and x_{ij}, is a good that is the output of trading activities between country i and country j. ε is a utility parameter, whereas the structure of the utility function captures the fact that the utility of the individuals increases by consuming a variety of goods, i.e. both x_i and x_{ij}. Nevertheless, each region can live in autarky, and therefore we do not impose trade unless it is a profitable option.

2.1.2 Local Production in Regions i and j

Each region produces only one good and for simplicity we assume that it produces only agricultural goods. The good x_i is the outcome of production in region i, Y_i, and is given by the following production function

$$x_i = \theta_i(1-z)Y_i \tag{2}$$

where

$$Y_i = \Gamma_i X_i^a L_i^{1-a} \tag{3}$$

where θ_i is the fraction of the region's population that is allocated to the production of the agricultural good Y_i, produced locally. z is the fraction of output that is required to develop trade infrastructure, if trade emerges as viable option. Γ_i denotes the natural land productivity in region i. Crucially it should not be perceived as productivity associated with technical progress but instead as region-specific productivity intrinsically linked to the climatic and geographic conditions of each region. X_i denotes the land endowment of region i, assumed constant,[15] and L_i is the labor endowment of each individual. For analytical convenience it is assumed that $\Gamma_i = X_i = L_i = 1$ and thus Eqs. (2) and (3) read

$$x_i = \theta_i(1-z)Y_i = \theta_i(1-z) \tag{4}$$

Equivalently it is assumed that the production function of region j is given by

$$Y_j = \Gamma_j X_j^a L_j^{1-a} \tag{5}$$

Γ_j denotes the natural land productivity in region j. X_j denotes the land endowment of region j, assumed constant, and L_j is the labor endowment for each individual. For analytical convenience it is assumed that $X_j = L_j = 1$. Since we do not model the behavior of region j we are not interested in the fraction of labor that is allocated to local production and thus

[15] In the context of this simplifying model, assuming that a region can expand its territory by conquering the neighboring region yields qualitatively similar results. The reason will become apparent in the subsection where the case of usurpation instead of trade is considered.

for simplicity we'll assume that it allocates all its labor force into productive activities. Upon implementing the above assumptions eq. (5) reads as

$$Y_j = \Gamma_j X_j^a L_j^{1-a} = \Gamma_j \tag{6}$$

2.1.3 Trade in Region i

Region i, which is assumed to be the only decision making region, also has a preference for consuming the good x_{ij} which is the outcome of trading activities, if this option is viable for the region. In particular, the equation for good x_{ij} is given by

$$x_{ij} = (1-\theta_i)(1-\omega+\Omega z)Y_j \tag{7}$$

where $(1-\theta_i)$ is the fraction of labor allocated to trading activities. ω is the cost of trading which captures costs associated with trade, e.g. distance of the two communities, geography and climate. Ωz captures the beneficial effect of trade related infrastructure in reducing the costs associated with trade.[16]

2.1.4 Utility Maximization

Combining Eqs. (1), (4), (6) and (2) the aim of the social planner is to maximize the following utility function, by choosing the optimal fraction of labor allocated to each activity, i.e. local production and intra-regional trade, θ_i

$$\max_{\theta_i} U_i = \max_{\theta_i}\ x_i(\varepsilon + x_{ij}) = \max_{\theta_i} \theta_i(1-z)(\varepsilon + (1-\theta_i)(1-\omega+\Omega z)\Gamma_j) \tag{8}$$

The utility maximization yields

$$\theta_i^* = \frac{1}{2}\left[1 + \frac{\varepsilon}{\Gamma_j(1-\omega+\Omega z)}\right] \tag{9}$$

where $0 \leq \theta_i^* \leq 1$. The solution suggests that under certain restrictions there can always be an incentive to trade. Lemma 1 describes the comparative static properties of the optimal solution.

Lemma 1 *(Comparative Statics) i) $\partial\theta_i^*/\partial\Gamma_j < 0$, i.e., the higher the productivity of region j, the higher the incentive for region i to trade with region j; ii) $\partial\theta_i^*/\partial\omega > 0$, i.e., the higher*

[16] For analytical convenience it is assumed that once trade takes place, region j "freely" provides the desired amount of the good Y_j required by region i. This simplifying assumption does not affect the qualitative results, only the magnitude of the results. Moreover it can be implicitly assumed that the terms of trade can be included in the parameter ω, i.e. assuming that ω incorporates the cost of exchange.

the cost of trade, the lower the incentive for region i to trade with region j; iii) $\partial\theta_i^*/\partial\Omega z < 0$, i.e., the better the trade the infrastructure linking the two regions, the stronger the incentive to trade.

Proof. Results (i)-(iii) can be obtained by taking the derivatives of θ_i^* with respect to each parameter. □

Lemma 2 provides the conditions for autarky.

Lemma 2 *(Conditions for Autarky) For an autarkic equilibrium to emerge, i.e. for $\theta_i^* = 1$ the following conditions should be satisfied: i) $\omega = 1+\Omega z$, i.e., if the cost of trade is sufficiently high, region i chooses not to involve in trade activity; ii) $\lim_{i\Gamma_j \to 0} \theta_i^* = 1$, i.e., as the productivity of region j goes to 0, and thus local production goes to 0, there is no trade between the two regions; iii) When the two goods produced are qualitatively the same, i.e. ,when $x_i = x_{ij}$ then $\theta_i^* = 1$, i.e. there is no incentive to trade when both regions produce the same goods.*

Proof. (i) As $\omega \to 1 + \Omega z \implies \theta_i^* \to 1$ (recall that $0 \leq \theta_i^* \leq 1$) and thus no trade takes place; (ii) Similarly, as $\Gamma_j \to 0 \implies \theta_i^* \to 1$, again suggesting that no trade takes place; (iii) For $x_i = x_{ij}$ it must hold that $1/(1-\theta_i^*)(1-\omega+\Omega z) = \Gamma_j$. Replacing Γ_j in Eq. (9) implies that $\theta_i^* = 2 + \varepsilon/1 + \varepsilon > 1$, i.e. no trade takes place if they two goods are the same. □

From Lemmas 1 and 2 three important remarks should be made that are in line with the proposed theory. First, as long as $\Gamma_j \neq \Gamma_i = 1$ there is always an incentive to trade, even if the productivity of region j, Γ_j, is lower than the productivity of region i (provided that the cost of trading is not forbidding, i.e., that $\omega \leq 1 + \Omega z$ and $\frac{\varepsilon}{\Gamma_j(1-\omega+\Omega z)} < 1$).

Second, as long as the two products are differentiated, i.e. $x_i \neq x_{ij}$, there is always an incentive for region i to trade with region j. Last, if trade is a viable option, the better the trade infrastructure the higher the incentive to trade. Naturally causality can ran both ways, however in this paper we seek to establish a causal effect running from the incentive to trade to the development of infrastructure.

Using Eq. (9) we can obtain the optimal level of trade infrastructure, Ωz, for any volume of trade.

Proposition 1 *(The Optimal Level of Trade Infrastructure) The optimal level of trade-related infrastructure, Ωz, that proxies for state formation, is given by*

$$\Omega z = \frac{1}{2}\left[\frac{\varepsilon}{2\theta_i^* - 1}\frac{1}{\Gamma_j} - (1-\omega)\right]$$

Noticeably, $\partial\Omega z/\partial\theta_i^* < 0$ suggesting that the higher the incentive to trade leads to the development of more extensive infrastructure.

In line with the proposed theory, the higher the volume of trade (partly driven by differences in productivity across regions as suggested by Lemma 1), the stronger the incentive to develop trade related infrastructure which can ultimately lead to the emergence of states. Other factors, incorporated in the parameter ω, can also have an effect on state formation (e.g. geography, transportation costs, etc.) and will be analytically explored in the empirical section of the paper.

2.2 Economy II: Political Unification

In this section we will briefly explore the case where a political unification between the two entities is undertaken. A simple way to manifest political unification between the two regions is to assume that they both invest money in developing the trade related infrastructure, however they both have full access to this infrastructure. A evident example would be that of a trade route linking the two regions. Both communities contribute money for the development of the route, thus their costs are cut down by half (in the case of a symmetric equilibrium), however they both have full access to the trade route.

A simplifying way to represent this is to assume that the production function is now given by

$$x_i^{PU} = \theta_i(1 - \frac{z}{2})Y_i \tag{10}$$

where the notation remains the same, with the only difference now being that the cost of developing infrastructure is not cut down by half, whereas the remaining cost in undertake by region j.

Importantly, despite the reduction of the cost, region i still has access to the full length of the road as implied by Eq. (2). Everything else remains the same in the structure of the model. Interesting, replicating the analysis above and maximizing 1 to obtain the optimal θ_i we still obtain the same solution as describe in Eq. (9)

$$\theta_i^{PU*} = \frac{1}{2}\left[1 + \frac{\varepsilon}{\Gamma_j(1 - \omega + \Omega z)}\right] \tag{11}$$

a result consistent with the simplifying structure of the model. Importantly though, if we compare the two utilities we notice that for the same amount of publicly provided trade infrastructure, collective utility is higher in the political unification case than in the case where no unification takes place.[17]

[17] This analysis will be formally undertaken in the last section of the model.

2.3 Economy III: Conflict

Assuming that the two regions cooperate is certainly a strong assumption. Undoubtedly, war has been one of the major drivers of state formation throughout the years, but certainly not the only one (CITE TILLY). The aim of this section is to model the case where region i attacks region j, whereas the other tries to defend itself. The outcome is that whereas the attacking regions can simply usurp the produced good (instead of exchanging it which is preferable from its viewpoint), nevertheless was implies some cost and moreover destruction may occur.

To closely follow the structure of the benchmark model and to make the results directly comparable we choose a very simple way to model war. Analytically the production function in now given by

$$x_i = \theta_i(1 - \tilde{z} - W)Y_i \tag{12}$$

where now there is a new cost added, C, that captures the cost of war (e.g. the cost of preparing for the war, potential destruction, etc.). Importantly, notice that we still assume a cost of developing infrastructure, $\tilde{z}$, since there is always the need to reach region j in order to invade it. Therefore, using Eqs. (12), (3) and $\Gamma_i = X_i = L_i = 1$ we obtain

$$x_i = \theta_i(1 - \tilde{z} - W) \tag{13}$$

Contrary to the benchmark model, x_{ij} is now not the outcome of trade, instead it is the outcome of conflict and usurpation. Therefore, the equation for good x_{ij} is given by

$$x_{ij} = (1 - \theta_i)(1 - \tilde{\omega} + \Omega\tilde{z})Y_j \tag{14}$$

where $(1 - \theta_i)$ is the fraction of labor allocated to conflict activities. $\tilde{\omega} \leq \omega$ is the cost of trading which is now assumed to be less than in the benchmark model. The purpose of this assumption is to capture the fact that the cost of usurping is lower than the cost of trading, since there is no need to exchange good. Moreover the cost of potential destruction and of conflict has already been incorporated in W. $\Omega\tilde{z}$ captures the side benefits of the war infrastructure.

Maximizing (1) using Eqs. (13) and (14) yields

$$\theta_i^{W*} = \frac{1}{2}\left[1 + \frac{\varepsilon}{\Gamma_j(1 - \tilde{\omega} + \Omega\tilde{z})}\right] \tag{15}$$

Comparing the new optimal value of θ_i^{W*} with that of the benchmark case, θ_i^*, does not yield straightforward results as it depends on the relative magnitude of the parameters $\tilde{\omega}$ and $\tilde{z}$. Moreover, it should be noticed that whereas the cost of war does not directly affect the optimal level of trade, an artifact of the simplifying structure of the model, yet it has an effect on the level of utility under any chosen level of trade.

The following section provides some comparisons across the three models.

2.4 Comparisons

In this section we will compare the optimal levels of trade, as emerging from each model, as well as the obtained utilities.

Lemma 3 *(Optimal Trade under Models I, II and III) i)* $\theta_i^* = \theta_i^{PU*}$ *as already highlighted the optimal choice of trade is not affected by the decision about political unification or not; ii)* $\theta_i^* = \theta_i^{PU*} \gtrless \theta$ *if* $-\omega + \Omega z \lesseqgtr -\tilde{\omega} + \Omega \tilde{z}$*, i.e., trade (in both the political unification case and in the no unification case) is preferred to conflict when a) the cost of trade is lower than the cost of war, and b) the trade related infrastructure is less costly related to the war related infrastructure.*

Due to the simplifying structure of the model the optimal choice of producing locally versus trade or war, is not as revealing as the comparison of utilities as implied by the optimal volum of trade in each model. Lemma 4 analytically describes the comparison across models in terms of utilities.

Lemma 4 *(Collective Utility under Models I, II and III) i)* $U_i^* < U_i^{PU*}$*, i.e. the utility in the no unification case (Model I) is lower than the utility in the political unification case (Model II) despite the fact that in both equilibria the optimal level of trade chosen is the same. The driver of this result is that the cost of developing infrastructure is split among the two regions, whereas the benefit is fully reaped by each region. ii)* $U_i^* \gtrless U_i^{W*}$ *and* $U_i^{PU*} \gtrless U_i^{W*}$ *depending on the parameter values of* $\tilde{\omega}, \omega$, $\tilde{z}$, z *and* X*, i.e., war may be preferred to trade under both the unification and the non-unification cases, depending on the relative costs of trade, infrastructure and conflict.*

The purpose of Lemma 4 is to underline that other forces, such as conflict, can be the drivers of state formation and consolidation. However, it also emphasizes that under plausible assumptions, trade and trade infrastructure may naturally emerge across regions as the optimal solution and lead to the formation of early states via the need to develop infrastructure and the unification of neighboring regions.

3 Empirical Strategy and Data

3.1 Empirical Strategy

First, the analysis explores all three hypotheses exploiting variations in land suitability for agriculture in a cross-section of countries. Second, the reduced form hypothesis is further established exploiting variations in climatic conditions using a panel dataset and exploiting within country variations.

3.1.1 The Effect of Variability in Land Suitability on Statehood: Cross Country Analysis

To establish the main hypothesis advanced by the paper, i.e. that geography has a persistent effect on the imminence of statehood, the analysis exploits variations in the variability of land suitability for agriculture, to account for cross-country variations in the emergence of states. In particular the analysis focuses on the index of statehood in the year 1000 CE, yet the robustness section establishes that the argument is valid for other historical periods as well. In examining the hypothesized effect of variability in land suitability on the emergence of states, the full set of controls is employed, i.e. geographical factors, distance from the nearest technological frontier in the year 1 CE and a fixed effect for Roman and Persian occupation. Moreover, the analysis controls for all the competing channels, as described in the first part of the empirical analysis, i.e. proximity to waterways, suitability of land for agriculture and suitability for storable crops (capturing the surplus hypothesis), years elapsed since the onset of the Neolithic transition and population density in the year 1 CE.

Analytically, the baseline regression specification employed to test the effect of variability in land suitability on statehood, using a sample of 117 countries, is described by

$$I_i = \alpha_0 + \alpha_1 V_i + \alpha_2 X_i + \alpha_3 \Delta_i + \varepsilon_i \tag{16}$$

where the subscript i is a country indicator; I_i is an index of statehood for the year 1000 CE[18]; V_i is an index of variability in land suitability; X_i is a vector of geographical and historical controls; Δ_i is a dummy variable for continental fixed effects and ε_i is a country specific error term.

Variability in Land Suitability and Statehood: The Trade Channel The third part of the empirical analysis, establishes that the main channel through which land variability reinforces the formation of states is via trade. The identifying assumption is that higher land variability is associated with stronger incentives to trade, and therefore the need for

[18] The construction of the index is described in the data section.

a state that will facilitate trade and will undertake the provision of the necessary infrastructure is more exigent. To capture the mediating factor of trade, the analysis employs a "horse race" regression between the measure of variability in land suitability and two proxies of trade. More analytically, these proxies are a) transportation in the year 1 CE, and b) medium of exchange in the year 1 CE. According to the theory, higher land variability provided more incentives for trade, incentives that were materialized into more sophisticated medium of exchange and medium of transportation. Interestingly, the results of the "horse race" regression, indicate that it is the proxies of trade that survive, despite the fact that the measure of variability in land suitability is more precisely measured.

Moreover to address the possibility that the relationship between statehood and the proxies of trade is spurious, driven by a third unobservable factor, and to mitigate reverse causality concerns, the analysis employed the proxies of trade in the year 1 CE and explores their effect on statehood in the year 1000 CE. This approach also underlines the persistent effect of trade on state formation.

In particular, the baseline regression specification used to test the channel of trade, is described by

$$I_{SH} = \alpha_0 + \alpha T_i + \alpha_1 V_i + \alpha_2 X_i + \alpha_3 \Delta_i + \varepsilon_i \tag{17}$$

where the subscript i is a country indicator. I_i is an index of statehood for the year 1000 CE[19]; T_i is a proxy for trade in the year 1 CE; V_i is an index of variability in land suitability; X_i is a vector of geographical and historical controls; Δ_i is a dummy variable for continental fixed effects and ε_i is a country specific error term.

The Effect of Variability on Statehood over Time This approach establishes that variability in land suitability was critical in generating states as long as the main mode of trade was intra-regional trade. In the process of development and as economies expanded their trade possibilities via long-distance and transatlantic trade, land variability became gradually less important in giving rise to states. To capture this effect this section adopts two alternative approaches. First it exploits variations in land variability across countries to establish that whereas variability was a significant determinant of states in the year 1000 CE as well as in the year 1500 CE, nevertheless it has no effect on the emergence of states in 1950 CE. In this approach, the analysis controls for a lagged value of statehood so as to distinguish whether variability in land has a direct effect on the emergence of states or whether it operates only via past statehood. Second, it employs a panel data approach where variability in land suitability is interacted with a time variable. Whereas the actual effect of land variability cannot be

[19] The construction of the index is described in the data section.

identified, as it is time invariant, however it is feasible to identify how this effect evolves over time while controlling for country fixed effect.

3.1.2 Climatic Variability and Statehood: Panel Analysis for the Period 500 CE - 2000 CE

This approach explores the intertemporal effect of climatic variability on statehood, employing a time varying measure of climatic variability available for the period 500-1500 CE.[20] Importantly this approach addresses the issue of potential omitted variable bias since it allows identifying the effect of climatic variability exploiting within country variation. Moreover climatic variability, for the era under examination, is completely exogenous and not prone to human intervention, a fact that allows to establish a casual effect.

The suggested measure differs from the measure of variability in land suitability for agriculture in that it is just aimed to capture the average temperature at the grid level. However, for reasons extensively analyzed in the data section, it is plausibly employed as a good proxy for variability in land suitability for agriculture.

In order to smooth out the effect of potential fluctuations of the climatic conditions, the measure of climatic variability is aggregated for every 500 year interval.[21] This approach also allows to capture climatic shocks that spanned over a large number of years, e.g. the Little Ice Age or the Medieval Warm Period. Similarly, the statehood index as constructed by Chanda and Putterman (2007), is aggregated for every 50 year interval. Moreover, the measure of statehood adopted is the one that assumes no discounting of past statehood.[22] Importantly, this ensures that the observations are independent to past observations.[23]

Analytically, the baseline regression specification employed to test the effect of variability in land suitability on statehood, using a sample of 93 countries, is described by

$$I_{it} = \alpha_0 + \alpha_1 C_{it} + \alpha_2 I_{it-1} + \alpha_3 X_{it} + \alpha 4 \Delta_i + \alpha_5 T_{it} + \alpha_6 T_t R_i + \varepsilon_{it} \tag{18}$$

where the subscript i is a country indicator and t is a time indicator (four intervals are used, i.e. $t1$ = 500 CE, $t2$ = 1000 CE, $t3$ = 1500 CE, $t4$ = 1950 CE); I_{it} is an index of statehood for each period t;[24] C_{it} is an index of climatic variability; $I_{i,t-1}$ is the index of state history in the period $t-1$; X_{it} is a vector of time varying variables across countries; Δ_i is a country dummy variable; T_t is a time dummy; $T_t R_i$ is an interactive term of the time period

[20] More details on the construction of the dataset are available in section on the data.

[21] Following Ashraf and Michalopoulos (2013) the average temperature is calculated at the grid level and then the measure is aggrefated at the country level

[22] A more lengthy discussion on discounting can be found in the data description section as well as on the robustness section and in particular on the analysis of Table 9.

[23] The results are robust to the use of a 50 year average of climatic variability.

[24] The construction of the index is described in the data section.

interacted with regional dummies, aimed to capture time varying factors at the continent level and ε_{it} is a country and time specific error term.

3.2 The Data

3.2.1 Dependent Variable

Statehood The statehood variable is the "State Antiquity" index developed and used by Chanda and Putterman (2007). It is a composite index capturing not only the existence or not of a state, but also the intensity of statehood. In particular it is a composite index, that is a multiple of three components:

$$I_{SH} = I_G \text{x} I_{FL} \text{x} I_T$$

where each component takes a value based on the related answer. More analytically, the questions addressed are, i) $I_G \equiv$ Is there a government above the tribal level?; ii) $I_{FL} \equiv$ Is this government foreign or locally based? and iii) $I_T \equiv$ What is the fraction of the modern territory ruled by this government?

The values are assigned as follows:

i) $I_G \equiv$ Is there a government above the tribal level?

$$\left\{ \begin{array}{c} \text{Yes} \\ \text{No} \end{array} \right\} \Longrightarrow I_G = \left\{ \begin{array}{c} 1 \\ 0 \end{array} \right\}$$

ii) $I_{FL} \equiv$ Is this government foreign or locally based?

$$\left\{ \begin{array}{l} \text{Foreign [e.g. colony]} \\ \text{Hybrid (local with foreign oversight)} \\ \text{Local} \end{array} \right\} \Longrightarrow I_{FL} = \left\{ \begin{array}{c} 0.5 \\ 0.75 \\ 1 \end{array} \right\}$$

and iii) $I_T \equiv$ Fraction of the modern territory, θ_T, ruled by this government

$$\theta_T \in \left\{ \begin{array}{c} [0, 0.1] \\ (0.1, 0.25] \\ (0.25, 0.5] \\ (0.5, 1] \end{array} \right\} \Longrightarrow I_T = \left\{ \begin{array}{c} 0.3 \\ 0.5 \\ 0.75 \\ 1 \end{array} \right\}$$

Therefore, as suggested by the construction of the index, the measure of statehood does not capture only the existence or non-existence of an autonomous state, but it gives a broader perspective on the capacity of statehood, reflecting the administrative structure and the level of autonomy. More importantly, given that all the variables in the analysis are employing countries as defined by their current borders, the state antiquity index is capturing this shortcoming by including the fraction of the modern territory was ruled by the government.

The combination of these three elements is particularly important, since it allows to trace the structure and the political organization of each region at any time period. Since this is an era where many forces are affecting the emergence of states this approach is highly useful since it allows to capture a wide range of forces that are operating at the same time.[25]

Another important element of the index, is that it captures past state history. In particular the index is constructed for all intervals of 50 years starting from the year 1 CE till the year 1950 CE. Whereas each interval is constructed without taking into account past history, all the years employed in this analysis (e.g. state history in 1000 CE, 1500 CE and 1950 CE) take into account all preceding intervals by discounting each interval at a rate of 5% per interval. Therefore this ensures that the index at each point in time can capture past history, an approach that is very useful to smooth out the effect of an exogenous and temporary shock on a state, that has otherwise existed for many years.[26]

The following table gives an example of values for the statehood index in the year 1000 CE:

	Ethiopia	China	Iceland
State history in 1000 CE	1	0.85	0

As described in Chanda and Putterman (2007), the value of 1 for Ethiopia captures the fact that Ethiopia had always been ruled by a domestic kingdom since the year 1 CE till the year 1700 CE. The value of 0.85 for China (in the year 1000 CE), reflects the fact that there has been a collapse of the centralized rule of T'ang dynasty, which led to the emergence of several, locally based, domestic kingdoms and dynasties. Finally, the value of 0 for Iceland, reflects the fact that in the year 1000 CE there is no political structure above the tribal level.

3.2.2 Independent Variables

Variability in Land Suitability Variability in land suitability is the standard deviation of land suitability for agriculture, i.e. the average deviation from the mean of average suitability for agriculture.

[25] An example would be the city of Venice, which despite the fact that it was part of the Roman Empire and later of the Kingdom of Lombardy, nevertheless from the 900 CE till 1200 CE it developed into a city state primarily due to the autonomy implied by its geographical position and the strong naval and commercial power that it had developed. Therefore, despite the fact that major forces determined the statehood status of Venice, nevertheless the role of geography is clear in determining the level of autonomy and the intensity of political integration.

[26] Crucially though the results are robust to the use of an index that is constructed without reflecting past history as well as to the use of different discount rates.

The land suitability measure is an index of the average suitability of land for cultivation, based on geospatial data on various ecological factors, related to climatic factors and soil quality. These factors include (i) growing degree days, (ii) the ratio of potential to actual evapotranspiration, (iii) soil carbon density, and (iv) soil pH. Therefore biophysical factors, such as topography and irrigation, and socioeconomic factors such as market price or incentive structure, which are important for determining whether land will be cultivated, are not part of the index.[27] The index is reported at a half-degree resolution by Ramankutty et al. (2002). The average of land quality is thus the average value of the index across the grid cells within a country. This measure is obtained from Michalopoulos (2012).

To replicate the examples of statehood mentioned above, the corresponding land variability indices are:

	Ethiopia	China	Iceland
State history in 1000 CE	1	0.85	0
Variability in Land Suitability	0.99	0.99	0.03

which suggests that highly variable countries in terms of land productivity, developed early states, whereas countries such as Iceland, with very low variability had not developed states by that time.

As has been lengthily explored in Litina (2013), one potential source of concern with respect to the measure of land suitability is whether current data on the suitability of land for cultivation reflect land suitability in the past. Importantly, the critical aspect of the data for the tested hypothesis is the ranking of countries with respect to their variability in land suitability as opposed to the actual measure of variability in land suitability. Hence the identifying assumption is that the ranking of countries as measured today, reflects the ranking in the past.

If intense cultivation and human intervention affected soil quality over time, this could have affected all countries proportionally and therefore it would introduce a non-systematic error. This would not only leave the ranking of countries with respect to variability in land suitability for agriculture unaffected, but would also enhance the difficulty to detect a significant effect on land suitability. Importantly, even in the presence of a systematic error, it would be implausible to argue that the ranking of countries has been reversed, based on two remarks similar to the ones made by Michalopoulos (2012). First, one of the two components of the index is based upon climatic conditions, which have not significantly changed during the period of examination.[28] Therefore, even if the characteristics of soil quality have significantly

[27] The argument for adopting such an approach is based upon the observation that at the global scale, climate and soil factors form the major constraints on cultivation, and adequately describe the major patterns of agricultural land (Ramankutty et al., 2002),

[28] Durante (2010) has examined at the relationship between climatic conditions for the years 1900-2000 and 1500-1900. In particular, he uses measures for average precipitation, average temperature,

changed over time, this would still have a limited effect on the total index of land suitability. Second, given that the measure of land suitability captures the average level of land suitability within a given country, it would be implausible to anticipate that deteriorations in land quality in particular segments of the country, could affect the average land quality of a country, to the extent that it would change its overall ranking.

To further alleviate potential concerns about the importance of the effect of human intervention on soil quality, two strategies are adopted: i) an alternative measure of dispersion of land suitability is employed, namely the range of land suitability for agriculture; and ii) the baseline regressions are repeated using each component of the variability in land suitability index separately, namely variability in climatic suitability and in soil suitability, with the emphasis laid on the climatic component.[29]

During this period three major climatic changes have occurred that could potentially affect statehood: i) the Medieval warm period (950 CE-1250 CE), ii) the Little Ice Age (1350 CE-1850 CE) and iii) droughts (500 CE-1500 CE). Given that these climatic "shocks" were not the result of human intervention, it would be plausible to argue that they did not systematically affect the index of climatic suitability. In addition, since these shock affected each country homogeneously, it would be plausible to argue that whereas a climatic shock of this type would affect the mean, albeit it would leave the dispersion unaffected. To further mitigate these concerns two additional robustness tests are made: i) a dummy for each shock is employed, and ii) given that climatic shocks affected each latitude differentially, the baseline analysis controls not only for absolute latitude but also for latitudinal zones, thereby mitigating any concerns associated with systematic errors within each zone. Reassuringly, the results remain largely intact.

Climatic Variability The measure of climatic variability is provided at the grid level by Mann et al. (2009), who have reconstructed surface temperature patterns over the interval 500 CE-2000 CE. To produce their reconstructions they employ a global proxy dataset that comprises treering, ice core, coral, sediment and other assorted proxy records. Whereas this dataset is only a rough approximation of actual climatic conditions throughout this era, nevertheless as the authors emphasize, it is the longer-term and larger-scale variations resolved by the reconstructions that are most meaningful. Therefore, for the purpose of this analysis, the data are generated at a scale as large as the country level and intervals of 500 years are employed, thereby rendering the analysis meaningful.[30]

---

precipitation variability and temperature variability. His findings confirm that regions with more variable climate in the present years were also characterized by more variate climate in the past, thereby reassuringly implying that climatic conditions have not significantly changed over time.

[29] This approach is reassuring since evidence suggests that climate has not severely changed during the last 2000 years (Durante, 2010).

[30] Extensive robustness analysis as to how precise the measurement of average temperature throughout this period is, has been conducted by the authors themselves in their paper.

Using GIS software, the climatic conditions data are constructed for a panel of 97 countries, for which the data is available, covering four time periods from 500 CE till 1950 CE. The measure of interest is climatic variability as reflected by the range of climatic variations within a country.

The suggested measure differs from the measure of variability in land suitability for agriculture in that it is just aimed to capture the average temperature at the grid level, not climatic factors conducive to agriculture. Nevertheless, temperature is a critical determinant of agricultural productivity. Importantly, as already argued in the data section, the index of suitability of land for agriculture is decomposed into a climatic and a soil component, and as is established in the robustness sections all the results are valid even when only the climatic component is employed. Moreover, the climatic component is viewed as a more robust proxy since it is less vulnerable to human intervention.

Therefore, it is argued in this section that variability in temperature is a good proxy for the climatic component of land suitability for agriculture and thus for land suitability as such. Particularly in a context where climate variations are intended to capture the incentive to trade and/or risk sharing attitudes, the use of this proxy is quite meaningful. Reassuringly the climatic component of the baseline measure of land suitability for agriculture, despite capturing only variations in temperature from the mean, is highly and positively correlated with the current measure of variability in average temperature across the globe (the correlation coefficient is higher than 0.4).

Proxies of Trade Data on a) transportation in the year 1 and b) medium of exchange in the year 1 are constructed from Peregrine's (2003) *Atlas of Cultural Evolution*, and aggregated at the country level by Ashraf and Galor (2011). Each of these three sectors is reported on a 3-point scale, as evaluated by various anthropological and historical sources. The level of technology in each sector is indexed as follows. In the transportation sector, the index is assigned a value of 0 under the absence of both vehicles and pack or draft animals, a value of 1 under the presence of only pack or draft animals, and a value of 2 under the presence of both. In the medium of exchange sector, the index is assigned a value of 0 under the absence of domestically used articles and currency, a value of one under the presence of only domestically used articles and the value of 2 under the presence of both. In all cases, the sector-specific indices are normalized to assume values in the [0,1]-interval. Given that the cross-sectional unit of observation in Peregrine's dataset is an archaeological tradition or culture, specific to a given region on the global map, and since spatial delineations in Peregrine's dataset do not necessarily correspond to contemporary international borders, the culture-specific technology index in a given year is aggregated to the country level by averaging across those cultures from Peregrine's map that appear within the modern borders of a given country.

4 Empirical Findings

4.1 Variability in Land Suitability and Statehood: Cross Country Analysis

The analysis in the preceding section suggested that, in line with historical evidence, the imminence of early states has been triggered by a number of factors. This section explores the hypothesized effect of variability in land suitability for agriculture on statehood. In particular, it first establishes that higher variability in land suitability for agriculture had a positive and persistent effect on the emergence of states. Second it establishes that this effect operates through the mediating channel of trade. Last, it establishes that the positive effect of variability diminishes over time.

4.1.1 The Effect of Variability in Land Suitability on the Emergence of States

The reduced form hypothesis advances the role of variability in land suitability on the advent of statehood. Exploiting variations in variability in land suitability in a sample of 117 countries, Column (1) of Table 1 establishes a positive and highly significant effect of variability of land for agriculture on the emergence of early states, while controlling for continental fixed effects. Column (2) introduces a number of exogenous geographical controls that capture the sedentism and the surplus hypothesis. The highly significant coefficient on variability in land suitability confirms the robustness of the hypothesis. Column (3) introduces historical controls such as distance to the nearest technological frontier in the year 1000 CE and a fixed effect for Persian and Roman occupation. Column (4) introduces additional channels that have already been explored earlier, starting the timing of the Neolithic, whereas Column (5) introduces population density in the year 1 CE. Reassuringly, the coefficient on variability, remains rather stable, decreasing somewhat in magnitude once all alternative channels are controlled for. Therefore the findings in Table 1 are confirming that variability in land suitability had a direct effect on the onset of statehood.

4.1.2 The Channel of Trade

This section uncovers the mechanism through which variability in land suitability affects the advent of early states. In particular, the idea that will be explored is that variability in land suitability generates more incentives for trade across regions. The desire for trade and the associated benefits, render the emergence of a state imperative, as a means to facilitate trade via undertaking the creation of the necessary social and institutional infrastructure.

Table 1: Main Hypothesis: The Effect of Land Variability on the Emregence of States

	(1)	(2)	(3)	(4)	(5)
		Dep. Var.:State History in 1000 CE			
Variability in Land Suitability	0.192**	0.291***	0.258***	0.255***	0.217**
	(0.096)	(0.083)	(0.092)	(0.093)	(0.096)
Average Land Suitability		0.680***	0.584**	0.596**	0.461*
		(0.228)	(0.240)	(0.245)	(0.259)
Fraction of Arable Land		-0.346	-0.291	-0.327	-0.479*
		(0.242)	(0.272)	(0.270)	(0.266)
Suitability for Cereals		-0.087***	-0.086***	-0.084**	-0.050
		(0.032)	(0.033)	(0.033)	(0.038)
Distance to the Nearest Coast/River		-0.222***	-0.198**	-0.191**	-0.182**
		(0.084)	(0.085)	(0.086)	(0.087)
% of Land within 100 km of Coast/River		-0.030	-0.005	-0.004	-0.054
		(0.116)	(0.116)	(0.117)	(0.118)
Average Ruggedness		-0.000	-0.000	-0.000	-0.000
		(0.000)	(0.000)	(0.000)	(0.000)
Average Elevation		0.048	0.056	0.047	0.038
		(0.082)	(0.081)	(0.082)	(0.082)
Total Area		14.598	16.392	15.730	20.844*
		(10.962)	(10.970)	(11.113)	(11.119)
Absolute Latitude		-0.007**	-0.006*	-0.006	-0.005
		(0.004)	(0.003)	(0.004)	(0.004)
% Land in Tropical and Subtropical Zones		-0.416***	-0.330***	-0.316**	-0.345**
		(0.105)	(0.124)	(0.128)	(0.135)
% Land Temperate Zones		-0.212*	-0.158	-0.160	-0.134
		(0.122)	(0.123)	(0.124)	(0.124)
Distance to Frontier in 1000 CE			-0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)
Years since Neolithic				0.014	-0.010
				(0.021)	(0.021)
Population Density in the Year 1 CE					0.064**
					(0.026)
Continental Dummies	Yes	Yes	Yes	Yes	Yes
Persian-Roman Occupation	No	No	Yes	Yes	Yes
Observations	117	117	117	117	117
R-squared	0.398	0.612	0.631	0.632	0.649

Summary: This table establishes that variability in land suitability has a direct effect on the emergence of states. The significant effect of land variability is established while controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

Notes: (i) Land suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate and soil suitability for cultivation; (ii) land suitability diversity is the range of the land suitability index; (iii) the set of continent dummies includes a fixed effect for Africa, the Americas, Australia, Europe and the Middle-East (iv) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (v) robust standard error estimates are reported in parentheses; (vi) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

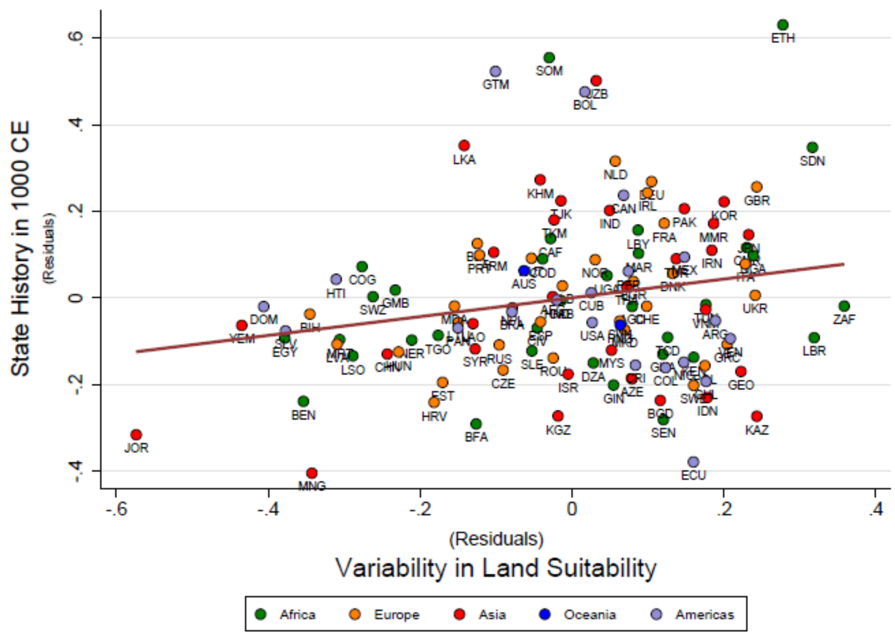


FIGURE 1: Land Variability Hypothesis-Conditional on controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

More analytically, Table 2 employs two proxies of trade: i) the level of sophistication of the medium of exchange in the year 1 CE, and ii) the level of sophistication of the means of transportation in the year 1 CE. In employing these measures an implicit assumption is made, i.e. that the incentives and the need for trade resulted in higher levels of sophistication for technologies associated with trade.

Given the potential endogeneity between the state index and the trade proxies, the analysis will exploit variations in the measures of trade for the year 1 CE, so as to mitigate the problem of reverse causality. Moreover, this approach accords with the historical persistence advanced by this research as the mechanism behind the emergence of state. In particular it is argued that variability in land suitability, persistently generated increased benefits from trade, leading the geographically diverse societies into a continuous effort to improve the means and terms of trade and through this process, leading ultimately to the emergence of states.

Column (1) replicates the results in Column (5) of Table 1, so as to facilitate comparisons. Column (2) establishes that, consistently with the predictions of the paper, higher variability in land suitability is associated with more advanced transportation technology in the year 1 CE, via providing more incentives for trade. In Column (3), once this measure of trade (transportation technology) is introduced in the baseline regression (i.e. Column (1)), both the significance and the magnitude of the coefficient of land variability drop completely, despite the fact that the proxy for trade is imprecisely measured. This suggests that the effect of land variability operated through the trade incentives it generates.

Similarly, Column (4) establishes the significant and positive effect of land variability on another proxy for trade, namely the level of sophistication of the means of exchange. Column (5) introduced this second measure of trade in the baseline analysis (e.g. Column (1)), and the coefficient of land variability drops completely both in magnitude and significance.

Overall, the analysis of this section suggests that one of the mechanisms via which land variability affects statehood is trade. More mechanisms conducive to states can be hypothesized, complementary to the role of trade. The robustness section provides a brief overview of alternative mechanisms.

4.1.3 The Effect of Variability on Statehood over Time

This section establishes that variability in land suitability was critical in generating states as long as the main mode of trade was inter-regional trade. In the process of development and as economies expanded their trade possibilities via long-distance and transatlantic trade, land variability became gradually less important in giving rise to states. To capture this effect this section adopts two alternative approaches. First it exploits variations in land variability across countries to establish that whereas variability was a significant determinant of states in the

Table 2: Mediating Factor: Testing Whether the Effect of Variablity Operates via Trade

	(1)	(2)	(3)	(4)	(5)
	State 1000 CE	Exch.1CE	State.1000CE	Transp. 1CE	State 1000 CE
Var. in Land Suit.	0.217**	0.512***	0.104	0.313***	0.112
	(0.096)	(0.153)	(0.088)	(0.114)	(0.085)
Med. Exch. in 1 CE			0.221***		
			(0.075)		
Med. Transp. in 1 CE					0.337***
					(0.119)
Cont. Dummies	Yes	Yes	Yes	Yes	Yes
Persi.-Rom. Occup.	Yes	Yes	Yes	Yes	Yes
Observations	117	117	117	117	117
R-squared	0.649	0.656	0.681	0.803	0.687

Summary: This table establishes that variability in land suitability has a direct effect on the emergence of states. The significant effect of land variability is established while controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

Notes: (i) Land suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate and soil suitability for cultivation; (ii) land suitability diversity is the range of the land suitability index; (iii) the indices of" Transportation in the Year 1" and "Medium of Exchange in the Year 1" are technology indices aggregated at the country level. Each of these two sectors is reported on a 3-point scale, as evaluated by various anthropological and historical sources; (iv) the set of continent dummies includes a fixed effect for Africa, the Americas, Australia, Europe and the Middle-East (v) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (vi) robust standard error estimates are reported in parentheses; (vii) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

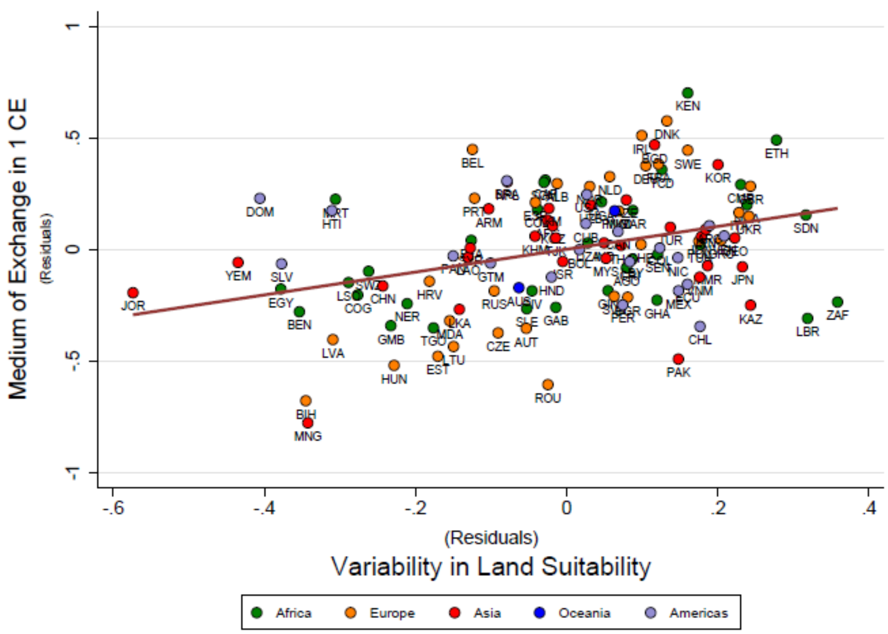


FIGURE 2: The effect of land variability on the medium of exchange technology-Conditional on controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

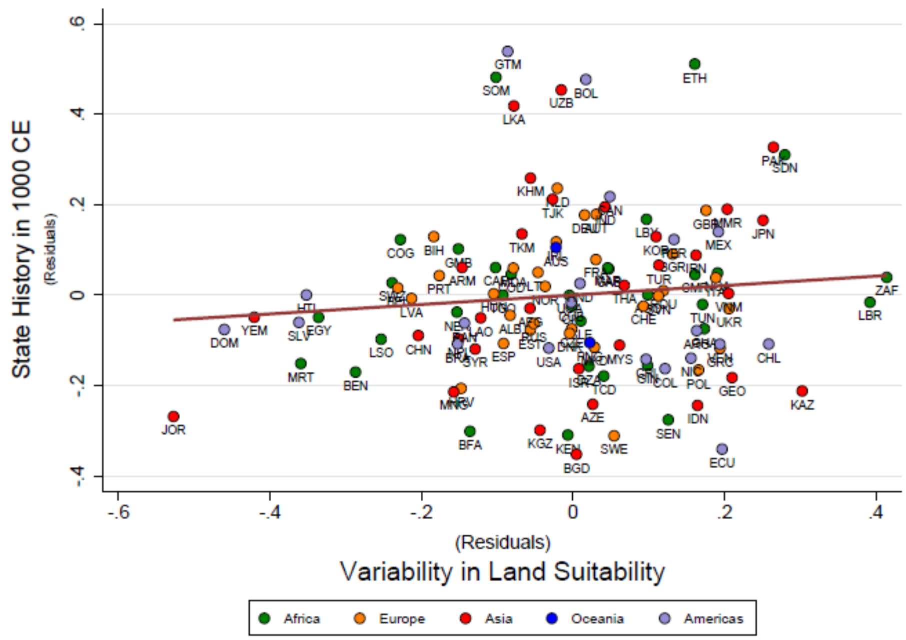


FIGURE 3: The channel of trade-Conditional on controlling for medium of exchange technology (the channel), various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

year 1000 CE as well as in the year 1500 CE, nevertheless it has no effect on the emergence of states in 1950 CE. In this approach, the analysis controls for a lagged value of statehood so as to distinguish whether variability in land has a direct effect on the emergence of states or whether it operates only via past statehood. Second, it employs a panel dataset by interacting variability in land suitability with a time variable so as to capture the effect of variability over time. The benefit of this approach is that it allows to control for country fixed effects, taking care for most of the unobserved heterogeneity.

More analytically, Table 3 employs the sample of 117 countries, that has been used in the baseline analysis and includes the full set of controls. Column (1) replicates the results in Column (5) of Table 1. Column (2) introduces as a control an index of statehood in the year 500 CE. Reassuringly, whereas the coefficient of statehood in 500 CE is highly significant and positively correlated with statehood in the year 1000 CE, nevertheless the coefficient on the variability of land remains significant reducing somewhat in magnitude. This result suggests that while inertia from past statehood is an important determinant, yet variability in land persistently affects the rise of states.

Column (3) employs the same sample of 117 countries, however it uses as dependent variable the index of statehood in the year 1500 CE. The coefficient on the variability of land suitability suggests that even in the year 1500 CE, land variability still confers a significant effect on statehood, while controlling for the full set of controls. When controlling for inertia from past statehood in Column (4), by introducing the index of statehood in the year 1000 CE, the effect of land variability reduces both in magnitude and significance, yet it is still detectable at conventional significance levels. As anticipated, inertia from past statehood is crucial as the coefficient of statehood in 1000 CE suggests.

Column (5) repeats the baseline analysis while using as an explanatory variable an index of statehood in the year 1950 CE while controlling for the full set of controls. Whereas the coefficient of variability in land is significant at the 5% level, yet this effects vanishes once the analysis controls for statehood in the year 1500 CE. The result suggests that land affects modern states only via its effect on past statehood, an outcome that is further reinforced by the highly significant coefficient associated with statehood in 1500 CE.

Overall Table 3 establishes that whereas variability was a crucial determinant for the rise of states, yet this effect dissipated over time.

A second approach that allows to capture the intertemporal effect of land variability on statehood, is to employ a panel dataset. Whereas the index of statehood is varying over time, as well as some historical controls such as population density and distance to the nearest technological frontier, yet the index of land variability is time invariant. This implies that one cannot identify the actual effect of land variability on statehood, however, by interacting the variability index with time it is possible to capture the evolution of this effect. Table 4 adopts

Table 3: Inertia from Past Statehood

	(1)	(2)	(3)	(4)	(5)	(6)
	State Hist. 1000 CE	State Hist. 1000 CE	State Hist. 1500 CE	State Hist. 1500 CE	State Hist. 1950 CE	State Hist. 1950 CE
Variability in Land Suit	0.217**	0.133**	0.238**	0.106**	0.205**	0.030
	(0.096)	(0.055)	(0.094)	(0.050)	(0.080)	(0.032)
State History 500 CE		0.012***				
		(0.001)				
State History 1000 CE				0.714***		
				(0.057)		
State History 1500 CE						0.734***
						(0.041)
Continental Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Persian-Roman Occup.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	117	117	117	117	117	117
R-squared	0.649	0.884	0.742	0.927	0.707	0.934

Summary: This table establishes that the effect of variability in land suitability dissipates over time. Whereas land variability is a crucial and independent determinant of statehood in the years 1000 CE and 15000, yet its effect on statehood in the year 1950 CE operates only through past statehood. The analysis controls for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

Notes: (i) Land suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate and soil suitability for cultivation; (ii) land suitability diversity is the range of the land suitability index; (iii) the set of continent dummies includes a fixed effect for Africa, the Americas, Australia, Europe and the Middle-East (iv) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (v) robust standard error estimates are reported in parentheses; (vi) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

this approach by employing a panel dataset of the 117 countries used in the baseline analysis for two periods, the year 1 CE and the year 1000 CE for which the full set of time variant controls is available. Column (1) in Table 4 regresses the index of statehood on the interactive term between time and variability in land suitability for agriculture, while controlling for country fixed effects. The negative and significant coefficient of the interactive term suggests that the effect of land variability on the emergence of states diminishes over time. The result remains robust to the inclusion of a control for population density and distance to the nearest technological frontier as suggested by evidence in Column (2). Column (3) explores whether the effect of land variability operates via trade as suggested by the baseline analysis. It thus introduces both the index for medium of exchange as well as the interaction between medium of exchange and time. The fact that the coefficient associated with land variability interacted with time drops, suggests that indeed the effect of variability on statehood operates through trade. Moreover, the positive and significant coefficient associated with the medium of exchange suggests that trade has a positive effect on statehood, yet this effect diminishes over time as suggested by the interactive term. Column (4) confirms these findings by employing as a proxy for trade the means of transportation.

4.2 Variability in Land Suitability and Statehood: Panel Analysis using Climatic Variability during the Period 500 CE - 2000 CE

This section addresses the issue of omitted variable bias employing a new dataset on climatic variability available at the grid level from Mann et al. (2009).

The suggested measure is aimed to capture the average temperature at the grid level and not climatic factors conducive to agriculture. Nevertheless, it is a good proxy for land suitability for agriculture since variations in temperature can capture the incentive to trade.

Column (1) Table 5 establishes that climatic variability has a positive and significant effect on statehood. The result is obtained without using any controls, i.e. the correlation is established based on cross country variation. Column (2) controls for country fixed effect, thereby taking care of any unobserved heterogeneity associated with time invariant country specific factors. Reassuringly, the positive and highly significant coefficient on climatic variability, which increased in magnitude, suggests that even when exploiting within country variations in climatic variability, it still confers a significant effect on the emergence of state thereby lending credence to the reduced for hypothesis established in the cross country analysis.

Column (3) introduces a dummy for each period $t2 = 1000$ CE, $t3 = 1500$ CE and $t4 = 1950$, in order to capture time specific shocks. The coefficient reduces in magnitude however it is still statistically significant at the 1% level. Column (4) adds more time varying

Table 4: The Intertemporal Effect of Land Variability on Statehood: A Panel Data Analysis

	(1)	(2)	(3)	(4)
		State History		
Variability in Land Suit X Time	-7.666***	-7.414***	-1.077	-0.693
	(0.929)	(1.067)	(2.315)	(2.368)
Population Density		-0.101	0.008	-0.034
		(0.208)	(0.755)	(0.746)
Distance to the Frontier		-0.004**	–0.001	-.001
		(0.002)	(0.003)	(0.003)
Medium of Exchange			25.112***	
			(8.157)	
Med. of Exch X Time			-19.066***	
			(3.019)	
Medium of Transportation				32.667***
				(11.938)
Med. of Trans X Time				-19.723***
				(2.993)
Country Fixed Effects	Yes	Yes	Yes	Yes
Persian-Roman Occup.	Yes	Yes	Yes	Yes
No. of Countries	117	117	117	117
R-squared	0.447	0.457	0.705	0.708

Summary: This table employs a panel data approach to establish that the effect of variability in land suitability dissipates over time. The analysis employs the interactive term between time and land variability while controlling for country fixed effects, population density and distance to the frontier.

Notes: (i) Land suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate and soil suitability for cultivation; (ii) land suitability diversity is the range of the land suitability index; (iii) the analysis employs two time periods, the year 1 CE and the year 1000 CE; (v) robust standard error estimates are reported in parentheses; (v) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

controls in the analysis, i.e. an interaction of regional dummies (where regions refer to Asia, Africa, Europe, Americas and Oceania similarly to the baseline analysis) in an attempt to capture time varying factors associated with the continent. The results remain relatively unaffected.

Column (5) controls for the mean level of temperature. The coefficient on the mean level of temperature is negative and significant suggesting that higher average temperature is associated with less statehood, potentially capturing the adverse effect of good climatic conditions on the need to cooperate and coordinate activities. Importantly, the coefficient on climatic variability is still highly significant and relatively stable. Column (6) controls for inertia from past state history, in order to ensure that the effect of climatic variability does not operate via the effect on past statehood. Reassuringly, the results remain intact both in terms of magnitude and of significance of the coefficient. Finally, Column (7) introduces in the analysis a control for the years elapsed since the onset of the Neolithic.[31] The timing of the Neolithic does not have a significant effect on statehood, whereas the coefficient on climatic variability remains unaffected.

Overall, the findings of this section suggest that climatic variability had a persistent effect of early states. Critically, by controlling for country fixed effects as well as for a number of time varying controls, it ensures any potential issue of omitted variable bias has been addressed. Moreover, the fact that climate was rather unaffected by human activity for the period under examination, is reassuring as to the causality of the effect.

5 Robustness

This section establishes the robustness of the results. In particular, it explores the validity of the index of diversity by subjecting the index to a number of tests and employing alternative indexes; it addresses the issue of potential omitted heterogeneity; it tests for spatial autocorrelation, and it establishes the validity of the estimation. Finally it extends the analysis beyond the year 1000 CE and it explores the persistence of the channel using a measure of state history in different time periods.

[31] The measure employed is the logarithm of the years elapsed since the Neolithic. The reason for not employing the same measure as in the baseline specification is because in its non-logarithmic form the measure is a linear transformation from one year to the other and thus drops from the analysis. Including it in a non-linear format allows the panel data analysis to capture the potential effect of the timing of the Neolithic.

Table 5: The Effect of Climatic Variability on State Formation

	(1)	(2)	(3)	(4)	(5)	(6)
				State History		
Clim. Var.	10.814***	15.466***	10.577***	9.749***	9.741***	9.846***
	(2.430)	(2.937)	(3.353)	(3.078)	(3.124)	(3.153)
Mean Temp.				-16.590***	-16.550***	-15.779***
				(5.878)	(5.893)	(5.904)
Lag State History					-0.039	-0.048
					(0.060)	(0.057)
Log Years Since Neol.						530.602
						(553.948)
Country Dum.	No	Yes	Yes	Yes	Yes	Yes
Time Dummies	No	Yes	Yes	Yes	Yes	Yes
Time X Reg. Dum.	No	No	Yes	Yes	Yes	Yes
Observations	372	372	372	372	372	372
Countries	93	93	93	93	93	93
R-squared	0.186	0.393	0.483	0.499	0.500	0.501

Summary: This table exploits within country variation in climatic variability to establish its significant effect on statehood. The analysis expands over 1500 years for a set of 93 countries and controls for country fixed effects, time fixed effects, regional fixed effects interacted with time dummies, average temperature, one period lag of state history and years elapsed since the onset of the Neolithic.

Notes: (i) Mean temperature is the average temperature of the country for each interval of 500 years. Following Ashraf and Michalopoulos (Ashraf and Michalopoulos, 2013), the average temperature is calculated at the grid level and then the measure is aggregated at the country level; (ii) climatic variability is the range of the average temperature measure; (iii) the set of time dummies includes a fixed effect for the years t1=500 CE, t2=1000 CE, t3=1500 CE and t4=1950 CE; (iv) the set of regional dummies interacted with time includes a fixed effect for Africa, the Americas, Australia, Europe and Oceania (v) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (vi) robust standard error estimates are reported in parentheses; (vii) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

Table 6: Robustness to Regional Fixed Effects

	(1)	(2)	(3)	(4)	(5)
	State 1000 CE	Exch 1CE	State.1000CE	Transp. 1CE	State 1000 CE
Var. in Land Suit.	0.174*	0.544***	0.039	0.316***	0.051
	(0.102)	(0.152)	(0.093)	(0.116)	(0.095)
Med. Exch. in 1 CE			0.248***		
			(0.075)		
Med. Transp. in 1 CE					0.388***
					(0.124)
Cont. Dummies	Yes	Yes	Yes	Yes	Yes
Pers.-Rom. Occup.	Yes	Yes	Yes	Yes	Yes
Observations	117	117	117	117	117
R-squared	0.623	0.671	0.661	0.807	0.672

Summary: This table establishes the robustness of the results to the use of alternative regional fixed effects. The significant effect of the land variability index is established while controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

Notes: (i) Land suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate and soil suitability for cultivation; (ii) land suitability diversity is the range of the land suitability index; (iii) the indices of "Transportation in the Year 1" and "Medium of Exchange in the Year 1" are technology indices aggregated at the country level. Each of these two sectors is reported on a 3-point scale, as evaluated by various anthropological and historical sources; (iv) the set of continent dummies includes a fixed effect for Latin America and the Caribbean, Sub-Saharan Africa, East Asia and Pacific Region, Europe and Central Asia, Middle East and North Africa and South Asia ; (v) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (vi) robust standard error estimates are reported in parentheses; (vii) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

5.1 Unobserved Heterogeneity and Spatial Autocorrelation

An attempt to deal with specific unobservables is already undertaken in the baseline regressions by including continental fixed effects, as well as by employing a proxy for climate which allowed to control for country fixed effects. Therefore all the results are robust to the fixed effects specification. An alternative attempt to capture unobserved heterogeneity, is to use regional fixed effects instead of continental fixed effects. The fixed effects that have been used are regional dummies for (i) Sub-Saharan Africa (ii) Middle East and North Africa, (iii) Europe and Central Asia, (iv) South Asia, (v) East Asia and Pacific and (vi) Latin America and the Caribbean. The results are robust to this specification as well (see Table 6).

5.1.1 Bilateral Approach

To generalize the results of the baseline analysis and to control for a larger number of unobservable characteristics and country fixed effects, this approach employs as the dependent variable absolute differences in statehood between pairs of countries, regressed on absolute differences in the variability of land suitability for agriculture, controlling for country fixed effects and a number of controls that capture differences between countries.

As Spolaore and Wacziarg (2009) mention this approach allows to make a more efficient use of a wealth of bilateral data as regressors. In particular the model to be estimated is:

$$|\Delta I_{ij}| = \alpha_0 + \alpha_1 |\Delta V_{ij}| + \alpha_2 |\Delta X_{ij}| + \alpha_3 \delta_i + \alpha_4 \delta_j + \varepsilon_{ij} \quad (19)$$

where the subscript i and j are country indicators; $|\Delta I_{ij}|$ is the absolute difference in the statehood index for the year 1000 CE between country i and j; $|\Delta V_{ij}|$ denotes absolute differences in the index of variability in land suitability; $|\Delta X_{ij}|$ denotes differences in a vector of geographical and historical controls; δ_i and δ_j are dummy variables for countries i and j respectively ε_{ij} is a pair specific error term.

The analysis in this section bares similarities to a cross-country gravity model, with the major difference being that observations of pairs comprising the same combination of countries (e.g. USA-UK and UK-USA), are symmetric. The sample features 13572 observations, constructed by using pairs of the same group of 117 countries (Nx$(N-1)$ pairs) used in the baseline analysis. Since the sample is symmetric, to avoid underestimating standard errors, they are clustered at the pair level.[32]

As Spolaore and Wacziarg (2009) mention, spatial correlation may result as the outcome of the construction of the dependent variable. To resolve this issue they follow the approach of Cameron et al. (2011) and rely on two-way clustering of the standard errors on the dimension of country i and country j. Their estimator allows for an arbitrary correlation between errors belonging to the same group thereby being applicable for cases where spatial correlation is a potential concern. The results of this approach are also robust to two-way clustering as well.

The controls employed in this section are the major set of controls that have been employed in the baseline analysis, i.e. various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic. Importantly though, all these controls in the current analysis reflect differences between countries. Moreover the analysis controls for contiguity (countries sharing a common

[32] An alternative approach would be to drop half the sample, by keeping only one side of the pair, but this would imply additional robustness controls to ensure the symmetry of the sample. Reassuring though, the results are robust to this specification as well.

border), distance between capitals and country fixed effects for both countries i and j that belong to each pair.

More analytically, Column (1) in Table 7 establishes the positive and significant effect of differences in land variability on differences in statehood while controlling for country i and j fixed effects. Column (2) introduces into the analysis differences in all the geographical and historical controls introduced in the baseline analysis. Column (3) introduces controls that are more applicable to bilateral models, i.e. a dummy variable for countries sharing a common border as well as distances between capitals. Reassuringly, the highly significant and positive coefficient suggests that the results of the main analysis are confirmed even after controlling for a larger number of unobservable characteristics.

The last two columns explore whether this effect of variability in land suitability operates via the scope for trade. The analysis thus conducts a horserace regression between variability in land suitability and medium of exchange in the year 1 CE (Column 4) and medium of transportation in the year 1 CE (Column 5). The fact that in both columns the coefficient associated with variability in land suitability reduces both in magnitude and significance suggests that the emergence of states operates via the scope for trade, confirming the findings of the baseline analysis.

Overall, the findings of this section shed more light on the main hypothesis of the paper. Even in the context of pairwise correlations, which allow for direct comparison between pairs of countries while exploiting a wealth of bilateral controls and the use of country i and j fixed effects, the results are reassuring as to the fact that differences in land variability can account for differences on the emergence of organized states.

5.1.2 Spatial Autocorrelation

Given the possibility that the disturbance terms in the baseline regression models may be non-spherical in nature, particularly since economic development has been spatially clustered in certain regions of the world, the robustness of the results has been tested from repeating the baseline analysis, with the standard errors of the point estimates corrected for spatial autocorrelation following the methodology of Timothy G. Conley (results not reported).

5.2 Validity of the State History Index

This section establishes the validity of the state history index. In particular it uses three alternative approaches. First, it employs a measure of state history for the year 1500 CE and replicates all the results of Table 2 employing the full set of relevant for the period controls. Second, it employs the measure of statehood for the years 1000 CE and 1500 CE without

Table 7: Bilateral Analysis: The Effect of Land Variability on the Emergence of States

	(1)	(2)	(3)	(4)	(5)
	Dep. Var.:State History in 1000CE				
Diffs in Variability in Land Suitability	0.049**	0.039**	0.037**	0.021	0.031*
	(0.021)	(0.018)	(0.018)	(0.017)	(0.017)
Diffs in Med. Exch. in 1 CE				0.183***	
				(0.008)	
Diffs. in Med. Transp. in 1 CE					0.212***
					(0.009)
Country i and j Dummies	Yes	Yes	Yes	Yes	Yes
Diffs in Geographical Controls	No	Yes	Yes	Yes	Yes
Historical Controls	No	Yes	Yes	Yes	Yes
Distance Between Capitals	No	No	Yes	Yes	Yes
Common Border	No	No	Yes	Yes	Yes
Observations	13572	13572	13572	13572	13572
Countries	117	117	117	117	117
R-squared	0.152	0.351	0.353	0.409	0.402

Summary: This table establishes that differences in variability in land suitability have a direct and independent effect on differences in the emergence of states. The significant effect of land variability is established while controlling for pairwise differences between countries on various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic, contiguity (common borders between catenaries), distances between capitals and fixed effects for country i and country j.
Notes: (i) Land suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate and soil suitability for cultivation; (ii) land suitability diversity is the range of the land suitability index; (iii) differences between countries reflect absolute differences; (iv) robust standard error estimates clustered at the pair of countries levels are reported in parentheses; (v) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

discounting for past state history. Last, it employs only one of the three elements of the index, namely the existence or not of statehood and replicates the analysis in Table 2.

Table 8 replicates the analysis using the index of statehood for the year 1500 CE. Column (1) regresses statehood in 1500 CE over the full set of controls. Notably, whereas all geographical controls remain the same, yet the measure of population density is in the year 1000 CE (to mitigate endogeneity issues) and distance to the nearest technological frontier is in the year 1500 CE. The coefficient on land variability suggests that it confers a significant effect on statehood in the year 1500 CE. Column (2) establishes that, consistently with the predictions of the paper, higher variability in land suitability is associated with a more advanced transportation technology in the year 1000 CE, due to providing more incentives for trade. In Column (3), once the measure of trade (transportation technology) is introduced in the baseline regression (i.e. Column (1)), both the significance and the magnitude of the coefficient of variability in land drop completely, despite the fact that the proxy for trade is imprecisely measured. The fact that it is the proxy for trade that survives, suggests that the effect of land variability operated through the trade incentives it generates.

Similarly, Column (4) establishes the significant and positive effect of land variability on another proxy for trade, namely the level of sophistication of the means of exchange in the year 1000 CE. Column (5) introduces this second measure of trade in the baseline analysis (e.g. Column (1)), and the coefficient of land variability drops completely both in magnitude and significance.

Second, Table 9 employs the non-discounted measure of statehood, i.e. a measure that does not capture inertia from past history. The results remain quite similar with the only difference being that the coefficient on statehood does not vanish completely once the trade measures are introduced. Yet the magnitude of the coefficient drops significantly. This is consistent, since the effect of trade proxies in the year 1 CE should have a cumulative effect on statehood in the year 1000 CE, through its persistent effect on past statehood.

Last, Table 10 employs only the first component of the index, i.e. a binary variable that indicates whether a state exists or not. The results are reassuring, since they suggest that even the existence of a state or not, is driven by variability in land suitability (see Column 1). Moreover, the remaining columns confirm that the effect of land variability operates via trade. However, the fact that the coefficient drops only in magnitude, potentially suggests that the degree of variability confers an important effect on the intensity and the extend of statehood as well, aspects of which are not captured by this binary index.

Table 8: The Effect of Land Variability on State History in the Year 1500 CE

	(1)	(2)	(3)	(4)	(5)
	State 1500	Exch.1000	State 1500	Transp. 1000	State 1500
Var. in Land Suit.	0.245**	0.187*	0.199**	0.213*	0.193**
	(0.098)	(0.111)	(0.083)	(0.112)	(0.082)
Med. Exch. in 1 CE			0.246**		
			(0.122)		
Med. Transp. in 1 CE					0.245**
					(0.121)
Cont. Dummies	Yes	Yes	Yes	Yes	Yes
Persi.-Rom. Occup.	Yes	Yes	Yes	Yes	Yes
Observations	125	125	125	125	125
R-squared	0.745	0.846	0.765	0.844	0.765

Summary: This table establishes that variability in land suitability has a direct and independent effect on the emergence of states in the year 1500 CE. The significant effect of land variability is established while controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

Notes: (i) Land suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate and soil suitability for cultivation; (ii) land suitability diversity is the range of the land suitability index; (iii) the indices of "Transportation in the Year 1" and "Medium of Exchange in the Year 1" are technology indices aggregated at the country level. Each of these two sectors is reported on a 3-point scale, as evaluated by various anthropological and historical sources; (iv) the set of continent dummies includes a fixed effect for Africa, the Americas, Australia, Europe and the Middle-East (v) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (vi) robust standard error estimates are reported in parentheses; (vii) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

Table 9: Mediating Factor: Testing Whether the Effect of Variablity Operates via Trade

	(1)	(2)	(3)	(4)	(5)
	State 1000 CE	Exch.1CE	State.1000CE	Transp. 1CE	State 1000 CE
Var. in Land Suit.	0.245**	0.187*	0.199**	0.213*	0.193**
	(0.098)	(0.111)	(0.083)	(0.112)	(0.082)
Med. Exch. in 1 CE			0.246**		
			(0.122)		
Med. Transp. in 1 CE					0.245**
					(0.121)
Cont. Dummies	Yes	Yes	Yes	Yes	Yes
Pers.-Rom. Occup.	Yes	Yes	Yes	Yes	Yes
Observations	0.745	0.846	0.765	0.844	0.765
R-squared	53.998	38.046	59.101	39.033	58.939

Summary: This table establishes that variability in land suitability has a direct and independent effect on the emergence of states. The significant effect of land variability is established while controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

Notes: (i) Land suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate and soil suitability for cultivation; (ii) land suitability diversity is the range of the land suitability index; (iii) the indices of "Transportation in the Year 1" and "Medium of Exchange in the Year 1" are technology indices aggregated at the country level. Each of these two sectors is reported on a 3-point scale, as evaluated by various anthropological and historical sources; (iv) the set of continent dummies includes a fixed effect for Africa, the Americas, Australia, Europe and the Middle-East (v) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (vi) robust standard error estimates are reported in parentheses; (vii) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

Table 10: Employing the non-Discounted Measure of State History

	(1)	(2)	(3)	(4)	(5)
	State 1000	Exch.1	State.1000	Transp. 1	State 1000
Var. in Land Suit.	0.489***	0.512***	0.352***	0.313***	0.369***
	(0.141)	(0.153)	(0.127)	(0.114)	(0.130)
Med. Exch. in 1 CE			0.267**		
			(0.114)		
Med. Transp. in 1 CE					0.383**
					(0.188)
Cont. Dummies	Yes	Yes	Yes	Yes	Yes
Pers.-Rom. Occup.	Yes	Yes	Yes	Yes	Yes
Observations	117	117	117	117	117
R-squared	0.703	0.656	0.723	0.803	0.724

Summary: This table establishes the robustness of the results when employing as the dependent variable a binary index that indicates the existence of a state or not. The significant effect of land variability is established while controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

Notes: (i) The measure of statehood employed in this table, is a binary variable that denotes the existence of a state or not, and it is a component of the composite state index used in the baseline analysis; (ii) land suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate and soil suitability for cultivation; (iii) land suitability diversity is the range of the land suitability index; (iv) the indices of "Transportation in the Year 1" and "Medium of Exchange in the Year 1" are technology indices aggregated at the country level. Each of these two sectors is reported on a 3-point scale, as evaluated by various anthropological and historical sources; (v) the set of continent dummies includes a fixed effect for Africa, the Americas, Australia, Europe and the Middle-East (vi) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (vii) robust standard error estimates are reported in parentheses; (viii) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

5.3 Artificial States

This section of the robustness explores an alternative proxy of statehood. In particular it employs the index of fractal dimension of each country, constructed by Alesina et al. (2011). As described in their research, this measures reflects how straight (and thus most likely artificial) or squiggly (and thus more likely natural) are the borders of a country. A fractal index of dimension one would denote a straight line, whereas an index of dimension two would capture a plane (and thus very squiggly borders).

Whereas this measure is not highly correlated with the measures of statehood, nevertheless it can be plausible inferred that the fractal dimension index of a country can implicitly capture how naturally the borders emerged.

This is precisely the reason why this section adopts this measure. While many artificial borders, particularly the ones in the African continent, are the outcome of colonization and political competition between western countries, yet the role of geography is critical in determining the natural borders of the country, affecting also indirectly the effectiveness of political decisions on a country that is primarily shaped into a state as driven by geography. For instance modern Ethiopia, which has been an independent kingdom for thousands of years and is also rather variable in terms of land suitability, has a fractal dimension index of 1.01 which is higher than that of other African countries that were artificially split.

Table 11 employs a sample of 128 countries, for which the full set of controls is available and establishes that higher diversity in land suitability is associated with a larger index of fractal dimension, thereby suggesting that the natural forces of a variable geography played a catalytic role in shaping current borders. In particular, Column (1) regresses the basic index of fractal dimension on variability in land suitability[33], while controlling for all the relevant controls (geographical and historical controls for the year 1500 CE which is the nearest date to modern borders) as well as a set of controls for colonial origins which is highly important for the drawing of artificial borders. The positive and highly significant coefficient associated with fractal dimension suggests that netting out the potential effect of all other geographical and political forces that could shape current borders, variability in land suitability plays an important role in determining the current shape of borders. Column (2) replicates the same analysis while using as a robustness an alternative measure of fractal dimension using 10 box sizes, while Column (3) uses a measure of fractal dimension using 9 box sizes. The results are robust to all three specifications highlighting the important role of land variability and reinforcing that it has a critical role in shaping countries borders and giving rise to states.

[33]Fractal dimension of the country's political (non-coastline) borders using all 12 box sizes.

Table 11: Employing a Binary Index of State History

	(1)	(2)	(3)
		Log Fractal Dimension	
Log Variability in Land Suitability	0.006***	0.003**	0.002*
	(0.002)	(0.001)	(0.001)
	(0.005)	(0.004)	(0.003)
Continental Dummies	Yes	Yes	Yes
Persian-Roman Occupation	Yes	Yes	Yes
Colonial Dummies	Yes	Yes	Yes
Observations	128	128	128
R-squared	0.660	0.634	0.665

Summary: This table establishes that variability in land suitability has a positive and significant effect on the index of fractal dimension of countries (i.e. less variable countries have more straight borders). The significant effect of land variability is established while controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

Notes: (i) The index of fractal dimension captures how straight are the borders of a country. A country whose borders are a straight line (artificial country) has an index of value 0. The more squiggly the borders, the higher the value of the fractal dimension index; (ii) land suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate and soil suitability for cultivation; (iii) land suitability diversity is the range of the land suitability index; (iv) the set of continent dummies includes a fixed effect for Africa, the Americas, Australia, Europe and the Middle-East (v) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (vi) robust standard error estimates are reported in parentheses; (vii) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

5.4 Validity of the Index of Diversity in Land Suitability

A potential source of concern with respect to the measure of variability in land suitability, is whether the index of variability as measured today is representative of the index as early as in the year 1000 CE. In particular there are two major sources of concerns, one associated with climatic changes that have occurred throughout this period, and the other being associated with the effect of human intervention on the index.

A number of arguments have been employed in Section 3.2.2 to address these concerns, suggesting that the identifying assumption, i.e. that the ranking of variability in land suitability as measured today reflects the ranking of land suitability in the past, is plausible.

Nevertheless, to further alleviate concerns about the effect of human intervention or climatic changes, the baseline analysis is repeated using each component of the land suitability index separately, namely variability in climatic suitability and variability in soil suitability.[34] Column (1) of Table 12 establishes the effect of variability in climatic suitability on statehood in the year 1000 CE whereas Columns (2) and (4) explore the mediating factor of trade, by augmenting the analysis with the two proxies of trade (medium of exchange and means of transportation respectively). Table 13 repeats the same analysis using the measure of variability in soil quality. Both tables in all columns employ the full set of controls. Reassuringly the results in both cases remain intact, which reinforces the validity of the index.[35]

A second robustness test is to employ an alternative measure of variability in land suitability. More analytically, a Gini index of land suitability, originally constructed by Michalopoulos (2012), is employed. As evidence in Table 14 suggests, repeating the baseline analysis using this alternative measure leaves the results unaffected.

An alternative test that ensures the validity of the index and the immunity of the results to the potential effect of climatic shocks, is to employ a dummy for each major climatic shock recorded the last 2000 years., i.e. the 'Little Ice Age' (1350 CE-1850 CE), the 'Medieval

[34] Soil suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of soil suitability for cultivation, such as soil carbon density and soil pH whereas climatic suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate suitability for cultivation such as growing degree days and the ratio of actual to potential evapotranspiration. Diversity in each of these indexes is capture by a measure of standard deviation.

[35] According to Durante (2010), who has explored the relationship between climatic conditions for the years 1900-2000 and 1500-1900, the regions with more variable climate in the present years were also characterized by more variate climate in the past. Reassuringly, this suggests that the measure of diversity in climatic suitability as measured today, reflects diversity in climatic suitability in the past. Moreover any climatic shock affected regions homogeneously thereby suggested that while a climatic shock may have affected the mean it could have a much less pronounced effect on the standard deviation of the index of climatic suitability for agriculture.

Table 12: Decomposition of the Index of Diversity: Climatic Component

	(1)	(2)	(3)	(4)	(5)
	State 1000	Exch.1	State.1000	Transp. 1	State 1000
Var. in Land Suit.	0.117*	0.185*	0.074	0.148**	0.065
	(0.068)	(0.099)	(0.063)	(0.072)	(0.065)
Med. Exch. in 1 CE			0.231***		
			(0.074)		
Med. Transp. in 1 CE					0.349***
					(0.120)
Cont. Dummies	Yes	Yes	Yes	Yes	Yes
Pers.-Rom. Occup.	Yes	Yes	Yes	Yes	Yes
Observations	117	117	117	117	117
R-squared	0.643	0.622	0.682	0.793	0.686

Summary: This table establishes that variability in climatic suitability has a direct and independent effect on the emergence of states. The significant effect of climatic variability is established while controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

Notes: (i) Climatic suitability for agriculture is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate suitability for cultivation, such as growing degree days and the ratio of actual to potential evapotranspiration; (ii) variability in climatic is the range of the climatic suitability index; (iii) the indices of "Transportation in the Year 1" and "Medium of Exchange in the Year 1" are technology indices aggregated at the country level. Each of these two sectors is reported on a 3-point scale, as evaluated by various anthropological and historical sources; (iv) the set of continent dummies includes a fixed effect for Africa, the Americas, Australia, Europe and the Middle-East (v) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (vi) robust standard error estimates are reported in parentheses; (vii) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

Table 13: Decomposition of the Index of Diversity: Climatic Component

	(1)	(2)	(3)	(4)	(5)
	State 1000	Exch.1	State 1000	Transp. 1	State 1000
Var. in Land Suit.	0.289**	0.652***	0.150	0.475***	0.133
	(0.112)	(0.156)	(0.102)	(0.111)	(0.101)
Med. Exch. in 1 CE			0.213***		
			(0.075)		
Med. Transp. in 1 CE					0.327***
					(0.123)
Cont. Dummies	Yes	Yes	Yes	Yes	Yes
Pers.-Rom. Occup.	Yes	Yes	Yes	Yes	Yes
Observations	117	117	117	117	117
R-squared	0.653	0.663	0.682	0.815	0.687

Summary: This table establishes that variability in soil suitability has a direct and independent effect on the emergence of states. The significant effect of climatic variability is established while controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

Notes: (i) Soil suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of soil suitability for cultivation, such as soil carbon density and soil pH; (ii) variability in soil suitability is the range of the soil suitability index; (iii) the indices of "Transportation in the Year 1" and "Medium of Exchange in the Year 1" are technology indices aggregated at the country level. Each of these two sectors is reported on a 3-point scale, as evaluated by various anthropological and historical sources; (iv) the set of continent dummies includes a fixed effect for Africa, the Americas, Australia, Europe and the Middle-East (v) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (vi) robust standard error estimates are reported in parentheses; (vii) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

TABLE 14: Robustness to Alternative Measures of Land Variability

	(1)	(2)	(3)	(4)	(5)
	State 1000	Exch 1	State 1000	Transp. 1	State 1000
Var. in Land Suit.	0.501**	1.203***	0.240	0.927***	0.191
	(0.209)	(0.284)	(0.219)	(0.210)	(0.209)
Med. Exch. in 1 CE			0.216***		
			(0.078)		
Med. Transp. in 1 CE					0.334***
					(0.125)
Cont. Dummies	Yes	Yes	Yes	Yes	Yes
Pers.-Rom. Occup.	Yes	Yes	Yes	Yes	Yes
Observations	117	117	117	117	117
R-squared	0.652	0.665	0.681	0.821	0.686

Summary: This table establishes the robustness of the results to the use of an alternative measure of land variability (land Gini index).The significant effect of the land Gini index is established while controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

Notes: (i) Land suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate and soil suitability for cultivation; (ii) land suitability diversity is a Gini index of the land suitability index; (iii) the indices of "Transportation in the Year 1" and "Medium of Exchange in the Year 1" are technology indices aggregated at the country level. Each of these two sectors is reported on a 3-point scale, as evaluated by various anthropological and historical sources; (iv) the set of continent dummies includes a fixed effect for Africa, the Americas, Australia, Europe and the Middle-East (v) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (vi) robust standard error estimates are reported in parentheses; (vii) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

Warm Period' (950 CE-1250 CE) and the droughts (500 CE-1500 CE). Table 15 replicates the baseline analysis while introducing fixed effects for all three climatic shocks. To construct these dummies a variety of resources have been employed.[36] The explanatory variable in this table is the composite measure of variability in land suitability while employing the full set of controls. Reassuringly, whereas the coefficient reduces somewhat in magnitude, yet the results remain largely unaffected.

Overall, the results of this section, establish the validity of the index on variability and suggest that the identifying assumption, i.e. that the ranking of variability in land suitability as measured today reflects the ranking of land suitability in the past, is plausible.

5.5 Validity of the Estimation

This section establishes that the main results are not driven by outliers. In Table 16 the baseline regressions are repeated employing the full set of controls, while weighting influential observations in the sample. The choice of influential observations is made by using Cook's D measure of influence.[37] Reassuringly all the results are robust to weighting influential observations.

5.6 Discussion of Alternative Channels

The main purpose of this paper is to establish a casual effect from climatic and land variability on the presence of states. A number of empirical tests have suggested that such an effect exists and is robust to a series of tests. A second argument advanced in the paper is that one of the mechanisms through which land and climatic variability is affecting states is that of trade. Correlations provided in the paper suggest that indeed trade is one of the mechanisms affecting the emergence of states. However a number of alternative mechanism could be plausibly linked to both the climatic variability and the emergence of state. Whereas the aim of the paper is to explore the trade mechanism, yet a discussion of alternative mechanisms is rather informative.

Risk Sharing The nexus between climatic variability and risk sharing mechanisms has been explored in the economics and sociology literature (Durante, 2010; Berg, 2007). Higher climatic variability, associated with higher uncertainty and risk for agricultural output, has historically been mitigated by developing risk sharing attitudes and institutions, among

[36] See Appendix B for an overview of the three major climatic shocks.

[37] This measure combines information on the residual and the leverage. The higher the Cook's D is, the more influential the point, whereas the convention cut-off point is 4/n where n denotes the number of observations.

Table 15: Robustness to Major Climatic Shocks

	(1)	(2)	(3)	(4)	(5)
	State 1000	Exch 1	State 1000	Transp. 1	State 1000
Var. in Land Suit.	0.168*	0.411***	0.063	0.263**	0.076
	(0.098)	(0.136)	(0.088)	(0.110)	(0.087)
Med. Exch. in 1 CE			0.255***		
			(0.083)		
Med. Transp. in 1 CE					0.350***
					(0.122)
Medieval Warm Period	0.106	-0.091	0.129*	-0.029	0.116*
	(0.068)	(0.071)	(0.067)	(0.064)	(0.068)
Little Ice Age	0.051	0.380***	-0.046	0.178**	-0.012
	(0.071)	(0.119)	(0.075)	(0.089)	(0.077)
Droughts	-0.151	-0.054	-0.137	-0.016	-0.146
	(0.135)	(0.214)	(0.123)	(0.139)	(0.126)
Cont. Dummies	Yes	Yes	Yes	Yes	Yes
Pers.-Rom. Occup.	Yes	Yes	Yes	Yes	Yes
Observations	117	117	117	117	117
R-squared	0.664	0.711	0.700	0.816	0.702

Summary: This table establishes the robustness of the results to the use of dummies that capture major climatic shocks, such as the Medieval Warm Period, the Little Ice Age and the droughts. The significant effect of the land variablity index is established while controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

Notes: (i) The major shocks that are captures in this table are the 'Little Ice Age' (1350 CE-1850 CE), the 'Medieval Warm Period' (950 CE-1250 CE) and the drougths (500 CE-1500 CE); (ii) land suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate and soil suitability for cultivation; (iii) land suitability diversity is a Gini index of the land suitability index; (iv) the indices of "Transportation in the Year 1" and "Medium of Exchange in the Year 1" are technology indices aggregated at the country level. Each of these two sectors is reported on a 3-point scale, as evaluated by various anthropological and historical sources; (v) the set of continent dummies includes a fixed effect for Africa, the Americas, Australia, Europe and the Middle-East (vi) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (vii) robust standard error estimates are reported in parentheses; (viii) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

Table 16: Robustness to Influential Observations

	(1)	(2)	(3)	(4)	(5)
	State 1000 CE	Exch 1CE	State.1000 CE	Transp. 1 CE	State 1000 CE
Var. in Land Suit.	0.178*	0.470***	0.097	0.300***	0.115
	(0.093)	(0.155)	(0.089)	(0.109)	(0.092)
Med. Exch. in 1 CE			0.206***		
			(0.060)		
Med. Transp. in 1 CE					0.273***
					(0.089)
Cont. Dummies	Yes	Yes	Yes	Yes	Yes
Pers.-Rom. Occup.	Yes	Yes	Yes	Yes	Yes
Observations	117	117	117	117	117
R-squared	0.706	0.628	0.760	0.784	0.745

Summary: This table establishes that variability in land suitability is robust to weighting influential observations using Cook's D measure. The significant effect of land variability is established while controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

Notes: (i) Land suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate and soil suitability for cultivation; (ii) land suitability diversity is the range of the land suitability index; (iii) the indices of "Transportation in the Year 1" and "Medium of Exchange in the Year 1" are technology indices aggregated at the country level. Each of these two sectors is reported on a 3-point scale, as evaluated by various anthropological and historical sources; (iv) the set of continent dummies includes a fixed effect for Africa, the Americas, Australia, Europe and the Middle-East (v) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (vi) robust standard error estimates are reported in parentheses; (vii) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

individuals and across regions. Thus it could be plausibly inferred that the need to develop this risk sharing mechanisms and institutions has facilitated interregional interaction and the emergence of states.

Ethnic Diversity Michalopoulos (2012) has established that higher geographic variability is associated with higher contemporary linguistic diversity. Ethnic diversity has a twofold effect on the emergence of states. On the one hand ethnic diversity adversely affects the emergence of states, due to the presence of many different groups that cannot easily be assimilated under one common state. On the other hand, the proposed hypothesis of Michalopoulos (2012) is that differences in land endowments gave rise to location-specific human capital, which in turn can be associated with different productive activities such as agriculture, or pastoral activities. The presence of these groups, each developing a comparative advantage in the production of region specific goods generates more incentive for trade among groups and thus facilitates the emergence of states.

6 Concluding Remarks

This research has empirically established the hypothesis that in early stages of development diversity in land suitability for agriculture had a persistent beneficial effect on the advent of early statehood. A high degree of diversity, and its association with potential gains from trade, accentuated the incentives to develop social, political and physical infrastructure that could facilitate interregional trade. Hence, the emergence of states, driven partly by facilitating the development of the desirable level of trade infrastructure, was expedited in more diverse geographical environments. Exploiting exogenous sources of variation in variability in land suitability for agriculture across and within countries, the research establishes that: i) the advent of statehood was expedited in regions characterized by a higher degree of variability in agricultural suitability and climatic conditions, ii) the effect of variability on statehood operates through the advancement of medium of exchange and transportation, suggesting that it is the pivotal role of states in facilitating trade that ultimately contributed to their emergence and consolidation, and iii) the effect of land variability on statehood dissipates over time.

The results are robust to controlling for all major theories advanced in the literature as to the emergence of states, such as sedentism, the role of agricultural surplus, population pressure and years elapsed since the onset of the Neolithic. The role of long lasting empires (e.g. the Roman empire), has been explored as well. Whereas most of these confounding factors are important determinants of statehood, yet the partial effect of land variability remains significant throughout.

References

Alesina, A., Easterly, W. and Matuszeski, J. (2011). Artificial States, *Journal of the European Economic Association* **9**(2): 246–277.

Allen, R. C. (1997). Agriculture and the Origins of the State in Ancient Egypt, *Explorations in Economic History* **34**(2): 135–154.

Ames, K. M. (2007). The Archaeology of Rank, *Handbook of Archaeological Theories* pp. 487–513.

Andersen, T. B., Bentzen, J., Dalgaard, C.-J. and Sharp, P. (2011). Religious Orders and Growth through Cultural Change in Pre-Industrial England. University of Copenhagen Discussion Paper No. 11-07.

Andersen, T. B., Dalgaard, C.-J. and Selaya, P. (2011). Eye Disease and Development. University of Copenhagen Discussion Paper No. 11-22.

Arnold, J. (1993). Labor and the Rise of Complex Hunter-Gatherers, *Journal of Anthropological Archaeology* **12**(1): 75–119.

Ashraf, Q. and Galor, O. (2011). Dynamics and Stagnation in the Malthusian Epoch, *American Economic Review* **101**(5): 2003–2041.

Ashraf, Q. and Galor, O. (2011b). Cultural Assimilation, Cultural Diffusion, and the Origin of the Wealth of Nations. NBER Working Paper No. 17640.

Ashraf, Q. and Galor, O. (2013). The 'Out of Africa'Hypothesis, Human Genetic Diversity, and Comparative Economic Development, *American Economic Review* **103**(1): 1–46.

Ashraf, Q. and Michalopoulos, S. (2013). Climatic Fluctuations and the Diffusion of Agriculture. NBER Working Paper.

Bates, R. (1983). Modernization, Ethnic Competition, and the Rationality of Politics in Contemporary Africa, *State Versus Ethnic Claims: African Policy Dilemmas* **152**: 171.

Berg, B. (2007). Volatility, Integration and Grain Banks: Studies in Harvests, Rye Prices and Institutional Development of the Parish Magasins in Sweden in the 18th and 19th Centuries. Stockholm School of Economics.

Cameron, C., Gelbach, J. and Miller, D. (2011). Robust Inference With Multiway Clustering, *Journal of Business and Economic Statistics* **29**(2).

Carneiro, R. (1994). A Theory of the Origin of the State, *The State: Critical Concepts* **1**: 433.

Chanda, A. and Putterman, L. (2007). Early Starts, Reversals and Catch-up in the Process of Economic Development, *The Scandinavian Journal of Economics* **109**(2): 387–413.

Childe, G. (1954). *What Happened in History?*, Harmondsworth.

Claessen, H. and Skalník, P. (1978). *The Early State*, Vol. 32, Walter de Gruyter.

Cohen, A. B. and Yosef, O. B. (2002). Early Sedentism in the Near East, pp. 19–38.

Diamond, J. (1997). *Guns, Germs and Steel: The Fates of Human Societies*, W. W. Norton & Co., New York, NY.

Durante, R. (2010). Risk, Cooperation and the Economic Origins of Social Trust: An Empirical Investigation. Brown University, Working Paper.

Fenske, J. (2013). Ecology, Trade and States in Pre-Colonial Africa, *Journal of the European Economic Association (forthcoming)* .

Gallup, J. L., Sachs, J. D. and Mellinger, A. D. (1999). Geography and Economic Development, *International Regional Science Review* **22**(2): 179–232.

Galor, O. (2011). *Unified Growth Theory*, Princeton University Press, Princeton, NJ.

Galor, O., Moav, O. and Vollrath, D. (2009). Inequality in Landownership, the Emergence of Human-Capital Promoting Institutions, and the Great Divergence, *Review of Economic Studies* **76**(1): 143–179.

Giuliano, P. and Nunn, N. (2013). The Transmission of Democracy: From the Village to the Nation-State, *The American Economic Review* **103**(3): 86–92.

Gosden, C. (1989). Debt, Production, and Prehistory, *Journal of Anthropological Archaeology* **8**(4): 355–387.

Halstead, P. (1989). The Economy has a Normal Surplus: Economic Stability and Social Change Among Early Farming Communities of Thessaly, Greece, *Bad Year Economics: Cultural Responses to Risk and Uncertainty* pp. 68–80.

Hariri, J. G. (2012). The Autocratic Legacy of Early Statehood, *American Political Science Review* **106**(3): 471–494.

Kosse, K. (1994). The Evolution of Large, Complex Groups: A Hypothesis, *Journal of Anthropological Archaeology* **13**(1): 35–50.

Lagerlöf, N.-P. (2012). Statehood, Democracy and Preindustrial Development, *Working Paper* .

Litina, A. (2013). The Geographic Origins of States, *Working Paper* .

Maddison, A. (2003). *The World Economy: Historical Statistics*, OECD, Paris.

Mann, M. (1986). *The Sources of Social Power, Vol. I: A History of Power from the Beginning to 1760 AD*, Cambridge: Cambridge University Press.

Mann, M., Zhang, Z., Rutherford, S., Bradley, R., Hughes, M., Shindell, D., Ammann, C., Faluvegi, G. and Ni, F. (2009). Global Signatures and Dynamical Origins of the Little Ice Age and Medieval Climate Anomaly, *Science* **326**(5957): 1256–1260.

Mayshar, J. and Moav, O. (2011). Transparency, Appropriability and the Early State.

McEvedy, C. and Jones, R. (1978). *Atlas of World Population History*, Penguin Books Ltd., New York, NY.

Michalopoulos, S. (2012). The origins of ethnolinguistic diversity, *American Economic Review* **102**(4): 1508–39.

Nordhaus, W. D. (2006). Geography and Macroeconomics: New Data and New Findings, *Proceedings of the National Academy of Sciences* **103**(10): 3510–3517.

Olsson, O. and Hibbs, Jr., D. A. (2005). Biogeography and Long-Run Economic Development, *European Economic Review* **49**(4): 909–938.

Özak, Ö. (2012). Distance to the Technological Frontier and Economic development, *Available at SSRN 1989216* .

Peregrine, P., Ember, C. and Ember, M. (2007). Modeling State Origins Using Cross-Cultural Data, *Cross-Cultural Research* **41**(1): 75–86.

Putterman, L. (2008). Agriculture, Diffusion, and Development: Ripple Effects of the Neolithic Revolution, *Economica* **75**(300): 729–748.

Ramankutty, N., Foley, J. A., Norman, J. and McSweeney, K. (2002). The Global Distribution of Cultivable Lands: Current Patterns and Sensitivity to Possible Climate Change, *Global Ecology and Biogeography* **11**(5): 377–392.

Spolaore, E. and Wacziarg, R. (2009). The Diffusion of Development, *Quarterly Journal of Economics* **124**(2): 469–529.

Steckel, R. and Rose, J. C. (2002). *The Backbone of History: Health and Nutrition in the Western Hemisphere*, Vol. 2, Cambridge University Press.

Testart, A., Forbis, R., Hayden, B., Ingold, T., Perlman, S., Pokotylo, D., Rowley-Conwy, P. and Stuart, D. (1982). The Significance of Food Storage Among Hunter-Gatherers: Residence Patterns, Population Densities, and Social Inequalities, *Current Anthropology* pp. 523–537.

Vollrath, D. (2011). The Agricultural Basis of Comparative Development, *Journal of Economic Growth* **16**(4): 343–370.

Appendices

A Historical Maps

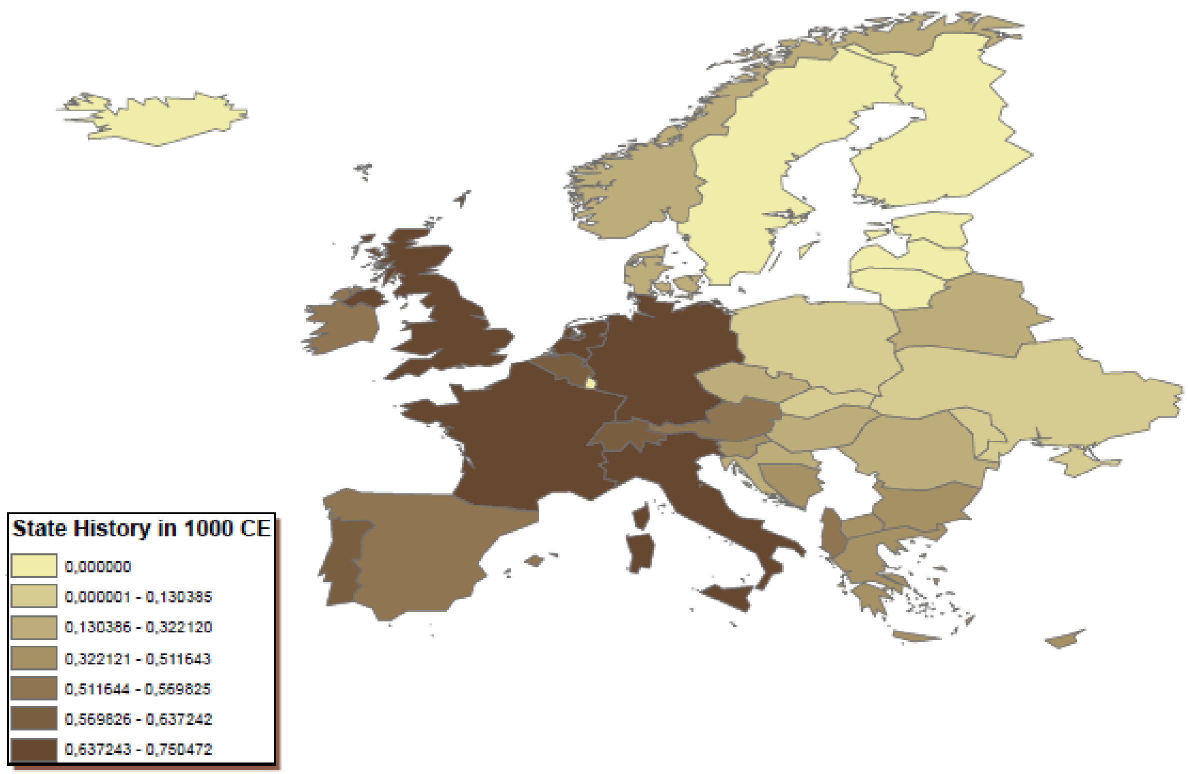


FIGURE 1: State History in Europe (1000 CE)

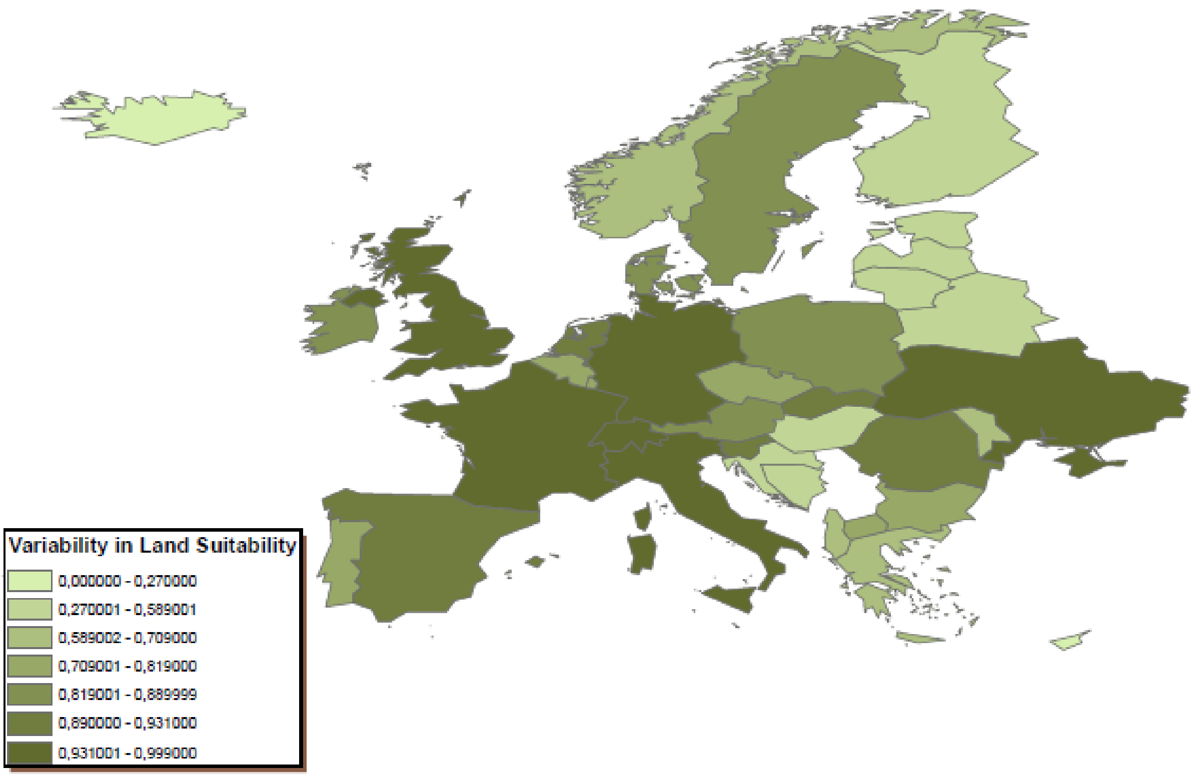


FIGURE 2: Variability in Land Suitability for Agriculture in Europe

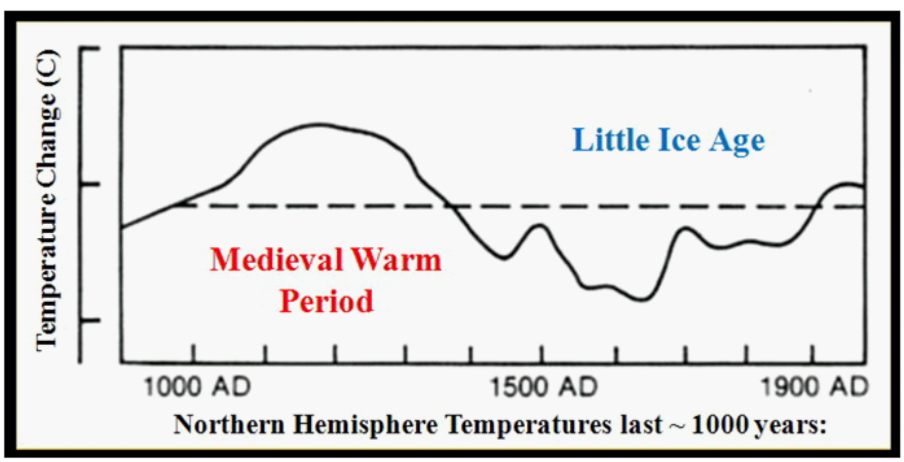
B Major Climatic Shocks

During the period that is being examined in the paper, three major climatic changes have occurred that could potentially affect statehood: i) the Medieval warm period (950 CE-1250 CE), ii) the Little Ice Age (1350 CE-1850 CE) and iii) droughts (500 CE-1500 CE).

The main characteristics of each of the climatic shocks are the following:

1) The Medieval warm period, extended from 950 CE till 1250 CE, and affected primarily the North Atlantic region, Southern Greenland, the Eurasian Arctic, and parts of North America. The most warm segment was between 950 CE and 1000 CE, however the mean temperature was $0.1C^0 - 0.2C^0$ lower than the mean temperature between 1960-1990. In addition, the Intergovernmental Panel on Climate Change (IPCEE) Third Assessment Report (2001) summarized the findings by arguing that "...current evidence does not support globally synchronous periods of anomalous cold or warmth over this time frame, and the conventional terms of 'Little Ice Age' and 'Medieval Warm Period' appear to have limited utility in describing trends in hemispheric or global mean temperature changes in past centuries".

2) The Little Ice Age, is extending from 1350 CE to 1850 CE. The IPCC report argued that the timing and areas affected by the Little Ice Age suggested largely independent regional climate changes, rather than a globally synchronous increased glaciation, however the consensus reached is that it affected primarily the Northern Hemisphere. Still however, research is inconclusive as to whether variations in temperature are sufficient to identify the period as "Little Ice Age".



iii) Various waves of droughts have been reported, that expand from approximately 500 CE to 1500 CE. They appear sporadically and non-systematically in different regions of the globe and they are mostly viewed as side effects of the "Medieval Warm Period" and of the "Little Ice Age". Archeologists argue that certain periods of draughts contributed to the collapse of Meso and South American civilizations, such as the Maya (900 CE-1000 CE), the Tula (1200 CE), the Tiwanaku (1100 CE) and the Wari (1150 CE).

C Alternative Theories on State Formation

The literature on the emergence of early states has suggested a number of alternative hypotheses. As most of these theories have not been empirically established, this section of the appendix will explore the dominant confounding factors as suggested by the literature. More analytically, the four alternative hypotheses that will be tested are: i) sedentism, i.e. that early civilizations developed in proximity to waterways, ii) the surplus hypothesis, iii) the population density hypothesis, and iv) the hypothesis that links the emergence of states to the timing of the Neolithic Revolution. The competing hypotheses that are tested in this section are employed as controls for the baseline analysis, in order to isolate the effect of variability in land suitability on state formation.

C.1 Sedentism

According to the sedentism hypothesis, the notions of state formation, stratification and social complexity became relevant only after populations settled in particular locations. Once a population had settled, early state formation took place through different mechanisms, e.g. surplus, stratification, etc. Whereas settlement could occur at any place that would be conducive to agriculture, nevertheless it has been argued that aquatic economies had an earlier tendency to sedentism and experienced higher population density (Mann, 1986).

To explore the suggested hypothesis, the analysis uses as the explanatory variable the average proximity to waterways. Table C.1 presents the results from testing the sedentism hypothesis, exploiting variations in average proximity to waterways for a sample of 117 countries for which the full set of controls is available.

Column (1) employs the measure of average distance to navigable waterways, while controlling only for continental fixed effects. The coefficient of distance to waterways is negative and significant suggesting that larger distance to waterways is associated with reduced incentives for states to emerge. Column (2) introduces into the analysis a set of geographical controls, namely the average fraction of land located within 100 km from water, average ruggedness, elevation, area of the country in sq. km, absolute latitude, and dummies for climatic zones. All these controls could have affected statehood via affecting climatic and geographical conditions conducive to the emergence of states. Nevertheless, most of these controls do not confer a statistically significant effect with the exception of a dummy variable about whether the country is located in tropical and subtropical zones, and whose coefficient is negative and significant. The reason for the sign of the coefficient could potentially reflect that the climate in these zones was neither conducive to agriculture nor to the emergence of surplus. The coefficient on proximity to waterways remains negative and significant and even increases in magnitude. Column (3) adds to the analysis some historical controls, such

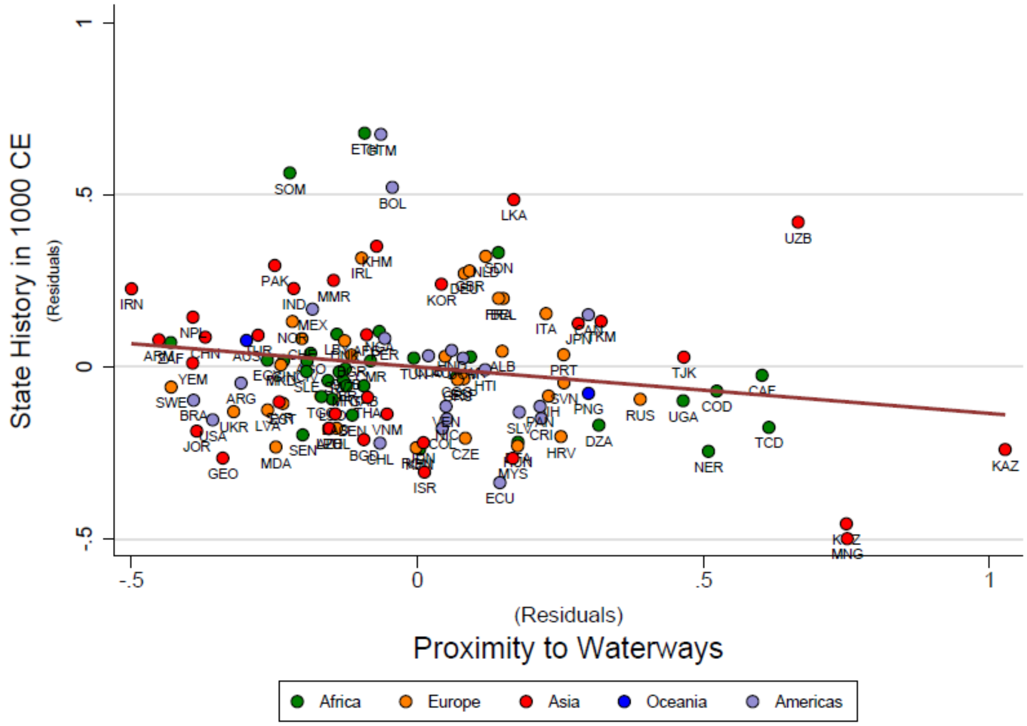


FIGURE C.1: Sedentism Hypothesis-Conditional on controlling for the fraction of land within 100km of coast/river, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

as distance to the nearest technological frontier in the year 1000 CE, as well as fixed effects for Roman or Persian occupation both of which could have facilitated statehood. Indeed the Roman occupation fixed effect has a positive and significant coefficient suggesting that Roman occupation has fostered statehood, potentially via imposing the administrative infrastructure inherent to the notion of the state. Whereas the coefficient on proximity to waterways reduces in magnitude and significance, yet its effect does not vanish. Overall, the results in Table C.1 suggest that sedentism, particularly when it occurred in proximity to water resources, has been an important determinant of the emergence of early states.

C.2 The Surplus Hypothesis

According to the hypothesis featuring the role of surplus, its availability expedited the emergence of an elite class, and the outcome of this increased stratification was the formation of the state. Whereas a number of factors can be associated with the emergence of surplus, nevertheless it could be plausibly assumed that two geographical pre-conditions must be satisfied: i) the existence of a fertile land that will allow for a large volume of agricultural

Table C.1: The Surplus Hypothesis: Testing the Sedentism Hypothesis

	(1)	(2)	(3)
	Dep. Var.:State History in 1000 CE		
Distance to the Nearest Coast/River	-0.140***	-0.176**	-0.135*
	(0.050)	(0.074)	(0.078)
% of Land within 100 km of Coast/River		0.009	0.020
		(0.113)	(0.112)
Average Ruggedness		0.000	0.000
		(0.000)	(0.000)
Average Elevation		0.033	0.033
		(0.084)	(0.080)
Total Area		16.307	19.709*
		(12.067)	(11.254)
Absolute Latitude		-0.011***	-0.009***
		(0.003)	(0.003)
% Land in Tropical and Subtropical Zones		-0.370***	-0.297**
		(0.107)	(0.122)
% Land Temperate Zones		0.001	0.016
		(0.132)	(0.117)
Distance to Frontier in 1000 CE			-0.000
			(0.000)
Continental Dummies	Yes	Yes	Yes
Persian-Roman Occupation	No	No	Yes
Observations	117	117	117
R-squared	0.419	0.533	0.573

Summary: This table establishes that sedentism (settlement of population in proximity to waterways) was a determinant of statehood in the year 1000 CE. The significant negative effect of distance to waterways on statehood is established while controlling for the fraction of land within 100km of coast/river, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

Notes: (i) Distance to the nearest coast/river captures the average distance in km for navigable coasts and/or rivers; (ii) the set of continent dummies includes a fixed effect for Africa, the Americas, Australia, Europe and the Middle-East (iii) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (iv) robust standard error estimates are reported in parentheses; (v) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

production; and ii) the ability to produce storable crops. Whereas the first factor is not sufficient alone, since during this era the Malthusian mechanism is in place, implying that larger production will be channeled towards larger population, the second factor is very important, since the ability to store agricultural production is directly associated with the notion of surplus.

Therefore, in accordance with historical evidence, the strategy adopted in order to explore the surplus channel is to employ as the independent variable i) an index of average land suitability for agriculture (as well as the fraction of arable land), and ii) an index of land suitability for storable crops. These variables are viewed as proxies capturing the potential for surplus.

Table C.2 presents the results from testing the surplus hypothesis, exploiting variations in average land suitability and in the fraction of arable land, for a sample of 117 countries for which the full set of controls is available. Column (1) employs as the explanatory variables an index of average land suitability and the fraction of arable land, while Column (2) introduces an index of suitability for rain-fed cereals. Both columns control for continental fixed effects only. The coefficient on land suitability for agriculture is positive and significant thereby suggesting that higher land suitability can be conducive for the emergence of states, operating via providing surplus. The coefficient on arable land is insignificant at conventional levels, while the coefficient on suitability for cereals is negative and significant at the 10% level, thereby not lending credence to the argument that storability of crops is associated with statehood. Yet, due to the high correlation of the index with the index of suitability, not much can be inferred as to the effect of the potential for storage. Column (3) introduces more geographical controls into the analysis, including proximity to waterways that reflects the sedentism hypothesis. The coefficients on average suitability and suitability for storable crops remain unchanged both in magnitude and significance. Finally Column (4) introduces distance to the nearest technological frontier in the year 1000 CE, as well as fixed effects for Roman and Persian occupation. The results remain intact thereby lending credence to the surplus hypothesis. Moreover, the sedentism hypothesis is also intact as well as the negative effect of the fixed effect for countries located to the tropics. Notably, this control could as well be associated with the surplus hypothesis as crops in these zones are not mostly non-storable.

C.3 The Population Density Hypothesis

According to the population density hypothesis, higher population density could boost statehood for two different reasons: i) by necessitating an increased degree of cooperation and centralized decision making that will allow society to be functional, and, ii) via accentuating the need for the emergence of an authority that can address the difficulties associated with

TABLE C.2: Testing the Surplus Hypothesis

	(1)	(2)	(3)	(4)
	Dep. Var.:State History in 1000CE			
Average Land Suitability	0.323**	0.448***	0.608***	0.497**
	(0.140)	(0.145)	(0.227)	(0.233)
Fraction of Arable Land	-0.516**	-0.313	-0.398	-0.332
	(0.218)	(0.270)	(0.243)	(0.278)
Suitability for Cereals		-0.060*	-0.064*	-0.064*
		(0.033)	(0.033)	(0.033)
Distance to the Nearest Coast/River			-0.227***	-0.186**
			(0.083)	(0.086)
% of Land within 100 km of Coast/River			-0.057	-0.029
			(0.116)	(0.116)
Average Ruggedness			-0.000	-0.000
			(0.000)	(0.000)
Average Elevation			0.089	0.083
			(0.086)	(0.084)
Total Area			22.447*	23.274**
			(11.548)	(11.393)
Absolute Latitude			-0.008**	-0.007**
			(0.003)	(0.003)
% Land in Tropical and Subtropical Zones			-0.393***	-0.317**
			(0.107)	(0.123)
% Land Temperate Zones			-0.169	-0.128
			(0.137)	(0.125)
Distance to Frontier in 1000 CE				-0.000
				(0.000)
Continental Dummies	Yes	Yes	Yes	Yes
Persian-Roman Occupation	No	No	No	Yes
Observations	117	117	117	117
R-squared	0.418	0.433	0.579	0.607

Summary: This table establishes that the potential for agricultural surplus was a determinant of statehood in the year 1000 CE. The significant negative effect of distance to waterways on statehood is established while controlling for the proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

Notes: (i) Land suitability (average land suitability and suitability for cereals) is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate suitability for cultivation, such as growing degree days and the ratio of actual to potential evapotranspiration, as well as ecological indicators of soil suitability for cultivation, such as soil carbon density and soil pH; (ii) arable land is the fraction of total land area that is arable for cultivation; (iii) the set of continent dummies includes a fixed effect for Africa, the Americas, Australia, Europe and the Middle-East (iv) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (v) robust standard error estimates are reported in parentheses; (vi) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

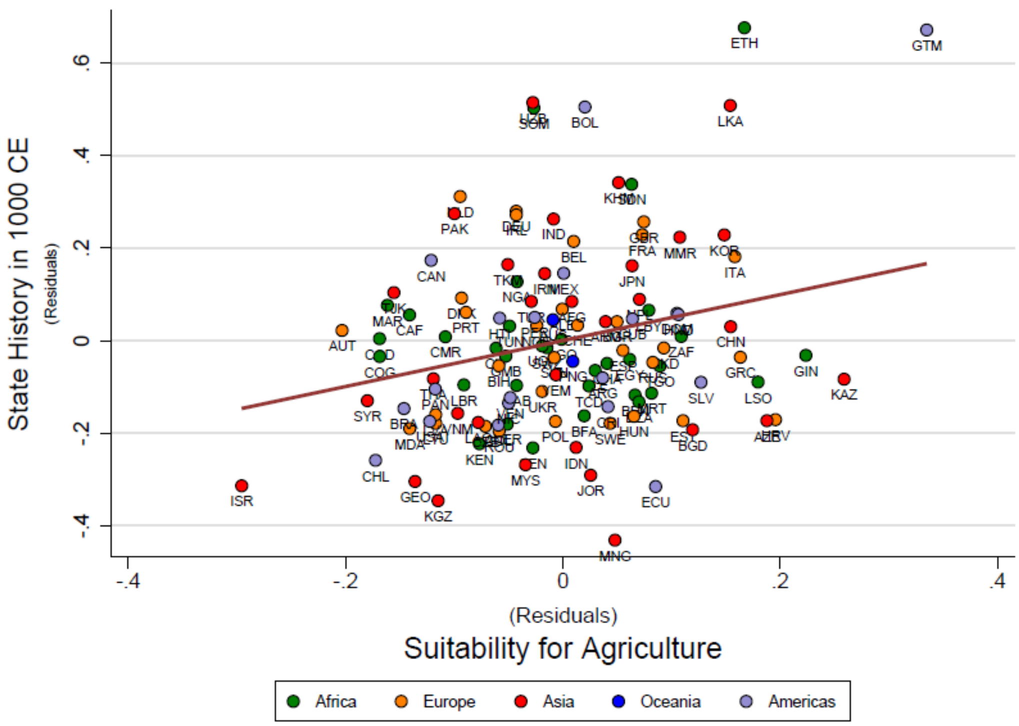


FIGURE C.2: Surplus Hypothesis-Conditional on controlling for proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

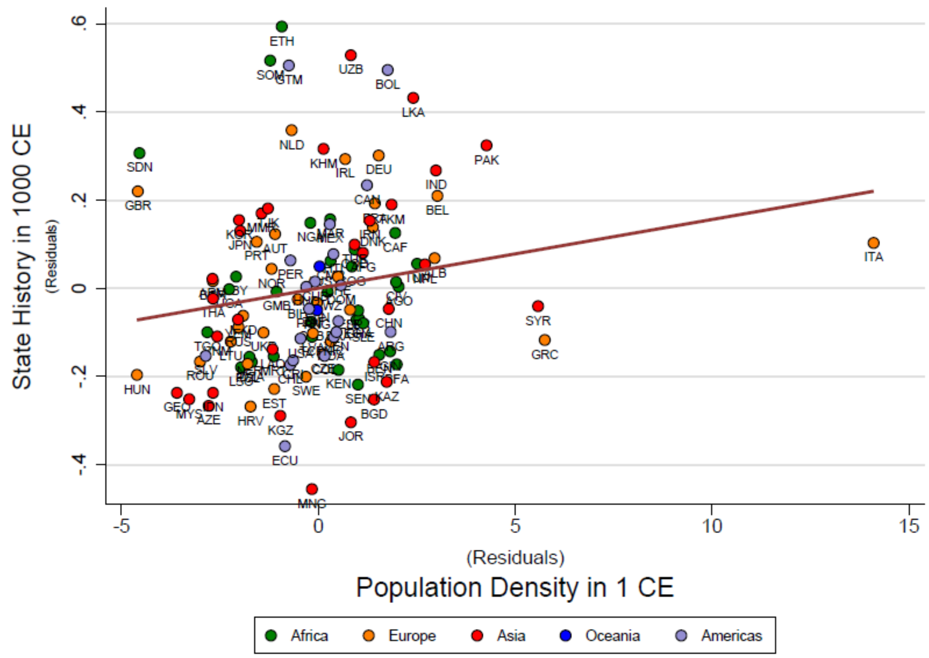


FIGURE C.3: Population Density Hypothesis-Conditional on controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

population pressure and social unrest. To explore this channel, this section will exploit variations in population density across a subset of 117 countries. Moreover, to mitigate reverse causality issues, in the absence of a good instrument for population density, the analysis will employ population density in the year 1 CE whereas the dependant variable is the index of statehood in the year 1000 CE.

Column (1) of Table C.3 suggests a positive and significant effect of population density on the advent on statehood, controlling only for continental fixed effects. Column (2) introduces exogenous geographical controls that include controls for the sedentism and the surplus hypothesis. Column (3) introduces controls on distance to the nearest technological frontier and fixed effects for Persian and Roman occupation. The coefficient on population density in the year 1 CE lends credence to the tested hypothesis, suggesting that a higher and more dense population was associated with earlier emergence of states.

C.4 The Neolithic Revolution Hypothesis

This section will explicitly explore the fourth hypothesis, that links the timing of the transition to agriculture to the advent of statehood. Interestingly, the literature argues that the timing

Table C.3: Testing the Population Density Hypothesis

	(1)	(2)	(3)
	Dep. Var.:State History in 1000CE		
Population Density in the Year 1 CE	0.030***	0.019***	0.016**
	(0.007)	(0.007)	(0.008)
Average Land Suitability		0.518**	0.449*
		(0.243)	(0.241)
Fraction of Arable Land		-0.559**	-0.512*
		(0.226)	(0.270)
Suitability for Cereals		-0.038	-0.038
		(0.035)	(0.038)
Distance to the Nearest Coast/River		-0.212**	-0.185**
		(0.084)	(0.086)
% of Land within 100 km of Coast/River		-0.051	-0.037
		(0.112)	(0.115)
Average Ruggedness		-0.000	-0.000
		(0.000)	(0.000)
Average Elevation		0.085	0.074
		(0.084)	(0.083)
Total Area		20.725*	21.759*
		(10.881)	(11.260)
Absolute Latitude		-0.006*	-0.006*
		(0.003)	(0.003)
% Land in Tropical and Subtropical Zones		-0.342***	-0.307**
		(0.114)	(0.125)
% Land Temperate Zones		-0.197	-0.148
		(0.131)	(0.125)
Distance to Frontier in 1000 CE			-0.000
			(0.000)
Continental Dummies	Yes	Yes	Yes
Persian-Roman Occupation	No	No	Yes
Observations	117	117	117
R-squared	0.467	0.605	0.620

Summary: This table establishes that population density was a determinant of statehood in the year 1000 CE. The significant negative effect of distance to waterways on statehood is established while controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

Notes: (i) The set of continent dummies includes a fixed effect for Africa, the Americas, Australia, Europe and the Middle-East (ii) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (iii) robust standard error estimates are reported in parentheses; (iv) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

of the Neolithic is not directly affecting statehood, but that it has an indirect effect either through allowing for the generation of surplus or via boosting a larger population. This section explores not only whether earlier transition to the Neolithic leads to states, but also whether this effect is direct or not.

Column (1) of Table C.4 suggests that the timing of the Neolithic had a positive effect on the advent of statehood, while controlling for continental fixed effects. The same positive effect persists once controlling for a set of geographical characteristic in Column (2) that incorporate the sedentism and the surplus hypothesis. Yet, the coefficient in Column (2) is much lower and weaker. Column (3) introduces distance to the nearest technological frontier as well as fixed effects for Persian and Roman occupation. Once these channels are controlled for, the effect of the Neolithic drops, thereby confirming that the effect of the Neolithic on the emergence of states operates indirectly via different channels. Finally, Column (4) controls for population density in the year 1 CE, yet the results remain unchanged with respect to Column (4)

Given the endogeneity of this channel, Column (6) employs an IV approach, instrumenting for the timing of the Neolithic with the number of prehistoric plants and animals as suggested byAshraf and Galor (2011). The coefficient on the timing of the Neolithic remains insignificant, thereby confirming that the timing of the Neolithic has a second order effect on statehood (Column (6) just replicates the results of Column (4) by restricting the sample to 79 countries for which the instrument for the channel of the Neolithic is available).

D Variable Definitions and Sources

Outcome Variables

Statehood Index in the Years 500 CE, 1000 CE, 1500 CE, 1950 CE. The statehood variable is using the "State Antiquity" index developed and used by Chanda and Putterman (2007). It is a composite index capturing not only the existence or not of a state, but also the intensity of statehood. In particular it is a composite index, that is a multiple of three components:

$$I_{SH} = I_G \text{x} I_{FL} \text{x} I_T$$

where each component takes a value based on the related answer. More analytically, the questions addressed are, i) $I_G \equiv$ Is there a government above the tribal level?; ii) $I_{FL} \equiv$ Is this government foreign or locally based? and iii) $I_T \equiv$ What is the fraction of the modern territory ruled by this government?

The values are assigned as follows:

Table C.4: Testing the Neolithic Hypothesis

	(1)	(2)	(3)	(4)	(5)	(6)
	Dep. Var.:State History in 1000CE				IV Estimates	
Years since Neolithic	0.056***	0.030*	0.017	-0.006	0.014	0.065
	(0.016)	(0.017)	(0.022)	(0.024)	(0.042)	(0.184)
Average Land Suitability		0.601**	0.512**	0.441*	0.660**	0.715*
		(0.231)	(0.237)	(0.246)	(0.323)	(0.370)
Fraction of Arable Land		-0.437*	-0.374	-0.509*	-0.678*	-0.777
		(0.238)	(0.271)	(0.272)	(0.394)	(0.513)
Suitability for Cereals		-0.064*	-0.062*	-0.037	-0.023	-0.018
		(0.033)	(0.033)	(0.038)	(0.052)	(0.057)
Distance to the Nearest Coast/River		-0.214**	-0.178**	-0.188**	-0.190	-0.136
		(0.083)	(0.087)	(0.085)	(0.165)	(0.212)
% of Land within 100 km of Coast/River		-0.037	-0.028	-0.038	0.026	0.070
		(0.116)	(0.117)	(0.115)	(0.144)	(0.215)
Average Ruggedness		-0.000	-0.000	-0.000	-0.000	-0.000
		(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Average Elevation		0.080	0.072	0.078	0.169	0.145
		(0.085)	(0.086)	(0.085)	(0.106)	(0.129)
Total Area		21.554*	22.406*	21.959*	26.888	19.985
		(11.367)	(11.431)	(11.283)	(19.529)	(35.158)
Absolute Latitude		-0.006*	-0.006*	-0.006*	-0.005	-0.004
		(0.004)	(0.003)	(0.003)	(0.005)	(0.005)
% Land in Tropical and Subtropical Zones		-0.305**	-0.301**	-0.312**	-0.264*	-0.249*
		(0.124)	(0.126)	(0.127)	(0.147)	(0.133)
% Land Temperate Zones		-0.135	-0.131	-0.148	-0.215	-0.229
		(0.137)	(0.127)	(0.125)	(0.194)	(0.205)
Distance to Frontier in 1000 CE			-0.000	-0.000	-0.000	-0.000
			(0.000)	(0.000)	(0.000)	(0.000)
Population Density in the Year 1 CE				-0.017*	-0.008	-0.001
				(0.008)	(0.011)	(0.025)
Continental Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Persian-Roman Occupation	No	No	Yes	Yes	Yes	Yes
Observations	117	117	117	117	79	79
R-squared	0.441	0.591	0.609	0.620	0.643	0.634

Summary: This table establishes that earlier transition to the Neolithic is associated with the onset of statehood only indirectly. This result is established while controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

Notes: (i) The measure of the timing of the Neolithic captures the number of years elapsed, since a country started practising agriculture; (ii) the set of continent dummies includes a fixed effect for Africa, the Americas, Australia, Europe and the Middle-East (iii) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (iv) robust standard error estimates are reported in parentheses; (v) *** denotes statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level, all for two-sided hypothesis tests.

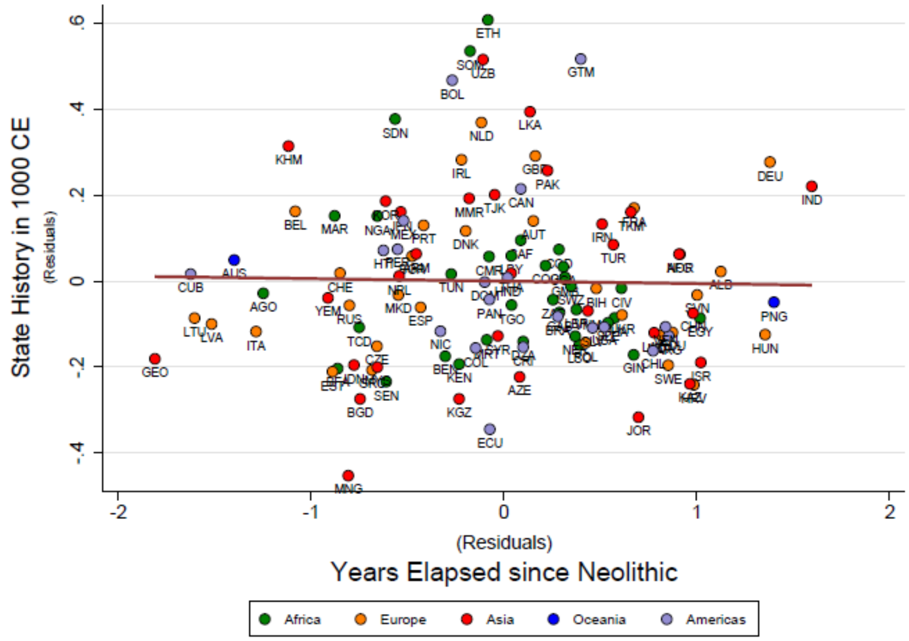


FIGURE C.4: Neolithic Hypothesis-Conditional on controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

i) $I_G \equiv$ Is there a government above the tribal level?

$$\left\{ \begin{array}{c} \text{Yes} \\ \text{No} \end{array} \right\} \Longrightarrow I_G = \left\{ \begin{array}{c} 1 \\ 0 \end{array} \right\}$$

ii) $I_{FL} \equiv$ Is this government foreign or locally based?

$$\left\{ \begin{array}{l} \text{Foreign [e.g. colony]} \\ \text{Hybrid (local with foreign oversight)} \\ \text{Local} \end{array} \right\} \Longrightarrow I_{FL} = \left\{ \begin{array}{c} 0.5 \\ 0.75 \\ 1 \end{array} \right\}$$

and iii) $I_T \equiv$ Fraction of the modern territory, θ_T, ruled by this government

$$\theta_T \in \left\{ \begin{array}{c} [0, 0.1] \\ (0.1, 0.25] \\ (0.25, 0.5] \\ (0.5, 1] \end{array} \right\} \Longrightarrow I_T = \left\{ \begin{array}{c} 0.3 \\ 0.5 \\ 0.75 \\ 1 \end{array} \right\}$$

The statehood index is developed for all countries for all intervals of 50 years starting at the year 1 CE till 1950 CE.

Communication in Year 1, Transportation in Year 1, Medium of Exchange in Year 1. Data on a) Communication in the year 1 CE b) Transportation in the year 1 CE c) Medium of Exchange in the year ,1 CE are constructed from Peregrine's (2003) Atlas of Cultural Evolution, and aggregated at the country level by Ashraf and Galor (2011). Each of these three sectors is reported on a 3-point scale, as evaluated by various anthropological and historical sources. The level of technology in each sector is indexed as follows. In the communications sector, the index is assigned a value of 0 under the absence of both true writing and mnemonic or non-written records, a value of 1 under the presence of only mnemonic or non-written records, and a value of 2 under the presence of both. In the transportation sector, the index is assigned a value of 0 under the absence of both vehicles and pack or draft animals, a value of 1 under the presence of only pack or draft animals, and a value of 2 under the presence of both. In the Medium of Exchange sector, the index is assigned a value of 0 under the absence of domestically used articles and currency, a value of one under the presence of only domestically used articles and the value of 2 under the presence of both. In all cases, the sector-specific indices are normalized to assume values in the [0; 1]-interval. Given that the cross-sectional unit of observation in Peregrine's dataset is an archaeological tradition or culture, specific to a given region on the global map, and since spatial delineations in Peregrine's dataset do not necessarily correspond to contemporary international borders, the culture-specific technology index in a given year is aggregated to the country level by averaging across those cultures from Peregrine's map that appear within the modern borders of a given country.

Geographical Variables

Land Suitability. A geospatial index of the suitability of land for agriculture based on ecological indicators of climate suitability for cultivation, such as growing degree days and the ratio of actual to potential evapotranspiration, as well as ecological indicators of soil suitability for cultivation, such as soil carbon density and soil pH. This index was initially reported at a half-degree resolution by Ramankutty et al. (2002). Formally, Ramankutty et al. (2002) calculate the land suitability index (S) as the product of climate suitability (Sclim) and soil suitability (Ssoil), i.e., S = Sclim Ssoil. The climate suitability component is estimated to be a function of growing degree days (GDD) and a moisture index (α) gauging water availability to plants, calculated as the ratio of actual to potential evapotranspiration, i.e., Sclim = f1(GDD)f2(α). The soil suitability component, on the other hand, is estimated to be a function of soil carbon density (Csoil) and soil pH (pHsoil), i.e. Ssoil = g1(Csoil)g2(pHsoil). The functions, f1(GDD), f2(α), g1(Csoil), and g2(pHsoil) are chosen by Ramankutty et al. (2002) by empirically fitting functions to the observed relationships between cropland areas, GDD, α, Csoil, and pHsoil. For more details on the specific functional forms chosen, the interested reader is referred to Ramankutty et al. (2002). Since Ramankutty et al. (2002) report the land suitability index at a half-degree resolution, Michalopoulos (2012) aggregates the index to the country level by averaging land suitability across grid cells within a country. This study employs the country-level aggregate measure reported by Michalopoulos (2012) as the control for land suitability in the baseline regression specifications for both historical population density and contemporary income per capita.

Land Suitability Diversity. The land suitability diversity measure is based on the range of the land suitability index, reported at a half-degree resolution by Ramankutty et al. (2002), across grid cells within a country. This variable is obtained from the data set of Michalopoulos (2012).

Land Suitability Gini. The land suitability Gini measure is a Gini index built using the land suitability index, reported at a half-degree resolution by Ramankutty et al. (2002), across grid cells within a country. This variable is obtained from the data set of Michalopoulos (2012).

Climatic Suitability. Climatic suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of climate suitability for cultivation such as growing degree days and the ratio of actual to potential evapotranspiration. This index was initially reported at a half-degree resolution by Ramankutty et al. (2002) whereas the country-level aggregate measure is obtained by Michalopoulos (2012).

Soil Suitability. Soil suitability is a geospatial index of the suitability of land for agriculture based on ecological indicators of soil suitability for cultivation, such as soil carbon density and soil pH. This index was initially reported at a half-degree resolution by Ramankutty et al. (2002) whereas the country-level aggregate measure is obtained by Michalopoulos (2012).

Absolute Latitude. The absolute value of the latitude of a country's approximate geodesic centroid as reported by the CIA's *World Factbook*.

Distance to Waterways. The distance, in thousands of kilometers, from a geospatial grid cell to the nearest ice-free coastline or sea-navigable river, averaged across the grid cells that are located within a country's national borders. This variable, developed by Gallup, Sachs and Mellinger (1999), is available from the online Research Datasets repository maintained by Harvard University's Center for International Development.

Percentage of Land within 100 km of Waterway. The percentage of a country's total land area that is located within 100 km of an ice-free coastline or sea-navigable river. This variable was originally constructed by Gallup et al. (1999) and is part of Harvard University's CID Research Datasets on *General Measures of Geography* available online.

Average Elevation. The average elevation of a country in thousands of km above sea level, calculated using geospatial elevation data reported by the G-ECON project (Nordhaus, 2006) at a 1-degree resolution. The measure is thus the average elevation across the grid cells within a country.

Average Ruggedness. The measure is the average degree of ruggedness across the grid cells within a country, calculated using geospatial elevation data reported by the G-ECON project (Nordhaus, 2006) at a 1-degree resolution. This variable is obtained from the data set of Michalopoulos (2012).

Total Land Area. The fraction of a country's total land area that is arable, as reported for the year 2000 by the World Bank's World Development Indicators online.

Arable Land. The fraction of arable land as reported by the World Bank statistics.

Percentage of Land in Tropical and Subtropical Climate Zones. The fraction of a country's total land area that is located in regions classi…ed as tropical or subtropical by the Koppen-Geiger climate classification system. This variable, developed by Gallup, Sachs and Mellinger (1999), is available from the online Research Datasets repository maintained by Harvard University's Center for International Development.

Medieval Warm Period. A dummy variable that takes the value 1 if the country has been affected by the "Medieval Warm Period" climatic shock (950 CE-1250 CE), and 0 otherwise The data are constructed by Litina.

Little Ice Age. A dummy variable that takes the value 1 if the country has been affected by the "Little Ice Age" climatic shock (1350 CE-1850 CE), and 0 otherwise. The data are constructed by Litina.

Droughts. A dummy variable that takes the value 1 if the country has been affected by the droughts (500 CE-1500 CE), and 0 otherwise . The data are constructed by Litina.

Historical Variables

Distance to Frontier in the Year 1, 1000 and 1500.: The distance, in thousands of kilometers, from a country's modern capital city to the closest regional technological frontier in the year 1500 CE, as reported by Ashraf and Galor (2011a). Specifically, the authors employ historical urbanization estimates from Tertius Chandler (1987) and George Modelski (2003) to identify frontiers based on the size of urban populations, selecting the two largest cities from each continent that belong to different sociopolitical entities.

Population Density in the Year 1, 1000, and 1500. Population density (in persons per square km) for given year is calculated as population in that year, as reported by McEvedy and Jones (1978), divided by total land area as reported by the World Bank's *World Development Indicators*. The cross-sectional unit of observation in McEvedy and Jones' (1978) data set is a region delineated by its international borders in 1975. Historical population estimates are provided for regions corresponding to either individual countries or, in some cases, to sets comprised of 2–3 neighboring countries (e.g., India, Pakistan, and Bangladesh). In the latter case, a set-specific population density figure is calculated based on total land area and the figure is then assigned to each of the component countries in the set. The same methodology is also employed to obtain population density for countries that exist today but were part of a larger political unit (e.g., the former Yugoslavia) in 1975. Historical population estimates are also available from Maddison (2003), albeit for a smaller set of countries than McEvedy and Jones (1978).

Years since Neolithic Revolution. The number of thousand years elapsed, until the year 2000 CE, since the majority of the population residing within a country's modern national borders began practicing sedentary agriculture as the primary mode of subsistence. This measure, reported by Putterman (2008), is compiled using a wide variety of both regional and country-specific archaeological studies as well as more general encyclopedic works on the transition from hunting and gathering to agriculture during the Neolithic.