

CREA Discussion Paper 2014-28

Center for Research in Economics and Management
University of Luxembourg

The Geographical Origins of Early State Formation

available online : http://www.fr.uni.lu/recherche/fdef/crea/publications2/discussion_papers

Anastasia Litina, CREA, Université de Luxembourg

November, 2014

For editorial correspondence, please contact: crea@uni.lu
University of Luxembourg
Faculty of Law, Economics and Finance
162A, avenue de la Faiencerie
L-1511 Luxembourg

The Geographical Origins of Early State Formation*

Anastasia Litina[†]

November 28, 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 land and climatic 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. To explore the hypotheses the analysis exploits exogenous sources of variation in a) the measure of land variability across countries, and b) 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 land and climatic conditions, ii) the effect of (land and climatic) variability on statehood operates partly through the advancement of technologies associated with trade, thus 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 for agriculture

JEL Classification Numbers: O50, O57, N70

*I am grateful to Jean Francois Carpentier, Antonio Cosma, Carl-Johan Daalgard, James Fenske, Oded Galor, Masashiye Hamano, Andreas Irmen, Marc Klemp, Battista Severgnini, Gautham Tripathi, Skerdi Zana and participants at the Copenhagen seminar series, CREA seminar series and FRESH meeting in Esjberg 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 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 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 land and climatic diversity, and its association with potential gains from trade, accentuated the incentives to develop social, political and physical infrastructure that could facilitate interregional interaction. Consequently, the emergence of states was expedited in areas where land suitability for agriculture and climatic conditions reflected larger variability.

Differences in the comparative advantage of 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.

¹The notion of state has been extensively debated in the related literature. While multiple definitions exist, the two central features in all definitions are i) the presence of a territorial base, and, ii) the tiers of decision making. Claessen and Skalnik (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).

²Wikipedia lemma on Ethiopia.

This research contributes primarily to the literature that explores the origins of statehood. 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 underlines the role of transparency in the production technology and its effect on the ability to extract resources from the population.

Analytically, the first hypothesis is widely known 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 driven by 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 crops³, 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) exogenous shocks (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 lead to increases in population. (Mayshar and Moav, 2011).⁴

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

³Diamond (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 and higher population densities.

⁴It 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).

⁵See Carneiro (1994); Kosse (1994).

the natural environment. Alternative determinants of a higher population density could be the transition to the Neolithic⁶, 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.

A second strand of the literature focuses not on the emergence of states as such but in the forces behind the centralization of states and the type of the emerging institutions. Fenske (2013) explores the forces behind the centralization of states in pre-colonial Africa and argues that trade supported class stratification between rulers and ruled. Giuliano and Nunn (2013); Hariri (2012); Lagerlöf (2012) explore the driving forces of the emergence of autocratic versus democratic regimes.

This paper contributes an additional theory to the emergence of states by advancing the hypothesis that it is the incentive 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 to trade to ample geographic and climatic variability.⁸ 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.

The theory also sheds light on the origins of the contemporary differences in income per capita across the globe.⁹ Recently, a number of researchers have traced the imprint of early states on the economic, cultural and institutional evolution of modern states (Chanda

⁶It 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).

⁷The 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).

⁸For a historical approach of trade see e.g., Borner and Severgnini (2011); Michalopoulos et al. (2013).

⁹Important contributions in the literature of comparative development are Galor and Weil (2000); Galor and Moav (2002); Galor et al. (2009); Ashraf and Galor (2011b); Özak (2012); Ashraf and Galor (2013a); Ashraf and Michalopoulos (2013); Ashraf and Galor (2013b); Fenske and Kala (2013); Michalopoulos and Papaioannou (2013).

and Putterman, 2007; Borcan et al., 2014). The hypothesis advanced by this research on the hypothesized effect of geography both on the incentives to trade and on its persistent effect on the advent of statehood, can implicitly account for differences in comparative development across the globe. Moreover the paper explicitly tests for the implications of the theory today by establishing a cultural effect of land variability on individual attitudes today.

The paper comprises two parts, a theoretical part and an empirical part. 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. Due to differences in land productivity, they produce different goods and they have the option either to produce locally and live in autarky or to trade with the neighboring region and thus increase the variety of the goods they consume. 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.

The empirical part exploits exogenous sources of variation in a) the measure of land variability across countries, and b) 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 land and climatic conditions, ii) the effect of (land and climatic) variability on statehood operates partly through the advancement of technologies associated with trade, thus 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 by the historical literature.¹⁰

¹⁰ 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.

Second, the reduced form hypothesis (i.e., the effect of climatic and land variability on the emergence of states) is further explored exploiting variations in climatic conditions using a newly constructed panel dataset and exploiting within country climatic variations. The empirical results are verified in a panel of countries for the period 500-1950 CE. This approach further mitigates concerns about omitted variable bias. Additional time varying controls such as regional and country trends are taken into account.

A discussion section analyzes a number of issues related to the main hypothesis. First the analysis explores and discusses several other channels that could link land variability to state formation. Section the section discusses whether it is regional or international trade that matters. Last, the implications of the paper for current economic outcomes are explored. The paper explicitly tests for the implications of the theory today by establishing a cultural effect of land variability on individual attitudes today. In particular, the analysis exploits a sample of immigrants from the European Social Survey to establish that higher land variability at the origin country increases the probability that an individual is employed in a trade related activity. This effect is established after controlling for a wide range of geographical and historical controls associated with both the origin and the host country as well as for individual controls.

The results are robust to a number of robustness tests, related to the validity of the statehood index and the land variability index, to the validity of the estimation (outliers) and omitted variable bias (spatial autocorrelation, regional fixed effects) as well as to alternative specifications. 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 can foster 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.¹¹ The costs assumed can

¹¹An 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.

be related to distance as well as to climatic and geographical conditions.

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 a viable option, 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.¹² 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.¹³ 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 issues.¹⁴

The aim of the social planner is to maximize the collective utility of the region.¹⁵ The utility function of the region i , U_i is described by

¹²For 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.

¹³This 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.

¹⁴The complexity hypothesis 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 channel of trade. Moreover it can be assumed that these channels can operate in a complementary fashion, e.g. the complexity hypothesis leads to a hierarchical structure within the community and once a leader is elected within the community, intra-regional trade can take place which further reinforces states.

¹⁵Assuming 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 in order to capture his rent-seeking attitude, nevertheless his economic incentives would still be in accordance with the incentives to trade and to facilitate trade.

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

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

$$x_i = \theta_i(1 - z)Y_i \quad (2)$$

where

$$Y_i = \Gamma_i X_i^a L_i^{1-a} \quad (3)$$

where θ_i ($0 \leq \theta_i \leq 1$) 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,¹⁶ 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

$$x_i = \theta_i(1 - z)Y_i = \theta_i(1 - z) \quad (4)$$

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

$$Y_j = \Gamma_j X_j^a L_j^{1-a} \quad (5)$$

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 of each individual. For analytical

¹⁶In 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. Thus whereas the possibility of war is not ignored altogether, the focus will remain on the mechanism of trade.

convenience it is assumed that $X_j = L_j = 1$. Since we do not model the behavior of the region j we are not interested in the fraction of labor that is allocated to local production and thus for simplicity we'll assume that it allocates all its labor force into productive activities. Upon implementing the above assumptions eq. (5) reads as

$$Y_j = \Gamma_j X_j^a L_j^{1-a} = \Gamma_j \quad (6)$$

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

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

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.¹⁷

2.1.4 Utility Maximization

Combining Eqs. (1), (7), (4) and (6) 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., between local production and intra-regional trade, θ_i

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

The utility maximization yields

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

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.

¹⁷For 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.

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 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. \square

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 be involved in trade activity; ii) $\lim_{\Gamma_j \rightarrow 0} \theta_i^* = 1$, 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 $\omega \rightarrow 1 + \Omega z \implies \theta_i^* \rightarrow 1$ (recall that $0 \leq \theta_i^* \leq 1$) and thus no trade takes place; (ii) Similarly, as $\Gamma_j \rightarrow 0 \implies \theta_i^* \rightarrow 1$, 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. \square

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 .

Third, if trade is a viable option, the better the trade infrastructure the higher the incentive to trade. Whereas in the theoretical model causality can ran both ways 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 (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

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

where the notation remains the same, with the only difference now being that the cost of developing infrastructure is now cut down by half, whereas the remaining cost is undertaken 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. (7). Everything else remains the same in the structure of the model. Interestingly, replicating the analysis of the previous section and maximizing 1 to obtain the optimal θ_i we still obtain the same solution as described in Eq. (9)

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

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.¹⁸

¹⁸This 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 not the only one. The aim of this section is to model the case where region i attacks region j , whereas the other region attempts 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 it 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 is now given by

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

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

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

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

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

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 pay for the "usurped" 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

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

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) $\theta_i^* = \theta_i^{PU^*}$ i.e., the optimal choice of trade volume is not affected by the decision on political unification or not; ii) $\theta_i^* = \theta_i^{PU^*} \geq \theta$ if $-\omega + \Omega z \leq -\tilde{\omega} + \Omega \tilde{z}$, 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 volume 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) $U_i^* \geq U_i^{W^*}$ and $U_i^{PU^*} \geq U_i^{W^*}$ 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 as well, 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 to unify neighboring regions.

3 Empirical Strategy and Data

3.1 Empirical Strategy

First, the analysis exploits variations in land suitability for agriculture in a cross-section of countries to explore three main hypotheses. Second, the reduced form hypothesis, i.e., the

effect of climatic variability on state formation, 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

The Effect of Variability in Land Suitability on State Formation 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.¹⁹ 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, that have been suggested in the historical literature, 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 lagged 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 for which the full set of controls is available, is described by

$$I_i = \alpha_0 + \alpha_1 V_i + \alpha_2 X_i + \alpha_3 \Delta + \varepsilon_i \quad (16)$$

where the subscript i is a country indicator; I_i is an index of statehood for the year 1000 CE²⁰; V_i is an index of variability in land suitability; X_i is a vector of geographical and historical controls; Δ 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 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

¹⁹The robustness section establishes that the argument is valid for other historical periods as well.

²⁰The construction of the index is described in the data section.

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

$$I_{SH} = \alpha_0 + \alpha T_i + \alpha_1 V_i + \alpha_2 X_i + \alpha_3 \Delta + \varepsilon_i \quad (17)$$

where the subscript i is a country indicator. I_i is an index of statehood for the year 1000 CE²¹; 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 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 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.

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.²² Importantly this approach addresses the issue of potential omitted variable bias since it allows

²¹The construction of the index is described in the data section.

²²More details on the construction of the dataset are available in the data section.

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.²³ 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.²⁴ Importantly, this ensures that the observations are independent to past observations.²⁵

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

$$I_{it} = \alpha_0 + \alpha_1 C_{it} + \alpha_2 I_{it-1} + \alpha_3 X_{it} + \alpha_4 R_i + \alpha_5 T_t + \alpha_6 T_t \Delta + \varepsilon_{it} \quad (18)$$

where the subscript i is a country indicator and t is a time indicator (four intervals are used, i.e. $t_1 = 500$ CE, $t_2 = 1000$ CE, $t_3 = 1500$ CE, $t_4 = 1950$ CE); I_{it} is an index of statehood for each period t ;²⁶ 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; R_i is a country dummy variable; T_t is a time dummy; $T_t \Delta$ is an interactive term of the time period 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:

²³Following Ashraf and Michalopoulos (2013) the average temperature is calculated at the grid level and then the measure is aggregated at the country level

²⁴A 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 12.

²⁵The results are robust to the use of a 50 year average of climatic variability.

²⁶The construction of the index is described in the data section.

$$I_{SH} = I_G \times I_{FL} \times I_T$$

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?

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

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

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

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

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

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.²⁷

²⁷An 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

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.²⁸ 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.²⁹

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 range of the measure of land suitability for agriculture, i.e. the difference between the maximum and the minimum value of land suitability for agriculture in each country.

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

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.

²⁸The statehood index is discounted according to the formula, $I_{SHi_Discounted} = \frac{\sum_{t=0}^{1950} (1.05)^{1-t} * I_{SHi}}{\sum_{t=0}^{1950} (1.05)^{1-t*50}}$, where

I_{SHi} is the non-discounted statehood index. The analysis assumes a 5% discount rate applied to each 50 year interval, following Bockstette et al., (2002). The discounting rates allow to take into account past statehood on the one hand, but attaching less importance to older values of the index on the other hand.

²⁹Crucially 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.

of the index.³⁰ 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 (2013a), 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.³¹ Therefore, even if the characteristics of soil quality have significantly 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

³⁰The 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),

³¹Durante (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, 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.

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.³²

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 shocks 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 tree-ring, 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.³³

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.

³²This approach is reassuring since evidence suggests that climate has not severely changed during the last 2000 years (Durante, 2010).

³³Extensive 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.

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 CE and b) 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 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.

Importantly the measures of technology employed as proxies for trade better capture the argument of the paper and the possibility of intra-regional trade. Alternative measures such as proximity to trade routes as in e.g., Özak (2010, 2012); Michalopoulos et al. (2013), whereas they are more accurate, however they mostly capture international trade and pre-existing, restructured or extended trade routes.

4 Empirical Findings

4.1 Variability in Land Suitability and Statehood: Cross Country Analysis

The analysis in the preceding section suggests 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 sedentism³⁴ and the surplus hypothesis.³⁵

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 a control for the timing of the Neolithic.³⁶ After introducing the full set of controls

³⁴According 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 is conducive to agriculture, nevertheless it has been argued that aquatic economies had an earlier tendency to sedentism and experienced higher population densities (Mann, 1986). To explore this hypothesis, the analysis employs two measures of proximity to waterways.

³⁵According to the hypothesis featuring the role of surplus, its availability expedited the emergence of an elite class. As a result of surplus, stratified societies emerged which eventually fostered the formation of states. 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 presence of fertile land that will allow for a large volume of agricultural production; and, ii) the ability to produce storable crops. The first factor is not sufficient since higher suitability does not necessarily imply storability of the agricultural products. The second factor is more important since it captures the potential production of storable goods. Crucially, the ability to store agricultural production is directly associated with the notion of surplus. Therefore, in order to explore the surplus hypothesis the analysis controls for three variables that proxy the potential surplus in agricultural production: i) An index of average land suitability for agriculture; ii) the fraction of arable land; and, iii) an index of land suitability for storable crops.

³⁶The Neolithic hypothesis links the timing of the transition to agriculture to the advent of statehood. Interestingly, the literature argues that the timing 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.

the coefficient on variability remains stable both in magnitude and in significance. Thus land variability is a significant determinant of state formation in the year 1000 CE.

The results in Column (4) also suggest that there are other significant channels as well. First, the channel of sedentism as suggested by the negative and systematically significant coefficient on distance to waterways. The higher is the distance to the nearest coast or river the less likely is the emergence of a state in the year 1000 CE. The controls accounting for the surplus hypothesis give mixed results. On the one hand, higher land suitability is positively correlated with state formation. On the other hand though, higher suitability for storable crops is negatively correlated with state formation, thus not lending credence to the surplus hypothesis.

There are two more interesting findings in Column (4). The first is that places located near the tropics are less likely to develop early states. One reason could be that the climate in these zones was neither conducive to agriculture nor to the emergence of surplus. Moreover the goods produced in these areas are less likely to be traded since they can hardly be stored for more than few days. A second interesting finding is that the transition to the Neolithic does not confer a statistically significant effect. A plausible interpretation for this finding is that the transition to the Neolithic affected state formation in several ways. The related literature argues that the Neolithic revolution had an indirect effect on state formation operating via e.g., the creation of surplus or sedentism. Once controlling for all these channels, which is the case in Column (4), the effect of the Neolithic dissipates and is potentially captured via some of the other controls.

TABLE 1 HERE

In terms of magnitude, Column (5) replicates the analysis in Column (4) and presents the beta coefficients. The beta coefficient of land variability, suggests that an one standard deviation increase in land variability is associated with a 0.2 standard deviation increase in the statehood index. The stronger effect appears to be the effect of land suitability followed by the sedentism hypothesis.

Overall, the findings in Table 1 are confirming that variability in land suitability had a direct effect on the onset of statehood. Figure 1 illustrates the scatter plot for the land variability hypothesis, conditional on controlling for the full set of controls in Column (4).

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 to 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 providing the necessary social and institutional infrastructure.

More analytically, Table 7 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. 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. This process ultimately lead to the emergence of states.

Column (1) of Table 7 replicates the results in Column (4) of Table 1 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. Column (3) runs a horserace between the measure of land variability and medium of exchange technology. Once this measure of trade (exchange 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. Interestingly this result emerges despite the fact that the proxy for trade is imprecisely measured whereas the measure of land variability is more precisely measured. This suggests that the effect of land variability operates through the trade incentives it generates.

Similarly, Column (4) establishes the significant and positive effect of land variability on a second proxy of trade, namely the level of sophistication of transportation technology. Higher land variability is associated with more advanced transportation technology. Column (5) run the horserace between land variability and the measure of transportation technology as potential determinants of state formation. Similarly, the coefficient of land variability drops dramatically and its significant reduces to the 10% level suggesting that the effect of land variability partly operates via trade.

TABLE 7 HERE

Most of the coefficients on the other channels remain unaffected by the introduction of the trade controls which implies that trade is not a channel via which the other hypotheses operates. The only exception is the coefficient on the fraction of land located near the tropics which also drops in magnitude and significance. This perhaps suggests that places located near the tropics mitigated the development of trade-related technologies. Thus the adverse

effect of the tropical zones on statehood partly operates via mitigating trade activities.

Figure 2 illustrates the effect of land variability on the technology of the medium of exchange. Figure 3 illustrates the effect of land variability after controlling for trade technology (medium of exchange). Both figures control for the full set of controls.

Overall, the analysis of this section suggests that one of the mechanisms via which land variability affects statehood is trade. The analysis does not preclude other potential mechanisms that fostered state formation. An extensive description of these mechanisms and their relation to the hypothesis advanced in this paper is provided in the robustness section.

4.1.3 The Intertemporal Effect of Variability on Statehood

This section establishes that variability in land suitability was critical in generating states as long as intra-regional trade was an important driver of regional growth. In the process of development, and as economies expanded their trade possibilities via long-distance and transatlantic trade³⁷, land variability became gradually less important in giving rise to states. To capture this effect this section 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.

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 (4) 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 an effect is

³⁷The role of long-distance trade is further discussed in the robustness section where the measure of distance to trade routes is explicitly controlled for.

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 1% level, yet this effect vanishes once the analysis controls for statehood in the year 1500 CE, in Column (6). The result suggests that land variability 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 land variability was a crucial determinant for the rise of states, yet this effect dissipated over time.³⁸

TABLE 3 HERE

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 of climatic variability available at the grid level by 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 plausibly assumed that variations in temperature can partly capture differences in agricultural suitability and thus the incentive to trade.

Column (1) in Table 4 establishes that climatic variability has a positive and significant effect on statehood. The analysis does not employ any controls. Column (2) controls for country fixed effects, thus controlling for most of unobserved heterogeneity associated with time invariant country specific factors. Reassuringly, the positive and highly significant coefficient on climatic variability remain whereas the coefficient even increases in magnitude. Thus, even when exploiting within country variation, climatic variability still confers a significant effect on the emergence of states thereby lending credence to the reduced form hypothesis established in the cross country analysis. More importantly, any unobservable that is constant at the country level has been controlled for.

TABLE 4 HERE

³⁸Our findings are in line with the literature that explores the origins of modern states. Smith (1986) explores the ethnic origin of modern nation and attributes them to a number of factors associates to the modes of production and administration as well as cultural factors.

Column (3) introduces a dummy for each period, i.e., $t_2 = 500$, $t_2 = 1000$ CE, $t_3 = 1500$ CE and $t_4 = 1950$, in order to capture time specific shocks. Column (4) adds more time varying controls such as the mean temperature. The results remain relatively unaffected. Moreover, the coefficient on the mean level of temperature is negative and significant suggesting that higher average temperature is associated with a lower statehood index, potentially capturing the adverse effect of good climatic conditions on the need to cooperate and coordinate activities (Litina, 2013b).

Table 5 tests the robustness of the baseline specification in Column (4) of Table 4.

TABLE 5 HERE

Column (1) of Table 5 employs an alternative measure of climatic variability, i.e., the measure of standard deviation from average temperature. The coefficient on climatic variability increases sizably in magnitude and remains significant at the 1% level. Column (2) replicates the analysis in Column (4) of Table 4 and introduces continental specific time trends. The set of regional dummies interacted with time includes a fixed effect for Africa, the Americas, Australia, Europe and Oceania. The results remain largely unchanged. Column (3) employs a time trend instead of a time fixed effect assuming a linear trend that homogeneously affected all countries, whereas Column (4) employs country specific time trends. The magnitude of the coefficients remains relatively stable and significant at the 5% level. Columns (5) and (6) exploit 300 year and 200 year intervals respectively to test the robustness of the results to the choice of shorter intervals for the climatic variable. The coefficient remains significant at the 1% and 5% respectively However its magnitude reduces gradually thus suggesting that the mechanism is weaker during shorter periods.³⁹

Overall, the findings of this section suggest that climatic variability had a persistent effect on early states.⁴⁰ Critically, by controlling for country fixed effects as well as for a number of time varying controls, it further mitigates concerns about omitted variables bias. Moreover, the fact that climate was rather unaffected by human activity for the period under examination, is reassuring as to claiming a causal effect.

5 Discussion

These section discusses a number of issues. First, a number of channels are discussed that could potentially link land variability to the emergence of states. Second, the analysis explores

³⁹Two reasons can account for this result. First, as Mann et al. (2009) argue the climate data, being reconstructions of climatic conditions, are more meaningful when they are used for large time periods. Second, the mechanism described in the paper may take long periods to materialize. Thus, intervals as small as 50 years may not be sufficient to allow the regional trade mechanism, driven by climatic variability, to be traced.

⁴⁰Recall that the baseline specification exploits 500 year intervals.

whether it is regional or international trade that links land variability and current state formation. Last, the section explores the implications of the reduced form hypothesis on current economic outcomes. Are the findings of the research of a pure theoretical interest or do they have implications for modern economies?

5.1 Alternative Channels

The main aim of the paper is to establish a casual effect from climatic and land variability to the emergence of states. The analysis in the paper suggests that such a reduced form effect is present. A second argument advanced in the paper is that one of the mechanisms through which land and climatic variability affects states is via 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 mechanisms could be potentially linked to both the climatic variability and the emergence of state. Whereas the aim of the paper is to focus on the trade mechanism, yet a discussion of alternative mechanisms is rather informative. Some of these mechanisms can be effectively controlled for, whereas some others, for which not relevant historical data is available, there is a discussion as to how they could interact with the suggested hypothesis.

Table 6 presents some additional results.

War One of the dominating theories of state formation is the one advanced by Tilly et al. (1985) who argue that war causes states. In this case it could be argued that land and climatic variability provided incentives to attack nearby regions and usurp, and not to exchange, the desired good. This case is certainly a possibility that cannot be precluded and could certainly be an additional channel that links variability with state formation. The theoretical model has already explored such a case and has provided the conditions under which war is not the optimal strategy, thus highlighting the fact that whereas war is a possibility it is not always the optimal one compared to trade. A second argument with the theory of war is that it is primarily applicable to the post-Medieval era primarily for Europe (Tilly et al., 1985), i.e., it is a theory for modern state formation. This is in line with the findings of this paper that suggests that land and climatic variability are crucial determinants for early rather than late state formation (see Table).

To formally test this possibility the analysis employs two proxies of war. The first is a measure of military technology in the year 1 CE employed by (Comin et al., 2010). In particular they use a measure derived by the Atlas of Cultural Evolution (Peregrine, 2003) that describes which metals were available in each society. This serves as a proxy for military

technology in each era.⁴¹ The index in the year 1 CE takes the value 1 if iron is used and 0 otherwise. The second measure is an index of military technology in the year 1500 CE. It is constructed by (Comin et al., 2010) and better captures technology in that era.⁴² Table 6 explores the war theory by controlling for military technology. Analytically, Column (1) introduces in the baseline analysis (i.e., Column (4) in Table 1) the measure of military technology in the year 1 CE.⁴³ The coefficient on military technology is insignificant, whereas the coefficient on land suitability remains unaffected. It is thus inferred that during that period, i.e., in the year 1000 CE, war was not a significant driver of early state formation. Column (2) employs the second proxy of war, i.e., military technology in the year 1500, to explore whether it affected state formation in year 1500 CE.⁴⁴ The coefficient of military technology is positive and significant at the 1% thus suggesting, in line with the related literature, that war in 1500 was a crucial determinant of state formation. However the coefficient of variability is also positive and significant at the 5% level and thus it suggests that the trade mechanism is still in place.

Population Density A second theoretical possibility to be explored is whether land suitability has an effect on population density and thus affected state formation via affecting population density and not via trade. Column (3) of Table 6 explores this possibility by controlling for population density in the year 1 CE. The coefficient on land variability remains relatively stable in magnitude and significant at the 5% level whereas the coefficient on population density is positive and significant is also positive and significant at the 1% level. Thus whereas population density appears to confer a statistically significant effect on early state formation, this effect does not appear to operate via land variability.

Stratification A third possibility is that land variability gave rise to an elite that ultimately led to state formation. The regression in Column (4) of Table 6 controls for a measure of class stratification (Peregrine, 2003).⁴⁵ The coefficient on land variability remains

⁴¹As they argue metallurgy was necessary for the development of advanced weaponry. E.g., stone weapons were succeeded by bronze swords to be later replaced by the iron swords.

⁴²The measure captures the presence of standing army, cavalry, firearms, muskets, filled artillery, warfare capable ships, heavy naval guns and ships. See the appendix for a detailed description of the variable.

⁴³Similar to the measures of trade, the military technology measure is employed with a lag to mitigate reverse causality issues and to capture the process of state formation as potentially driven by wars.

⁴⁴In this regression since the measure of military technology in 1000 CE is not available, the contemporaneous measure is employed. Thus in this case reverse causality cannot be addressed and the results are interpreted as mere correlations.

⁴⁵Social stratification is a measure of social complexity and captures the number of classes within a society. It is constructed from Peregrine's (2003) Atlas of Cultural Evolution. The level of stratification is indexed as follows. The index is assigned a value of 1 for egalitarian societies, a value of 2 for two social classes and a value of 3 for three or more social classes. The index is constructed for the year 1 CE.

positive and significant at the 1% level whereas the magnitude remains stable. The coefficient on stratification suggests that whereas more stratified societies are more likely to develop states, yet this effect does not operate via variability in land suitability.

Column (5) controls for all channels simultaneously. The coefficient on land variability, when controlling for all channels slightly increases in magnitude and remains significant at the 1% level.

TABLE 6 HERE

Ethnic Diversity Michalopoulos (2012) has established that higher geographic variability is associated with higher contemporary ethnic and linguistic diversity. Ethnic diversity could have a twofold effect on the emergence of states. On the one hand ethnic diversity could adversely affect the emergence of states, due to the presence of many different groups that cannot easily be assimilated under one common ruler. On the other hand, the hypothesis established by 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 could give incentives to the emergence of states. In the absence of a control for past ethnic diversity the analysis cannot explicitly control for this channel. However, intuitively it can be inferred that if the presence of many ethnic groups (driven by high land variability as established in Michalopoulos (2012)) mitigated state formation, then controlling for ethnic diversity (had a historical control been available) would further reinforce the results of the paper. Analytically, if land variability has two effects on state formation, a negative effect (driven by fostering ethnic diversity) and a positive effect (driven again by fostering ethnic diversity and thus production specialization and thus the incentives to trade), then netting out the negative effect would reinforce the positive effect.

5.2 International or Regional Trade

A question that arises with respect to the suggested mechanism of trade is whether it concerns international or regional trade. Thus far it has been assumed that the proxies of trade, i.e., transportation and exchange technology primarily capture regional trade. The reason for this assumption is related to how the dataset is constructed. The cross-sectional unit of observation in the ACE dataset (Peregrine, 2013) is an archaeological culture that may occupy only a region of the current unit of analysis, i.e., the current borders. To construct the country level index for each technology in a given year, the index is aggregated to the country level

by averaging across those cultures from Peregrine’s map that are located within the modern boundaries. Thus the aggregated measure reflects the regional levels of each technology.

In order to capture international trade the analysis employs the measure of proximity to trade routes by Özak (2010, 2012). The distance is measured in weeks from the capital of each country in the Old World to trade routes used between 500 BCE and 1900CE. The trade routes in the dataset are primarily perceived as long-distance arteries that extend over tens or hundreds of kilometers. Thus, this measure is a better proxy for international trade.

TABLE 7 HERE

Table 7 explicitly explores the possibility that it is international trade that matters. Column (1) replicates the baseline analysis in Column (4) of Table 1 using as the dependent variable the statehood index in 1500 CE.⁴⁶ The coefficient on land variability suggests that land variability has a significant positive effect on state formation. Column (2) regresses the measure of distance to trade routes on the land variability index. The coefficient is insignificant at conventional levels thus implying that land variability did not have an effect on the formation of long-distance trade routes. Recall that this was not the case for regional trade. When regressing the proxies of regional trade on land variability, the latter conferred a statistically significant effect on the development of regional trade-related technology (see Table 7). Column (3) runs a horserace between the measures of land variability and proximity to trade routes and their effect on state formation. The coefficient of land variability retains its significance, where the coefficients of proximity to trade routes is insignificant. Our interpretation here is that international trade does not foster state formation.

Overall, it is argued that it is regional trade that matters for state capacity. The need to interact and exchange goods with nearby regions (not necessarily located nearby international trade routes) fosters cooperation between neighboring regions which ultimately may administratively link the various areas. On the contrary, Table 7 suggests that international trade does not play a role in unifying regions. Intuitively, this appears plausible as international trade routes do not necessarily entail regional cooperation in order to be constructed. Instead, they could be reconstructed on the remnants of former military routes or of former routes connecting several parts of an empire. Moreover international trade often reflected a trend in demand of some goods, such as demand for spices or tea. Thus any emerging international routes aimed to cover this demand and not to cover regional need. Consequently, regional cooperation was not an issue at stake in the case of international trade. On the contrary, regional trade, as perceived in the context of the current study, is initiated as a means to obtain a larger variety of goods some of which may be necessary for subsistence. Hence,

⁴⁶Note that this is a more relevant measure as the distance to trade routes is aggregated for the whole period 500 BCE-1900 CE.

reaching the neighboring region is mainly reflecting a necessity not as much a trend. And in this case, cooperation and ultimately stronger political and administrative ties mutually benefit both regions.

5.3 Implications of the Research for Current Economic Outcomes

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; Chanda et al., 2014). The majority of this literature associates past statehood with current positive economic outcomes, whereas Borcan et al. (2014) establish that the relationship between the extended state history index and current levels of economic development has an inverted u shape.

Having take stock of the significant effect of state history on modern growth, this subsection takes the analysis to a different direction, i.e., it explores the cultural implications of the explored mechanism. Analytically, it associates land variability with the probability that an individual is engaged in trade-related activities, after controlling for a number of individual and country characteristics. The underlying assumption is that individual coming from countries with a long tradition of trade (partly driven by land variability) are more likely to be engaged in trade.

5.3.1 Land Variability and Trade Related Attitudes

Data The analysis employs data from five waves of the European Social Survey (2004-2012), a cross sectional survey conducted in a number of European countries. One element in the construction of the dataset is that it provides an immigrant identifier that allows to trace immigrants up to the second generation, as well as concrete information about the mother's and the father's country of origin. This element is crucial since it allows researchers to exploit the event of immigration in order to explore the evolution of cultural traits. The identifying assumption in these studies is that when immigrants move to a host country their current attitudes are no longer directly affected by the economic or the institutional environment at the country of origin. Thus, any effect of the origin country on immigrants' attitudes operates indirectly via culture (Fernández and Fogli, 2009).

The Variables The main dependent variable is a binary variable taking the value 1 if the individual (immigrant) is employed in a trade-related sector and the value of 0 otherwise. The ESS also provides information about the age of the respondent, the gender, employment and family status, the highest level of education achieved, level of income, parental and spousal education, citizenship, belonging to a discriminated group or not, and whether the individual voted or not in the last election.

Appendix C provides a detailed description of all the variables used in the baseline analysis and the robustness section.

Empirical Strategy The reduced form model is

$$P_{jhit} = \alpha_0 + \alpha_1 \mathbf{V}_i + \alpha_2 \mathbf{I}_j + \alpha_3 \mathbf{B}_i + \alpha_4 \mathbf{\Phi}_h + \alpha_5 \mathbf{T}_t + \varepsilon_{jhit} \quad (19)$$

where P is the probability that an immigrant j , residing in the host country h , with ancestry i , who participated in the t th ESS round is currently employed in a trade related sector. V_i is the measure of land variability at the ancestry country i . The analysis controls for a vector of individual controls such as age, age square, gender, employment and family status, and educational level. B_i is the vector of the full set of country of ancestry controls employed in the baseline analysis. Φ_r is a vector of host country fixed effects that captures all time invariant unobserved heterogeneity at the host country level. \mathbf{T}_t is a vector of ESS round fixed effects aimed to capture round specific shocks that could affect individual responses. ε_{jri} is an individual specific error term. The standard errors are corrected for clustering at the dimension of the country of origin.⁴⁷

Building on the channel of trade established in an earlier section of the paper, one would anticipate that $\alpha_1 > 0$, i.e., that higher land variability at the origin country is associated with a higher probability that the individual is involved in trade related activities. Indeed, as will become clear below this is the obtained outcome, even after controlling for a number of individual, ancestry and host country characteristics. The underlying assumption is that land variability is associated with a culture of trade. Therefore, individuals coming from states that were reinforced by the incentives to trade are more likely to manifest a culture of trade and thus be engaged in trade activities.

Empirical Results Table 8 establishes that land variability of the origin country is associated with a higher probability that an individual (immigrant) is engaged in trade activities. Analytically, Column (1) controls for the full set of controls that were used in the baseline analysis.⁴⁸ All these controls refer to the origin country. Column (2) controls for host country fixed effects and ESS round fixed effects thus controlling for host country characteristics as well as shocks that may be associated with the timing of the survey. Column (3) further augments the analysis with a set of individual controls, such as age, gender, employment status, etc. Column (4) introduces two additional controls of the origin country

⁴⁷Double clustering at the dimension of i) the host and the origin country, and ii) the origin country and the ESS round, yields similar results (results not reported in the paper).

⁴⁸That is for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects.

that are of more relevance to the current economic situation, i.e., current income per capita (year 2000) and current volume of trade as a percentage of GDP (year 2000). The analysis suggests that in all four columns land variability at the origin country has a positive effect on the probability that the individual is employed in a trade related occupation.

TABLE 8 HERE

Column (5) restricts the sample to second generation immigrants and replicates the analysis of Column (5). The results remain intact. The restriction of the sample to that of the second generation immigrants is reassuring i) as to selection issues; and, ii) as to establishing the cultural component associated with land variability. Taking into account that a large number of the country of origin characteristics have been controlled for and that these immigrant no longer live in their origin country (and are thus not directly affected by it institutionally or financially) any traced effect is attributed to the cultural effect of the origin country.

6 Robustness-Cross Country Analysis

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.

6.1 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 9).

TABLE 9HERE

6.1.1 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 land variability, 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:

$$|\Delta I_{ij}| = \alpha_0 + \alpha_1 |\Delta V_{ij}| + \alpha_2 |\Delta X_{ij}| + \alpha_3 \delta_i + \alpha_4 \delta_j + \varepsilon_{ij} \quad (20)$$

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 bilateral 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 ($N \times (N - 1)$ pairs) used in the baseline analysis. Since the sample is symmetric, to avoid underestimating standard errors, they are clustered at the pair level.⁴⁹

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, years elapsed since the onset of the Neolithic. Importantly though, all these controls in the current analysis reflect differences between countries.

⁴⁹An 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.

Moreover the analysis controls for contiguity (countries sharing a common 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 10 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 differences in all the geographical and historical controls introduced in the baseline analysis. Column (3) introduces controls that are relative 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 horseshoe 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 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 in line with the findings of the cross country analysis.

6.1.2 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 by replicating 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).

6.2 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 7 while 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 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 of Table 7.

Analytically, Table 11 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 (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. In Column (2), 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 (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.

TABLE 11 HERE

Second, Table 12 employs the non-discounted measure of statehood, i.e. a measure that does not capture inertia from past history. The results remain quite similar. There are two main differences though. First the magnitude of the coefficients is larger, thus taking the non-discounted values of the statehood index magnifies the effect of land variability on state formation. Second, the coefficient on statehood does not completely vanish once the trade measures are introduced. It dropped significantly in magnitude yet it retains significance at the 1% level. This finding is reassuring as to the fact that land variability has a cumulative effect on state formation via trade. Only when past statehood is accounted for via discounting, and thus the persistent effect of land variability on state formation, do the coefficients drop completely (see Table 1)

TABLE 12 HERE

Last, Table 13 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 13 HERE

6.3 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 paper, this measure 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 plausibly inferred that the fractal dimension index of a country can implicitly capture how naturally the borders emerged.

TABLE 14 HERE

This is 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 in 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 any other African countries that were artificially split.

Table 14 employs a sample of 134 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 suitability⁵⁰, while controlling for the baseline set of controls 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 a more "natural" current shape of borders. Column (2) replicates the same analysis where it employs 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.

⁵⁰Fractal dimension of the country's political (non-coastline) borders using all 12 box sizes.

6.4 Validity of the Index of Diversity in Land Suitability

6.4.1 Decomposition of the Index

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.⁵¹ Column (1) of Table 15 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 16 repeats the same analysis using the measure of variability in soil quality. Both tables employ the full set of controls. Reassuringly the results in both cases remain intact, which reinforces the validity of the index.⁵²

TABLE 15 HERE

TABLE 16 HERE

6.4.2 Climatic Shocks

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

⁵¹Soil 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.

⁵²According 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.

recorded the last 2000 years., i.e. the 'Little Ice Age' (1350 CE-1850 CE), the 'Medieval Warm Period' (950 CE-1250 CE) as well as the droughts (500 CE-1500 CE). Table 17 replicates the baseline analysis while introducing fixed effects for the three climatic shocks. To construct these dummies a variety of resources have been employed.⁵³ 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.

TABLE 17 HERE

6.4.3 Alternative Measures of Land Variability: Standard Deviation and Gini Indices

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 18 suggests, replicating the baseline analysis using this alternative measure leaves the results 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.

TABLE 18 HERE

Table 19 replicates the baseline analysis using a measure of standard deviation. The coefficient is significant at the 10% level while employing the full set of controls.

TABLE 19 HERE

Overall, the results in this section suggest that other measure of land variability can give rise to similar results as to the effect of land variability on state formation. The same is true for the trade proxies as well, i.e., in all cases trade appears to be one of the prevailing channels of state formation.

6.4.4 Alternative Sources of the Land Variability Index

This section establishes the robustness of the results to the use of alternative sources of the land variability index.

Table 20 employs the FAO-GAEZ dataset. In particular the analysis employs Plate 27 which provides an index that evaluates climate, terrain slope and soil constraints. Higher

⁵³See Appendix B for an overview of the three major climatic shocks.

values of the index denote lower suitability for cultivation. The land variability measure is the range of the index.

TABLE 20 HERE

6.5 Influential Observations

6.5.1 Middle East and North Africa

A plausible concern is whether the results are driven by a number of states are all based around the fertile parts of desert. The cross country analysis takes this concern into account by controlling for i) continental fixed effects in the baseline analysis, ii) country fixed effects in the panel section, and, iii) regional fixed effects in the robustness section.

However in the context of the cross country analysis this concern is further addressed by excluding the countries that fall into the region of the Middle East and North Africa. The remaining sample is 107 countries. The results reported in Table 21 establish that our findings are not driven by early states located around the fertile parts of the dessert.

TABLE 21 HERE

An alternative approach would be to recalculate the land variability index by eliminating very low fertility area, such as deserts, tundra or taiga. Intuitively this approach would further strengthen the results. Recall that the measure of land variability is the measure of range of land suitability for agriculture, i.e., $range_{land\ suitability} = \max_{land\ suitability} - \min_{land\ suitability}$. Eliminating grid cells with low land suitability would lower the minimum values of land suitability in each country and thus this would lower the range of land suitability in each country. Therefore, for each given value of state the measure of land variability would be lower. This would be even more pronounced for high values of state formation as in the case of Middle East and Northern Africa countries. The new regression line would be steeper suggesting that deserts introduce an attenuation bias.

6.6 Alternative Specifications

Given the importance of the control on the size of a country, the analysis in this section employs several specification with variants of the control on area. Table 22 addresses this issue by controlling for several variant of the area control. Analytically Column (1) presents the baseline results, Column (2) controls for the logarithm of area, Column (3) for the logarithm of area and the quadratic term whereas Column (4) controls for area and its squared term. The results in Table 22 suggest that the magnitude and in most cases the significance of the coefficient on land suitability remain largely unaffected, whereas the coefficients on area are in all three cases insignificant.

TABLE 22 HERE

6.6.1 Outliers

This section establishes that the main results are not driven by outliers. In Table 23 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.⁵⁴ Reassuringly all the results are robust to weighting influential observations.

TABLE 23 HERE

6.7 Extended State History Data

A recent paper by Borcan et al. (2014) extended the state antiquity index till 3500 BCE. The data is not yet in the public domain therefore in the current version of this paper testing the robustness of the results to the extended dataset is not yet feasible. There are several argument why the extended dataset would not pose a threat to the current identification strategy.

First, the robustness section replicates all the results i) using the non-discounted data; and ii) using only the first component of the index (i.e., whether a state existed or not in the year 1000CE). Therefore, the argument of the paper is applicable even when not taking past history into account.

Second, as to the discounted index it could be argued that in the context of the cross country analysis, where the measure of statehood in 1000CE is considered, the results could potentially be stronger with the extended index. The reason is that, if anything, the index is downward biased particularly for older states. For instance, as the authors mention in the case of countries such as Iran or Egypt, state history extends at least 8 centuries before the year 1 CE. Therefore, one would anticipate that for the same value of the land variability index in e.g., Egypt, the value of the state history index would be higher, thus attenuating the estimated coefficient downwards.

Finally, when it come to the case of the panel analysis, the measure of climatic variability does not extend beyond the year 500 CE therefore the analysis cannot be extended before that year either.

⁵⁴This 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.

7 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 a number of important historical controls derived by dominant theories in the historical 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) and climatic shocks, 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.
- 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. (2013a). Genetic Diversity and the Origins of Cultural Fragmentation, *American Economic Review* **103**(3): 528–33.
- Ashraf, Q. and Galor, O. (2013b). 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.
- Borcan, O., Olsson, O. and Putterman, L. (2014). State History and Economic Development: Evidence from Six Millennia, *SSRN 2464285* .
- Borner, L. and Severgnini, B. (2011). Epidemic Trade, *Technical report*, School of Business and Economics Discussion Paper: 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., Cook, C. J. and Putterman, L. (2014). Persistence of Fortune: Accounting for Population Movements, There was no Post-Columbian Reversal, *American Economic Journal: Macroeconomics* **6**(3): 1–28.
- 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.
- Comin, D., Easterly, W. and Gong, E. (2010). Was the Wealth of Nations Determined in 1000 B.C.?, *American Economic Journal: Macroeconomics* **2**(3): 65–97.
- 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) .
- Fenske, J. and Kala, N. (2013). Climate, Ecosystem Resilience and the Slave Trade.
- Fernández, R. and Fogli, A. (2009). Culture: An Empirical Investigation of Beliefs, Work, and Fertility, *American Economic Journal: Macroeconomics* **1**(1): 146–177.
- 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. and Moav, O. (2002). Natural Selection and the Origin of Economic growth, *Quarterly Journal of Economics* **117**(4): 1133–1191.
- 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.
- Galor, O. and Weil, D. N. (2000). Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond, *American Economic Review* **90**(4): 806–828.

- 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. (2013a). The Geographic Origins of States, *Working Paper* .
- Litina, A. (2013b). Natural Land Productivity, Cooperation and Comparative Development, *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.
- Michalopoulos, S., Naghavi, A. and Prarolo, G. (2013). Trade and Geography in the Origins and Spread of Islam, *NBER w18438* .

- Michalopoulos, S. and Papaioannou, E. (2013). Pre-colonial ethnic institutions and contemporary african development, *Econometrica* **81**(1): 113–152.
- Nordhaus, W. D. (2006). Geography and Macroeconomics: New Data and New Findings, *Proceedings of the National Academy of Sciences* **103**(10): 3510–3517.
- Özak, Ö. (2010). The Voyage of Homo-Economicus: Some Economic Measures of Distance. Mimeo, Brown University.
- Özak, Ö. (2012). Distance to the Technological Frontier and Economic development, *Available at SSRN 1989216* .
- Peregrine, P. (2003). *Atlas of Cultural Evolution*.
- 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.
- Smith, A. D. (1986). *The Ethnic Origins of Nations*, Basil Blackwell Oxford.
- 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.
- Tilly, C., Evans, P. B., Rueschemeyer, D. and Skocpol, T. (1985). *War Making and State Making as Organized Crime*, Cambridge University Press.
- Vollrath, D. (2011). The Agricultural Basis of Comparative Development, *Journal of Economic Growth* **16**(4): 343–370.

Tables and Figures

This section present the figures and the tables in the main text.

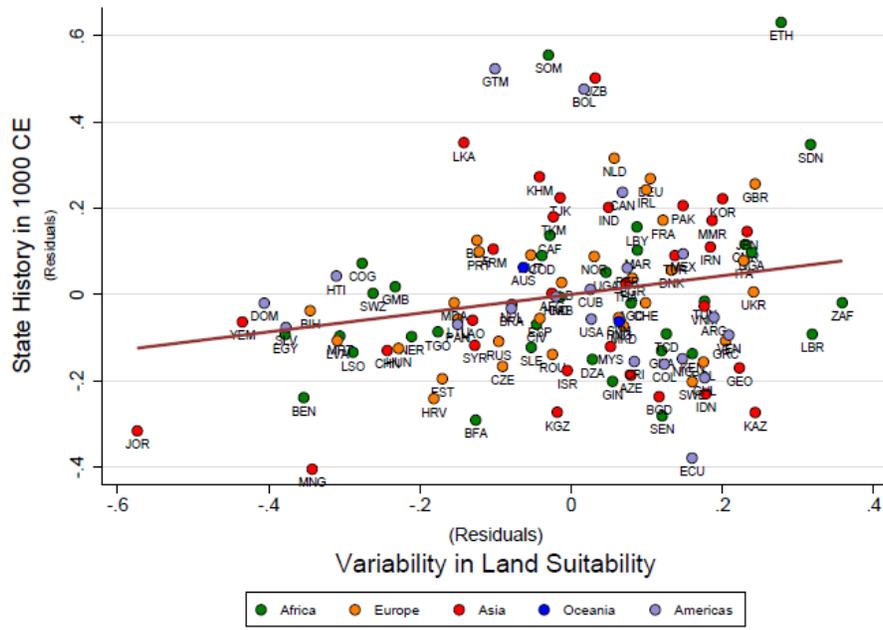


FIGURE 1: Land Variability Hypothesis-Conditional on controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population density, years elapsed since the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

TABLE 1: Main Hypothesis: The Effect of Land Variability on the Emrengence of States

| | (1) | (2) | (3) | (4) | (5) |
|--|------------------------------------|----------------------|----------------------|---------------------|---------------------|
| | Dep. Var.:State History in 1000 CE | | | | |
| | Beta Coefficients | | | | |
| Land Variability | 0.192** (0.096) | 0.313*** (0.089) | 0.279*** (0.097) | 0.272*** (0.099) | 0.202*** (0.099) |
| Average Land Suitability | | 0.763*** (0.229) | 0.638*** (0.242) | 0.645** (0.246) | 0.488** (0.246) |
| Fraction of Arable Land | | -0.468* (0.267) | -0.381 (0.298) | -0.438 (0.290) | -0.196 (0.290) |
| Suitability for Cereals | | -0.094*** (0.034) | -0.092*** (0.034) | -0.087** (0.033) | -0.311** (0.033) |
| Distance to the Nearest Coast/River | | -0.281*** (0.090) | -0.249*** (0.088) | -0.226** (0.090) | -0.346** (0.090) |
| % of Land within 100 km of Coast/River | | -0.032 (0.118) | -0.008 (0.120) | -0.009 (0.121) | -0.010 (0.121) |
| Average Ruggedness | | -0.000 (0.000) | -0.000 (0.000) | -0.000 (0.000) | -0.107 (0.000) |
| Average Elevation | | 0.066 (0.086) | 0.069 (0.084) | 0.050 (0.085) | 0.087 (0.085) |
| Total Area | | 15.664 (10.993) | 16.027 (10.862) | 14.159 (10.905) | 0.103 (10.905) |
| Absolute Latitude | | -0.003 (0.033) | 0.001 (0.029) | 0.007 (0.029) | 0.025 (0.029) |
| % Land in Tropical and Subtropical Zones | | -0.327*** (0.115) | -0.242* (0.122) | -0.224* (0.123) | -0.296* (0.123) |
| % Land Temperate Zones | | -0.343*** (0.130) | -0.267* (0.146) | -0.253* (0.145) | -0.334* (0.145) |
| Distance to Frontier in 1000 CE | | | -0.000 (0.000) | -0.000 (0.000) | -0.098 (0.000) |
| Years since Neolithic | | | | 0.028 (0.021) | 0.200 (0.021) |
| 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 | 0.594 | 0.618 | 0.624 | 0.624 |

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.

TABLE 2: Mediating Factor: Testing Whether the Effect of Variability Operates via Trade

| | (1) | (2) | (3) | (4) | (5) |
|----------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | State 1000 CE | Exch.1CE | State.1000CE | Transp. 1CE | State 1000 CE |
| Land Variability | 0.272*** (0.099) | 0.504*** (0.151) | 0.156 (0.096) | 0.314*** (0.113) | 0.163* (0.092) |
| Med. Exch. in 1 CE | | | 0.231*** (0.086) | | |
| Med. Transp. in 1 CE | | | | | 0.345*** (0.122) |
| Full Set of Contorls | Yes | Yes | Yes | Yes | Yes |
| Observations | 117 | 117 | 117 | 117 | 117 |
| R-squared | 0.624 | 0.660 | 0.658 | 0.803 | 0.663 |

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.

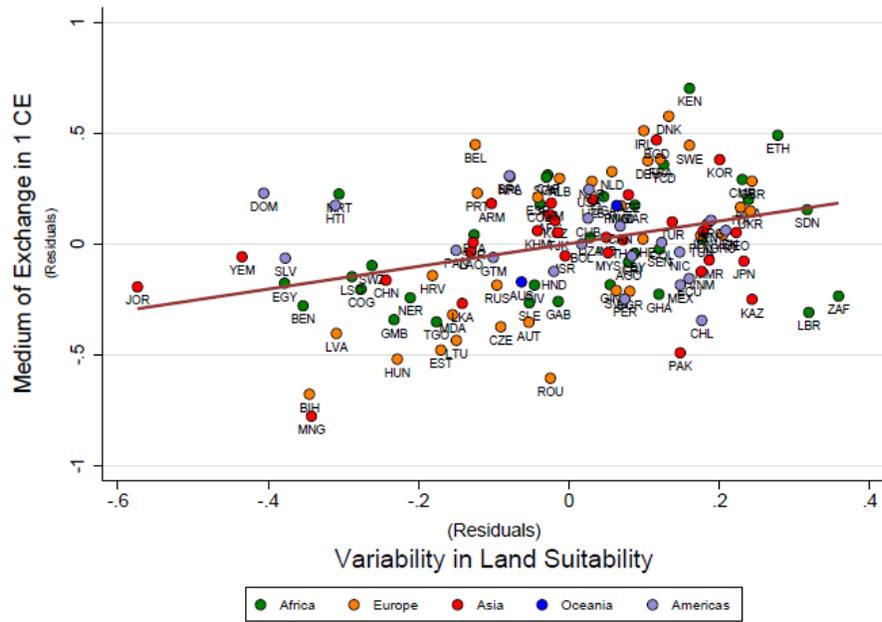


FIGURE 2: The effect of land variability on the medium of exchange technology-Conditional on controlling for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population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

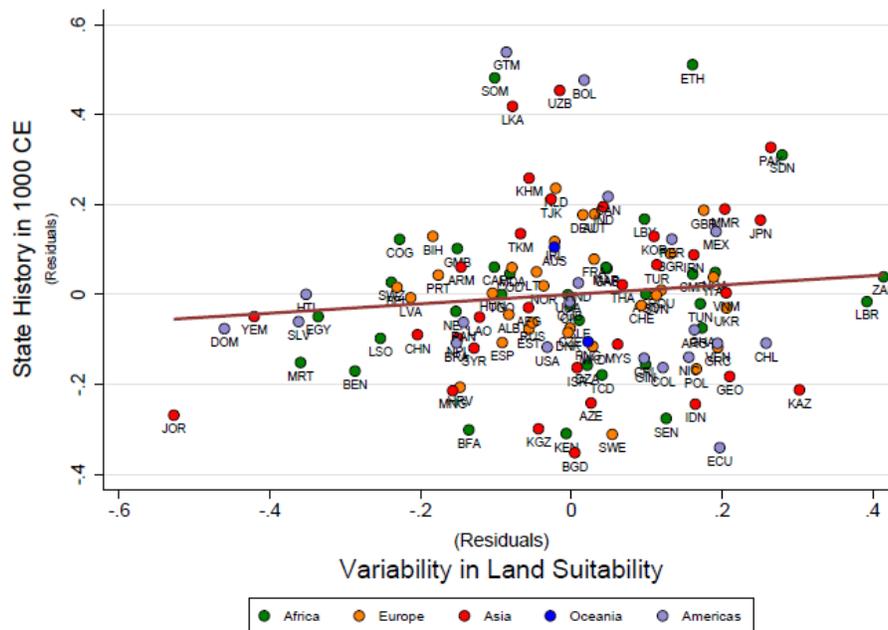


FIGURE 3: The channel of trade-Conditional on controlling for medium of exchange technology (the channel), various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, population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

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 |
| Land Variability | 0.272*** (0.099) | 0.172*** (0.059) | 0.305*** (0.089) | 0.110** (0.049) | 0.271*** (0.075) | 0.038 (0.032) |
| State History 500 CE | | 0.714*** (0.060) | | | | |
| State History 1000 CE | | | | 0.711*** (0.050) | | |
| State History 1500 CE | | | | | | 0.762*** (0.037) |
| Full Set of Controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 117 | 117 | 117 | 117 | 117 | 117 |
| R-squared | 0.624 | 0.870 | 0.729 | 0.931 | 0.677 | 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.

TABLE 4: The Effect of Climatic Variability on State Formation

| | (1) | (2) | (3) | (4) |
|------------------------------|----------------------|----------------------|----------------------|-----------------------|
| | State History | | | |
| Climatic Variability (Range) | 11.957*** (2.354) | 26.731*** (2.820) | 12.177*** (2.511) | 12.557*** (2.362) |
| Mean Temperature | | | | -17.653*** (5.535) |
| Country FE | No | Yes | Yes | Yes |
| Time FE | No | No | Yes | Yes |
| Countries | 102 | 102 | 102 | 102 |
| R-squared | | 0.188 | 0.386 | 0.407 |

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 102 countries and controls for country fixed effects, time fixed effects and a time varying measure of average temperature. .

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) 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.

TABLE 5: The Effect of Climatic Variability on State Formation

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|------------------------|
| | State History | | | | | |
| | Clim. St. Dev. | Continents | Time Trend | Country Time Trend | 300 Y Intervals | 200 Y Intervals |
| Clim. Var. (Range) | | 7.703*** (2.673) | 11.929*** (2.278) | 11.346** (5.684) | 6.301*** (2.231) | 4.215** (1.950) |
| Clim Var. (St. Dev) | 38.638*** (10.155) | | | | | |
| Mean Temp. | -18.427*** (5.597) | -17.325*** (5.758) | -16.264*** (5.290) | -25.002** (12.119) | -18.009*** (4.994) | -11.451*** (4.269) |
| Country FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Time FE | Yes | Yes | No | Yes | Yes | Yes |
| Continent Trend | No | Yes | No | No | No | No |
| Time Trend | No | No | Yes | No | No | No |
| Country Trend | No | No | No | Yes | No | No |
| Countries | 102 | 102 | 102 | 102 | 102 | 102 |
| R-squared | 0.403 | 0.474 | 0.404 | 0.642 | 0.347 | 0.319 |

Summary: This table establishes the robustness of the climatic variability panel section. The analysis controls for alternative measures of climatic variability (standard deviation from average temperature), continental time trends and country specific time trends.

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 in Columns (2)-(6) and the standard deviation of the average temperature measure in Column (1); (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: Alternative Channels of State Formation

| | (1) | (2) | (3) | (4) | (5) |
|----------------------|---------------------|---------------------|---------------------|--------------------|---------------------|
| | State 1000 CE | State.1500 CE | State 1000 CE | State 1000 CE | State 1000 CE |
| | War Channel | | Population Density | Stratification | All Channels |
| Land Variability | 0.336*** (0.106) | 0.234** (0.092) | 0.221** (0.099) | 0.243** (0.097) | 0.300*** (0.101) |
| Milit. Tech. 1 CE | 0.516*** (0.178) | | | | 0.306* (0.166) |
| Milit. Tech. 1500 CE | | 0.374*** (0.100) | | | |
| Pop. Density 1 CE | | | 0.084*** (0.027) | | 0.046 (0.028) |
| Stratification 1 CE | | | | 0.153** (0.072) | 0.167** (0.076) |
| Full Set of Controls | Yes | Yes | Yes | Yes | Yes |
| Observations | 104 | 96 | 117 | 116 | 104 |
| R-squared | 0.651 | 0.812 | 0.650 | 0.647 | 0.684 |

Summary: This table explores a number of alternative channels via which land variability could affect state formation such as war, stratification and population density.

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 variability is the range of the land suitability index; (iii) the index of military technology in the year 1CE takes the value 1 if iron is used and 0 otherwise; (iv) the index of military technology in the year 1500 CE captures the presence of standing army, cavalry, firearms, muskets, filled artillery, warfare capable ships, heavy naval guns and ships; (v) population density comes from Mc Evedy and Jones (1978) and gives the relevant estimates for the year 1 CE; (vi) social stratification in the year 1 CE is a measure of social complexity that captures the number of classes within a society. The index is assigned a value of 1 for egalitarian societies, a value of 2 for two social classes and a value of 3 for three or more social classes; (vii) the set of continent dummies includes a fixed effect for Africa, the Americas, Australia, Europe and the Middle-East (viii) a single continent dummy is used to represent the Americas, which is natural given the historical period examined; (ix) robust standard error estimates are reported in parentheses; (x) *** 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 7: The Trade Channel: International or Regional Trade?

| | (1) | (2) | (3) |
|--------------------------|---------------------|-------------------------|---------------------|
| | State 1500 | Distance to Trade Route | State 1500 |
| Land Variability | 0.421*** (0.125) | -0.208 (0.271) | 0.403*** (0.118) |
| Distance to Trade Routes | | | -0.0847 (0.0635) |
| Cont. Dummies | Yes | Yes | Yes |
| Persi.-Rom. Occup. | Yes | Yes | Yes |
| Observations | 103 | 103 | 103 |
| R-squared | 0.725 | 0.908 | 0.738 |

Summary: This table establishes that the mechanism linking land variability and state formation is not that of international trade. The analysis controls for the full set of baseline controls, i.e., for various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance to 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 variability is the range of the land suitability index; (iii) distance to trade routes is measured in weeks from the capital of each country in the Old World to trade routes used between 500BCE and 1900CE. The trade routes are long-distance arteries that extend over tens or hundreds of kilometres; (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 8: The Cultural Effect of (Origin) Land Variability on the Choice of a Trade-Related Occupation

| | (1) | (2) | (3) | (4) | (5) |
|----------------------------|--------------------------|---------|----------|---------|---------|
| | Trade Related Occupation | | | | |
| Land Variability (Origin) | 0.004* | 0.007** | 0.007*** | 0.006** | 0.008** |
| | (0.002) | (0.003) | (0.003) | (0.003) | (0.003) |
| Income Per Capita (Origin) | | | | 0.000 | 0.000 |
| | | | | (0.001) | (0.001) |
| Trade % of GDP (Origin) | | | | -0.000 | -0.000 |
| | | | | (0.000) | (0.000) |
| ESS Round FE | No | Yes | Yes | Yes | Yes |
| Host Country FE | No | Yes | Yes | Yes | Yes |
| Baseline Controls (Origin) | No | Yes | Yes | Yes | Yes |
| Individual Controls | No | No | Yes | Yes | Yes |
| Second Generation Imm. | No | No | No | No | Yes |
| No. of Origin Countries | 134 | 134 | 134 | 134 | 134 |
| No. of Host Countries | 34 | 34 | 34 | 34 | 34 |
| Obs. | 26550 | 26550 | 26550 | 26550 | 10690 |
| R-sq. | 0.000 | 0.005 | 0.007 | 0.008 | 0.011 |

Summary: This table establishes the presence of a cultural effect associated with land variability at the origin country. Analytically, higher land variability at the origin is associated with a higher probability that the immigrant is occupied in a trade related activity. The analysis controls for all the baseline controls of the origin country (various measures of agricultural suitability, proximity to waterways, average ruggedness, average elevation, total area, absolute latitude, distance from the nearest technological frontier, years elapsed since the onset of the Neolithic and fixed effects for climatic zones, Persian and Roman occupation and unobserved continental fixed effects), individual characteristics (age, age square, gender, educational level, family and employment status) as well as for ESS round and host country fixed effects.

Notes: (i) Corruption is measured by the ICRG Index. The index takes values from 0-6 with 6 indicating the most corrupt country; (ii) The variables "Trust in Parliament", "Trust in the Legal System", "Trust in Politicians" and "Trust in the Political Parties" refer to the host country and take values from 0-10 with 0 denoting "no trust at all", and 10 denoting "complete trust"; (iii) Robust standard error estimates, clustered at the dimension of the country of origin, 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.

TABLE 9: Robustness to Regional Fixed Effects

| | (1) | (2) | (3) |
|-------------------------------|--------------------|---------------------|---------------------|
| | State 1000 CE | | |
| Land Variability | 0.225** (0.102) | 0.090 (0.090) | 0.110 (0.094) |
| Medium of Exchange 1 CE | | 0.252*** (0.080) | |
| Medium of Transportation 1 CE | | | 0.379*** (0.120) |
| Full Set of Controls | Yes | Yes | Yes |
| Observations | 117 | 117 | 117 |
| R-squared | 0.612 | 0.651 | 0.658 |

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, 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 variability 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.

TABLE 10: 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 Land Variability | 0.049** (0.021) | 0.035* (0.018) | 0.033* (0.018) | 0.017 (0.017) | 0.027 (0.017) |
| Diffs in Med. Exch. 1 CE | | | | 0.191*** (0.008) | |
| Diffs. in Med. Transp. 1 CE | | | | | 0.223*** (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 | 0.317 | 0.319 | 0.380 | 0.374 |

Summary: This table establishes that differences in land variability 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, 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 variability 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.

TABLE 11: The Effect of Land Variability on State History in the Year 1500 CE

| | (1) | (2) | (3) |
|----------------------|---------------------|---------------------|---------------------|
| | State 1500 | | |
| Land Variability | 0.347*** (0.091) | 0.305*** (0.085) | 0.299*** (0.087) |
| Med. Exch. 1 CE | | 0.237** (0.113) | |
| Med. Transp. 1 CE | | | 0.232* (0.118) |
| Full Set of Controls | Yes | Yes | Yes |
| Observations | 129 | 129 | 129 |
| R-squared | 0.723 | 0.739 | 0.738 |

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, 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 variability 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 12: Mediating Factor: Testing Whether the Effect of Variability Operates via Trade

| | (1) | (2) | (3) |
|----------------------|----------------------|----------------------|----------------------|
| | State 1000 CE | | |
| Land Variability | 20.038*** (5.917) | 12.459** (4.949) | 13.100*** (4.727) |
| Med. Exch. 1 CE | | 15.051*** (5.030) | |
| Med. Transp. 1 CE | | | 22.104*** (6.890) |
| Full Set of Controls | Yes | Yes | Yes |
| Observations | 117 | 117 | 117 |
| R-squared | 0.617 | 0.658 | 0.663 |

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, 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 variability 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 13: Employing the non-Discounted Measure of State History

| | (1) | (2) | (3) |
|-------------------|---------------------|---------------------|---------------------|
| | State 1000 | | |
| Land Variability | 0.489*** (0.141) | 0.352*** (0.127) | 0.369*** (0.130) |
| Med. Exch 1 CE | | 0.267** (0.114) | |
| Med. Transp. 1 CE | | | 0.383** (0.188) |
| Cont. Dummies | Yes | Yes | Yes |
| Pers.-Rom. Occup. | Yes | Yes | Yes |
| Observations | 117 | 117 | 117 |
| R-squared | 0.703 | 0.723 | 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, 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 variability 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.

TABLE 14: Employing a Binary Index of State History

| | (1) | (2) | (3) |
|----------------------|-----------------------|---------------------|-------------------|
| | Log Fractal Dimension | | |
| Land Variability | 0.016** (0.008) | 0.013*** (0.005) | 0.006* (0.003) |
| Full Set of Controls | Yes | Yes | Yes |
| Colonial Dummies | Yes | Yes | Yes |
| Observations | 134 | 134 | 134 |
| R-squared | 0.636 | 0.634 | 0.657 |

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, 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 variability 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.

TABLE 15: Decomposition of the Index of Diversity: Climatic Component

| | (1) | (2) | (3) |
|---------------------------------------|------------|----------|----------|
| | State 1000 | | |
| Land Variability (Climatic Component) | 0.138* | 0.092 | 0.084 |
| | (0.071) | (0.066) | (0.068) |
| Med. Exch. 1 CE | | 0.251*** | |
| | | (0.084) | |
| Med. Transp.1 CE | | | 0.368*** |
| | | | (0.123) |
| Full Set of Controls | Yes | Yes | Yes |
| Observations | 117 | 117 | 117 |
| R-squared | 0.612 | 0.657 | 0.660 |

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, 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) land variability (climatic variability) 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 16: Decomposition of the Index of Diversity: Soil Component

| | (1) | (2) | (3) |
|-----------------------------------|---------------------|--------------------|--------------------|
| | State 1000 | | |
| Land Variability (Soil Component) | 0.361*** (0.113) | 0.216* (0.115) | 0.207* (0.114) |
| Med. Exch. 1 CE | | 0.219** (0.090) | |
| Med. Transp.1 CE | | | 0.327** (0.130) |
| Full Set of Controls | Yes | Yes | Yes |
| Observations | 117 | 117 | 117 |
| R-squared | 0.631 | 0.660 | 0.664 |

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, 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) land variability (soil component) 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 17: Robustness to Major Climatic Shocks

| | (1) | (2) | (3) |
|----------------------|--------------------|---------------------|---------------------|
| | State 1000 | | |
| Land Variability | 0.236** (0.101) | 0.118 (0.094) | 0.130 (0.091) |
| Med. Exch. 1 CE | | 0.276*** (0.088) | |
| Med. Transp. 1 CE | | | 0.380*** (0.118) |
| Medieval Warm Period | 0.109 (0.077) | 0.141** (0.070) | 0.134* (0.069) |
| Little Ice Age | 0.022 (0.076) | -0.075 (0.075) | -0.044 (0.077) |
| Droughts | -0.148 (0.133) | -0.130 (0.134) | -0.141 (0.128) |
| Full Set of Controls | Yes | Yes | Yes |
| Observations | 117 | 117 | 117 |
| R-squared | 0.636 | 0.678 | 0.681 |

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 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) The major shocks that are captured in this table are the 'Little Ice Age' (1350 CE-1850 CE), the 'Medieval Warm Period' (950 CE-1250 CE) and the droughts (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 18: Robustness to Alternative Measures of Land Variability-Gini

| | (1) | (2) | (3) |
|-------------------------|--------------------|---------------------|---------------------|
| | State 1000 | | |
| Land Variability (Gini) | 0.535** (0.241) | 0.255 (0.263) | 0.209 (0.254) |
| Med. Exch. 1 CE | | 0.239*** (0.090) | |
| Med. Transp. 1 CE | | | 0.357*** (0.130) |
| Full Set of Controls. | Yes | Yes | Yes |
| Observations | 117 | 117 | 117 |
| R-squared | 0.618 | 0.654 | 0.657 |

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, 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 variability 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.

TABLE 19: Robustness to Alternative Measures of Land Variability-Standard Deviation

| | (1) | (2) | (3) |
|---------------------------------------|------------|----------|----------|
| | State 1000 | | |
| Land Variability (Standard Deviation) | 0.642* | 0.338 | 0.341 |
| | (0.341) | (0.319) | (0.313) |
| Med. Exch. 1 CE | | 0.242*** | |
| | | (0.087) | |
| Med. Transp. 1 CE | | | 0.358*** |
| | | | (0.123) |
| Full Set of Controls. | Yes | Yes | Yes |
| Observations | 117 | 117 | 117 |
| R-squared | 0.618 | 0.654 | 0.657 |

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, 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 variability is the standard deviation from the mean value 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 20: Robustness to Alternative Sources of Land Variability-FAO GAEZ

| | (1) | (2) | (3) |
|----------------------------------|--------------------|---------------------|---------------------|
| | State 1000 | | |
| Land Variability (FAO GAEZ) | 0.064** (0.027) | 0.042 (0.028) | 0.041 (0.027) |
| Mean Land Suitability (FAO GAEZ) | 0.037 (0.064) | 0.007 (0.064) | -0.005 (0.064) |
| Med. Exch. 1 CE | | 0.236*** (0.085) | |
| Med. Transp. 1 CE | | | 0.346*** (0.126) |
| Full Set of Controls. | Yes | Yes | Yes |
| Observations | 117 | 117 | 117 |
| R-squared | 0.590 | 0.627 | 0.630 |

Summary: This table establishes the robustness of the results to the use of an alternative source of land variability (FAO GAEZ). 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, 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 (FAO GAEZ) is a geospatial index of the suitability of land for agriculture based on climate, soil and terrain slope constraints. Larger values of the index denote lower suitability for agriculture due to high constraints; (ii) land variability is the range of the land suitability index (FAO GAEZ); (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 21: Robustness: Eliminating Middle East and North Africa

| | (1) | (2) | (3) |
|----------------------|-------------------|---------------------|---------------------|
| | State 1000 CE | | |
| Land Variability | 0.213* (0.121) | 0.080 (0.115) | 0.101 (0.116) |
| Med. Exch. 1 CE | | 0.233*** (0.086) | |
| Med. Transp. 1 CE | | | 0.339*** (0.122) |
| Full Set of Controls | Yes | Yes | Yes |
| Observations | 107 | 107 | 107 |
| R-squared | 0.608 | 0.646 | 0.650 |

Summary: This table establishes the robustness of the results to the exclusion of the countries located in the Middle east and North Africa. 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, 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 variability 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 22: Robustness to Alternative Specifications of the Country Size Control

| | (1) | (2) | (3) | (4) |
|-------------------------------------|---------------------|-------------------|-------------------|----------------------|
| Dep. Var.: State History in 1000 CE | | | | |
| Land Variability | 0.272*** (0.099) | 0.224* (0.119) | 0.247** (0.12) | 0.266*** (0.1) |
| Total Area | 14.159 (10.905) | | | 22.971 (28.696) |
| Log Total Area | | 0.026 (0.023) | 0.123 (0.101) | |
| Log Total Area Squared | | | 0.007 (0.006) | |
| Total Area Squared | | | | -678.41 (1996.75) |
| Full Set of Controls | Yes | Yes | Yes | Yes |
| Observations | 117 | 117 | 117 | 117 |
| R-squared | 0.624 | 0.623 | 0.639 | 0.624 |

Summary: This table establishes that the baseline analysis is robust to a number of alternative specifications of the control on the size of a country. 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, 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 variability 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.

TABLE 23: Robustness to Influential Observations

| | (1) | (2) | (3) |
|----------------------|--------------------|---------------------|---------------------|
| | State 1000 | | |
| Land Variability | 0.232** (0.091) | 0.108 (0.077) | 0.156* (0.081) |
| Med. Exch. 1 CE | | 0.211*** (0.053) | |
| Med. Transp. 1 CE | | | 0.247*** (0.080) |
| Full Set of Controls | Yes | Yes | Yes |
| Observations | 117 | 117 | 117 |
| R-squared | 0.705 | 0.806 | 0.785 |

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, 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 variability 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.

Appendices

A Historical Maps

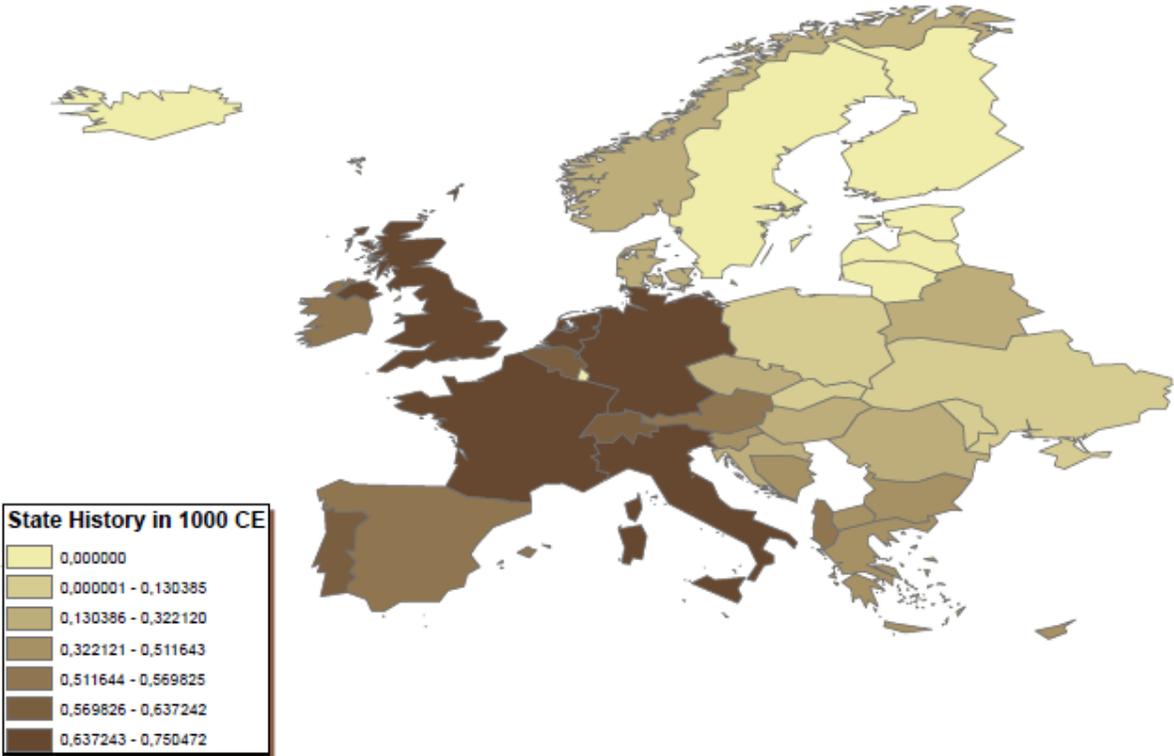


FIGURE 1: State History in Europe (1000 CE)

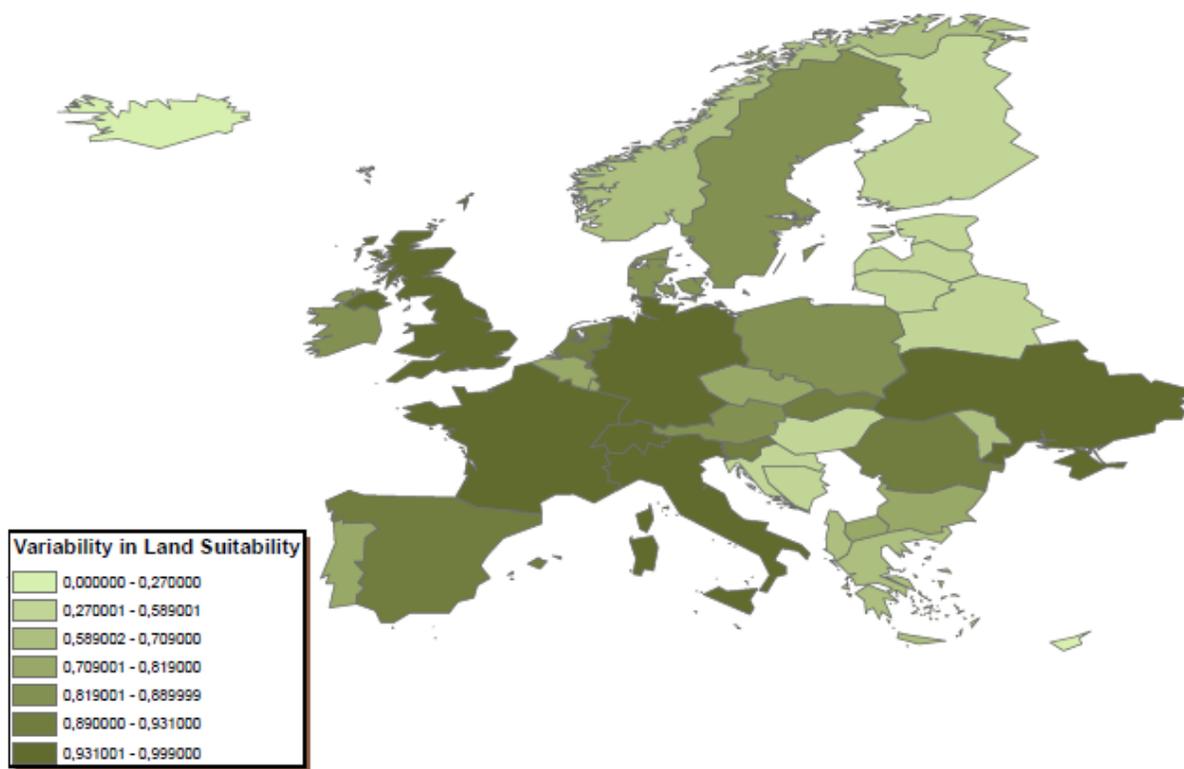


FIGURE 2: Variability in Land Suitability for Agriculture in Europe

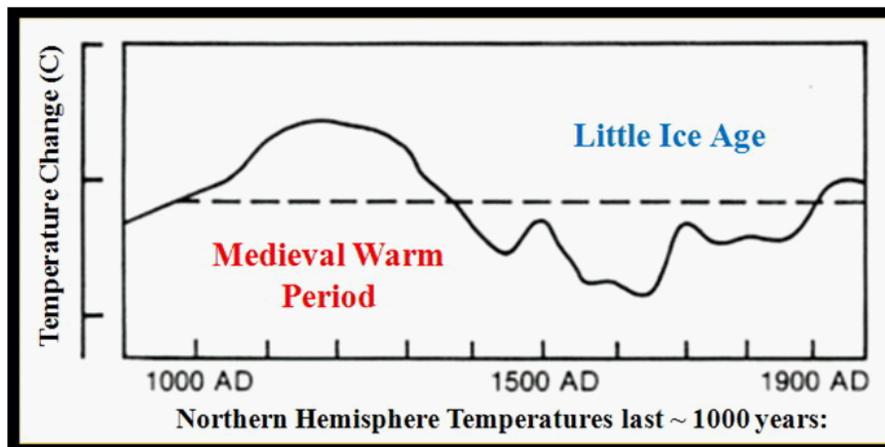
B Major Climatic Shocks

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 Variable Definitions and Sources

C.1 Cross Country Analysis Variables

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:

$$I_{SH} = I_G \times I_{FL} \times I_T$$

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?

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

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

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

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

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

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 (S_{clim}) and soil suitability (S_{soil}), i.e., $S = S_{clim} S_{soil}$. 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., $S_{clim} = f_1(\text{GDD})f_2(\alpha)$. The soil suitability component, on the other hand, is estimated to be a function of soil carbon density (C_{soil}) and soil pH (pH_{soil}), i.e. $S_{soil} = g_1(\text{C}_{soil})g_2(\text{pH}_{soil})$. The functions, $f_1(\text{GDD})$, $f_2(\alpha)$, $g_1(\text{C}_{soil})$, and $g_2(\text{pH}_{soil})$ are chosen by Ramankutty et al. (2002) by empirically fitting functions to the observed relationships between cropland areas, GDD, α , C_{soil}, and pH_{soil}. 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 classified 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 Puterman (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.

Social Stratification. Social Stratification is a measure of social complexity and captures the number of classes within a society. It is constructed from Peregrine's (2003) *Atlas of Cultural Evolution*. The level of stratification is indexed as follows. The index is assigned a value of 1 for egalitarian societies, a value of 2 for two social classes and a value of 3 for three or more social classes. The index is constructed for the year 1 CE.

C.2 Panel Analysis Variables

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). See above for more details. The analysis employs the index for every 500 year interval, i.e, for the years 500 CE, 1000 CE, 1500 CE and 1950 CE.

Geographical Variables

Climatic Variability. The measure of average temperature 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 tree-ring, ice core, coral, sediment and other assorted proxy records. The measure is defined as anomalies relative to the 1961–1990 reference period mean.

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. Using GIS software, the climatic conditions data are constructed for a panel of 97 countries, for which the data is available, covering the averages of four time periods from 500 CE till 1950 CE.

Climatic Variability. The measure of interest is climatic variability is the range of the average temperature of Mann et al. (2009) within a country.

C.3 ESS Variables

Outcome Variables

Trade Related Occupation. "Trade Related Occupation." is a binary variable that takes the value 1 if the individual is employed in a trade related occupation and 0 otherwise. The codes that are classified as a trade related occupation are the following: "1420", "3324", "1224", "1314", "3420", "3421" (derived by the third version of the International Standard Classification of Occupations-ISCO88).

Individual Controls

Age. The age of the respondent.

Gender. The gender of the respondent.

Family Status. Family status is a binary variable taking the value 0 if the individual lives with a partner and 1 otherwise.

Employment Status. Employment status is a binary variable taking the value 0 if the individual is employed and 1 otherwise.

Level of Education. The higher level of education attained by the respondent. The questionnaire distinguishes seven different levels of education (less than lower secondary, lower secondary, lower tier upper secondary, upper tier upper secondary, advanced vocational, lower tertiary BA level, higher tertiary > MA level).

Parental and Spouse Educational Level. The higher level of education attained by the respondents' father, mother and spouse. The questionnaire distinguishes seven different levels of education (less than lower secondary, lower secondary, lower tier upper secondary, upper tier upper secondary, advanced vocational, lower tertiary BA level, higher tertiary > MA level).

Individual Income. Individual income measures the reported income of the immigrant. The variable has 12 gradations.