Essays on Cultural and Historical Determinants of Economic Development

Gian Luca Tedeschi

Thesis submitted to the University of Nottingham
for the Degree of Doctor of Philosophy

Supervised by:
Abigail Barr & Roberto Bonfatti

School of Economics
University of Nottingham
January 2021

Acknowledgements

First and foremost, I would like to thank my supervisors Abigail Barr and Roberto Bonfatti for their invaluable help and guidance. Many words of gratitude also go to the faculty, students and administrative staff of the School of Economics, who have accompanied me through the five years I spent at the University of Nottingham. I am also grateful to Oded Galor and the people at Brown University for welcoming me during my visit, which provided inspiring discussions and helped me shape my research. This thesis would not exist without funding from the Economic and Social Research Council, which I gratefully acknowledge. Finally, special thanks go to my family, friends and my partner Monika for their never-ending support.

Abstract

This thesis is composed of three chapters, presenting research at the intersection of cultural economics, economic history and political economy. Relying on historical and contemporary data sources, and through rigorous empirical analysis, the thesis explores the origins and persistent influence of historical characteristics on economically-relevant features across the world, within countries and across individuals.

Chapter 1 advances the hypothesis that precolonial state institutions had an impact on attitudes towards modern-day state institutions in African countries. To address the question empirically, I combine contemporary individual-level survey data with historical data on precolonial political centralisation by ethnic group. By employing an identification strategy based on the current location of survey respondents and the disease environment of the ethnic groups' historical homelands, I determine that individuals belonging to ethnic groups that were characterised by a precolonial state show significantly higher levels of trust in modern-day state institutions. By examining a variety of potential mechanisms and alternative explanations, and through numerous robustness checks, I establish the role played by culturally embodied norms and beliefs transmitted intergenerationally within ethnic groups in determining the impact of precolonial states.

Chapter 2 studies the role played by politics in shaping the Italian railway network, and its impact on long-run growth patterns. Examining a large state-planned railway expansion that took place during the second half of the 19th century in a recently unified country, we first study how both national and local political processes shaped the planned railway construction. Exploiting close elections, we show that a state-funded railway line is more likely to be planned for construction where the local representative is aligned with the government. Furthermore, our analysis indicates that the actual path followed by the railways was shaped by local pork-barreling, with towns supporting elected candidates more likely to see a railway crossing their territory. Finally, we explore the long-run effects of the network expansion on economic development. Employing population censuses for the entire 20th century, we show that politics at a critical juncture played a key role in explaining the long-run evolution of local economies.

Chapter 3 studies the joint evolution of kin ties and state institutions across the world and over time. First, I study the geographical forces that have historically driven the evolution of both kin ties and state institutions, based on the assumption that historical societies featured an equilibrium between kin ties and formal institutions reflecting the scope of prevailing economic relations. I exploit historical ethnographic data and combine it with measures of climate intertemporal volatility and spatial variability to establish that the need to protect from weather shocks historically led to weaker kin ties and more developed state institutions. Second, I explore the long-run persistence of historical differences in kin ties and state institutions. Employing country-level institutional quality measures from the 19th century to the present and linking modern countries' populations with the characteristics of their historical ancestors, I determine that countries whose populations' ancestors were characterised by more developed state institutions and weaker kin ties are associated with modern institutions of significantly higher quality.

Contents

List of Figures

List of Tables

Chapter II - The Political Economy of the Italian Railway Expansion, 1879-1890

CHAPTER I

STATE HISTORY, INTERGENERATIONAL TRANSMISSION AND INSTITUTIONAL TRUST IN AFRICA

1 Introduction

Recognition of the state's legitimacy and confidence in its institutions are fundamental requirements for its effective functioning (Lipset, 1959; Almond and Verba, 1963; Easton, 1965; Putnam, 1993, 2000; Fukuyama, 1995). It is commonly accepted that trust in the state and its representatives is determined by current political, economic and institutional features; nonetheless, ample research in economics has documented that historical factors can not only have an impact on current economic and institutional outcomes but also shape persistent cultural traits and impact current attitudes and values (Alesina, Giuliano, and Nunn, 2013; Nunn and Wantchekon, 2011; Becker, Boeckh, Hainz, and Woessmann, 2016; Guiso, Sapienza, and Zingales, 2016; Giuliano and Nunn, 2017). Although since the late 19th century, virtually the entire world's land and population has been separated into states, there is a great deal of variation across the world in the processes that led to state formation. In this aspect, Africa stands out since hundreds of ethnic groups that pursued very different lifeways as recently as five generations ago currently live together in the setting of colonially established modern countries. Recent studies examined the long-run impact of precolonial states on economic development (Gennaioli and Rainer, 2006, 2007; Michalopoulos and Papaioannou, 2013) and the far-reaching consequences of the creation of artificial states during the Scramble for Africa (Alesina, Easterly, and Matuszeski, 2011; Michalopoulos and Papaioannou, 2016). What impact did the ethnic groups' precolonial state institutions, characterised by a significant heterogeneity that ranged from acephalous societies to states with a centralised and hierarchical organisation, have on the attitudes of individuals with respect to modern states?

In this paper, I advance the hypothesis that African ethnic groups' precolonial state institutions shaped attitudes towards the modern state through culturally embodied beliefs transmitted intergenerationally within ethnic groups. Combining survey data from African countries with historical ethnographic data, I establish that modern-day individuals whose ancestors were exposed to precolonial state institutions show overall higher confidence in current state institutions. More specifically, I employ data from the 2005 and 2008 rounds of the Afrobarometer surveys, providing information on attitudes and socio-economic characteristics of respondents, to construct an Institutional Trust Index (ITI) based on a series of questions on trust in state institutions. To measure ethnic groups' precolonial state institutions, I employ the database developed from Murdock (1959, 1967), mapping the spatial distribution of ethnic groups in Africa and quantifying their economic, political and cultural characteristics. Specifically, by using the "Jurisdictional Hierarchies Beyond the Local Community Level" index from the *Ethnographic Atlas*, I construct a measure categorising ethnic groups as either politically centralised or stateless in the precolonial period.

To isolate a causal effect of the ethnic groups' precolonial political centralisation on modern-day individuals' institutional trust, I address three significant challenges to identification. Of these challenges, the first relates to contemporary and historical factors potentially driving the estimated relationship between the ethnic group's precolonial political centralisation and the individual's institutional trust. To deal with this issue, first, I include a series of controls relating to the individuals' socio-economic characteristics, namely age, gender, education, religion and living conditions. Second, I include country fixed effects, accounting for time-invariant, country-specific features. Third, I account for subnational factors, such as the different conditions of urban areas, presence of public goods, ethnic fractionalization, presence of majority and minority groups and distance from the country capital. Fourth, I account for historical ethnic group features such as settlement structure, agricultural dependence, intensity of agriculture, relevance of indigenous slavery, number of slaves taken from the ethnic group during the slave trade and geographic features of the area historically inhabited by the ethnic group, that is absolute latitude, mean temperature and precipitation, ruggedness, distance to waterways and agricultural productivity of the land. Furthermore, to account for unobservable ethnicity-related factors, I employ cultural province indicators capturing spatial, cultural and genealogical clusters of ethnic groups' features. Fifth, to control for the potentially confounding influence of colonisation on the measures of political development contained in the *Ethnographic Atlas*, I employ time of ethnographic record indicators, allowing for a within-period comparison isolating unobserved factors that might be related with earlier or later ethnographic records.

The second empirical challenge relates to historical state formation also influencing current economic and institutional development, which, in turn, may have influenced trust in state institutions. To isolate the effect of precolonial political centralisation on current development, I employ an epidemiological approach exploiting cultural beliefs and values internal to the individual. Using Murdock's (1959) ethnolinguistic map of Africa to determine the area ethnic groups historically inhabited and employing georeferenced data on the location of respondents from the Afrobarometer surveys, I determine whether individuals are living on the historical homeland of their ethnic group. By exploiting individuals living outside of their ethnic group's historical homeland I mitigate the confounding effect of historical state formation working through current economic and institutional development, and capture the culturally embodied impact of the ethnic group's precolonial political centralisation.

The third empirical challenge relates to precolonial political centralisation, as constructed from the *Ethnographic Atlas*, being an imprecise measure of the ethnic groups' state history, which could lead to an attenuation bias. The source of measurement error is due to the information contained in *Ethnographic Atlas* indicating the presence of a precolonial state, but not how long it existed, thus potentially biasing the results towards zero. To deal with this bias, I exploit data on the historical disease environment and instrument my measure of precolonial political centralisation with an environmental feature that potentially shaped historical state formation in Africa, namely the presence of the tsetse fly. As argued by Alsan (2015), the presence of the tsetse fly affected African precolonial agricultural practices, patterns of subsistence, population density and the probability of ethnic groups being politically centralised. I thus employ a measure of tsetse presence on the ethnic group's historical homeland, obtained by combining data on the suitability of the environment to the tsetse fly with Murdock's (1959) ethnolinguistic map of Africa, as an instrument for the ethnic group's precolonial political centralisation. Therefore, by exploiting one of the factors determining whether ethnic groups had favourable conditions for early state formation, I aim to retrieve the true effect of precolonial states. Furthermore, the instrumental variable strategy allows me to deal with any residual omitted factors that may cause the political centralisation measure to be endogenous.

Through the analysis, I establish that individuals belonging to ethnic groups that were politically centralised in the precolonial period show significantly higher levels of trust in current state institutions. This result is consistent across OLS and IV estimates, although OLS yields estimates that appear to be biased downwards. Moreover, I obtain similar patterns when I focus on the subsample of individuals living outside their ethnic group's historical homeland, with estimates being qualitatively identical to the full sample estimates. Overall, findings are consistent with the initial hypothesis of ethnic groups' precolonial state institutions shaping attitudes of individuals towards modern-day state institutions.

To establish whether culturally embodied norms and beliefs transmitted intergenerationally within ethnic groups played a pivotal role in determining the impact of precolonial state institutions, I rule out a number of potential alternative mechanisms and explanations. First, I examine whether the uncovered effect of precolonial political centralisation is a function of the power held by an individual's ethnic group in the country. In other words, if ethnic groups that were politically centralised in the precolonial period are more likely to hold power in modern African countries, individuals may be more trusting of state institutions because positions of power are more likely to be occupied by co-ethnics. Employing the Ethnic Power Relations (EPR) dataset, providing country-level time-varying data on politically relevant ethnic groups and their access to executive state power, I construct variables denoting the power status of the individual's ethnic group. Results show that belonging to a group that holds a significant amount of power in the country indeed increases the individual's institutional trust but without affecting the estimated impact of the ethnic group's precolonial political centralisation. Second, I study the role played by horizontal transmission of beliefs, both between and within ethnic groups. Constructing a measure of the average precolonial political centralisation of individuals living in the same location as the respondent but belonging to other ethnic groups, I find no significant evidence of between-groups horizontal transmission. To establish the role played by within-group horizontal transmission, I interact the ethnic group's precolonial political centralisation with the share of the district's population belonging to the same ethnic group as the respondent. Results reveal a heterogeneous effect of co-ethnics presence, with the impact of precolonial state history increasing with the proportion of individuals belonging to the same ethnic group. Third, to account for the potential selection of individuals into more developed areas, I employ a series of current ethnic homeland fixed effects to obtain a within-location comparison. Results from regressions comparing individuals who are subject to closely similar economic and institutional conditions confirm the impact of precolonial political centralisation.

I then pursue a number of strategies to determine whether the results are robust and imply a causal relationship. First, I undertake a falsification exercise to determine whether the measure of the historical disease environment I employ as instrument, tsetse fly presence, is capturing the effect of unobserved factors related to the colonial rule. Drawing from the literature on how colonisation strategies were affected by settlers' mortality, I employ a measure of the prevalence of malaria in the historical ethnic homeland as an instrument for the group's precolonial political centralisation. The stark difference in results obtained from the two instrumental variable strategies provides suggestive evidence on how tsetse and malaria are two fundamentally different aspects of the disease environment in terms of their relation to precolonial states and colonisers' strategy. Second, I verify the robustness of results to changing the classification of precolonial political centralisation. Using Murdock's (1967) original classification of political institutions, I find that estimation results are essentially unaltered. Third, I employ an alternative proxy for precolonial states. Using data from Chandler (1987) on the location of African cities in 1800, I employ an indicator for the presence of a large city on the area inhabited by the ethnic group. Estimation results from using this alternative measure are comparatively similar to the ones obtained using the precolonial political centralisation variable. Fourth, I show the robustness of results when employing an alternative version of the Institutional Trust Index based on an enlarged set of survey questions, and additionally run separate regressions employing the single institutional trust survey questions, confirming baseline results. Fifth, to ensure there are no inference issues deriving from my treatment of standard errors, I employ a number of clustering techniques accounting for various degrees and forms of spatial correlation, finding no relevant differences. Finally, I restrict the comparison to individuals within the same region or interviewed in the same survey round through fixed effects specifications, showing a robust estimated impact of the ethnic group's precolonial political centralisation on the individual's institutional trust.

This paper builds on different strands of the literature. First, I contribute to the broader literature on the relationship between historical factors and cultural norms (see Nunn, 2012 and Gershman, 2017 for reviews). Among this diverse literature, this article adds to a large number of empirical studies arguing for the persistence of cultural norms over long periods of time and how these norms act as a mechanism through which historical factors shape modern differences in economic outcomes (Nunn, 2008; Nunn and Wantchekon, 2011; Voigtländer and Voth, 2012; Alesina, Giuliano, and Nunn, 2013; Giuliano and Nunn, 2013; Becker, Boeckh, Hainz, and Woessmann, 2016; Michalopoulos, Putterman, and Weil, 2019). My paper contributes to this line of work by uncovering long-run determinants of confidence in state institutions. The results in this paper also complement the literatures studying the roots of state formation (Litina, 2014; Alsan, 2015; Mayshar, Moavz, Neeman, and Pascali, 2016; Galor and Klemp, 2017) and the impact of historical states on a variety of modern outcomes, such as economic development (Englebert, 2000; Gennaioli and Rainer, 2006, 2007; Michalopoulos and Papaioannou, 2013), cultural norms (Putnam, 1993; Guiso, Sapienza, and Zingales, 2016; Lowes, Nunn, Robinson, and Weigel, 2017) and conflict (Depetris-Chauvin, 2015; Heldring, 2018). In particular, I provide evidence of the impact of precolonial states on how individuals relate to modern states. I also contribute to a group of studies measuring cultural differences across societies with different methodologies, such as using lab experiments (Henrich, 2004; Henrich, Boyd, Bowles, Camerer, Fehr, Gintis, McElreath, Alvard, Barr, Ensminger, et al., 2005) or studying natural settings where people from different cultural backgrounds face the same decision in the same environment (Fernandez, 2007; Giuliano, 2007; Fernandez and Fogli, 2009; Algan and Cahuc, 2010; Giavazzi, Petkov, and Schiantarelli, 2014). I build on this literature by studying individuals with very diverse ethnic backgrounds as fellow citizens of modern African countries. Another relevant body of research to which I relate studies the relationship between culture and institutions (Greif, 1994; Tabellini, 2008a,b, 2010; Alesina and Giuliano, 2015; Bisin and Verdier, 2017), with this paper examining the long-run relationship between ethnicity-related attitudes and precolonial institutions. On a broader scale, this work relates to the literature on the historical institutional origins of contemporary development (Diamond, 1997; Acemoglu, Johnson, and Robinson, 2001, 2002; Bockstette, Chanda, and Putterman, 2002; Dell, 2010; Michalopoulos and Papaioannou, 2014).

The remainder of the paper is organised as follows. Section 2 describes the data. Section 3 describes the analytical framework. Section 4 presents main results, addresses alternative mechanisms and explanations, and provides robustness checks. Section 5 provides a summary of the arguments and concluding remarks.

2 Data

2.1 Afrobarometer Surveys

I employ individual-level data from the third and fourth rounds of the Afrobarometer surveys, conducted between 2005 and 2008. Afrobarometer is an independent and non-partisan research project conducted by the Ghana Centre for Democratic Development (CDD), the Institute for Democracy in South Africa (IDASA), the Institute for Empirical Research in Political Economy (IERPE) with support from Michigan State University (MSU) and the University of Cape Town, Centre for Social Science Research (UCT/CSSR).[1] Implemented by national partners, Afrobarometer measures economic conditions and the political atmosphere in African countries. The questionnaire is standardised to facilitate comparison between the countries covered. These nationally representative surveys are based on interviews conducted in the local languages on a random sample of either 1,200 or 2,400 individuals per country. The third round of surveys covers the following 18 countries: Benin, Botswana, Cape Verde, Ghana, Kenya, Lesotho, Madagascar, Malawi, Mali, Mozambique, Namibia, Nigeria, Senegal, South Africa, Tanzania, Uganda, Zambia, Zimbabwe. The fourth round of surveys is conducted in the same countries as the previous round, with the addition of Burkina Faso and Liberia. A map of the 20 countries included in the fourth round of surveys is shown in Figure C.1.

Pooling the two rounds provides a potential sample of 53,110 respondents. For the sake of the analysis, only individuals matchable to an ethnic group for which ethnographic data is available (see section 2.2) are used. Therefore, individuals belonging to one of the following categories were removed: (i) respondent listed "other" as his ethnicity, (ii) respondent listed his country as his ethnicity, (iii) the ethnicity is not an indigenous African ethnicity, (iv) the ethnicity could not be cleanly matched to the ethnographic data or (v) the ethnicity could be matched but no ethnographic data is available. Additionally, for Cape Verde, the ethnicity of the respondent is not recorded in the surveys. Finally, due to the lack of data on tsetse suitability (see section 2.3) for Madagascar, this country is also dropped to keep the sample balanced throughout the analysis. The aforementioned data limitations result in a potential sample of 37,287 respondents.

To measure individual attitudes towards the state, I consider the reported trust with respect to four institutional figures: the president, the parliament, the police and courts of law. For

[1]The interested reader can refer to the Afrobarometer website at Afrobarometer.org.

these trust questions, respondents were asked the following: "How much do you trust each of the following? The President / The Parliament / The Police / Courts of Law". The answer categories were (i) "Not at all", (ii) "Just a little", (iii) "Somewhat", and (iv) "A lot". Respondents also had the option of answering "Don't know". Furthermore, for a number of respondents, the answer was reported as "Missing". Respondents with "Don't know" or "Missing" answers for any of the preceding trust questions are dropped from the sample. Figure 1 illustrates the distribution of responses to each question. These four variables enable me to consider the attitudes of respondents with respect to representatives of the different powers of the state, both at a national level (President, Parliament) and at a more local level (Police, Courts of Law). Moreover, to capture a common underlying determinant, I aggregate these four measures by employing principal component analysis and extract the first principal component. I normalise the resulting variable, Institutional Trust Index (ITI), to be in $[0, 1]$. This index constitutes the dependent variable which I will use in the empirical analysis. Appendix C.1 provides details concerning construction and distribution of the index. Due to "Don't know" or "Missing" answers to the single trust questions, the Institutional Trust Index has non-missing values for 33,419 respondents.

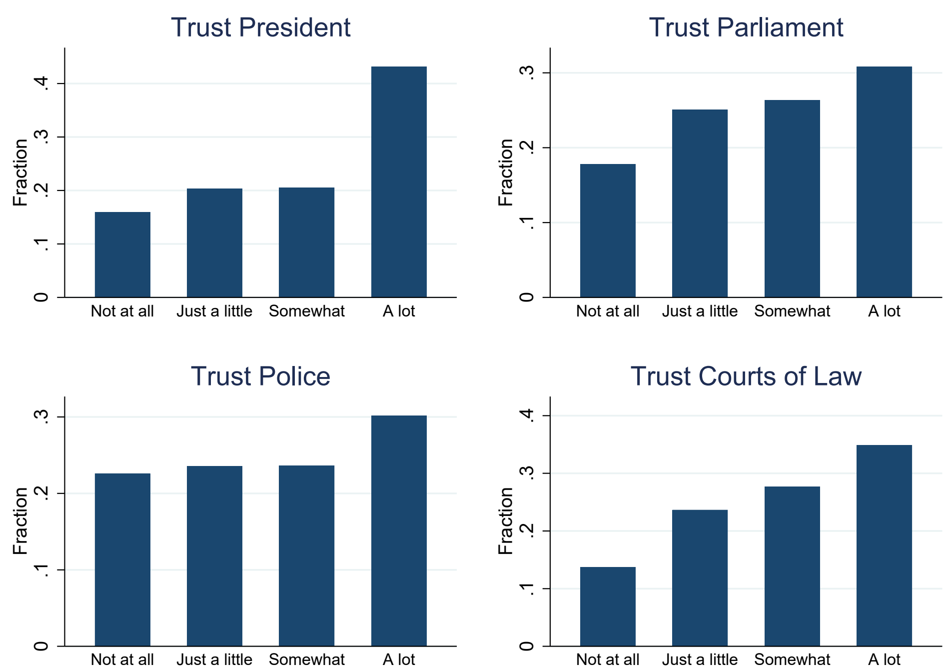


FIGURE 1. Distribution of responses to trust questions from the Afrobarometer surveys

2.2 Precolonial Political Centralisation

I employ historical data on African ethnic groups based on the work of the anthropologist George Peter Murdock. Figure 2(A) shows Murdock's (1959) ethnolinguistic map of Africa, which portrays the spatial distribution of ethnicities across Africa at the beginning of European colonisation in the mid- to late 19th century, including 843 historical ethnic homelands.

Following Nunn and Wantchekon (2011) and Deconinck and Verpoorten (2013), I match the ethnicities as reported by respondents in the Afrobarometer surveys with the classification of ethnic groups constructed and mapped by Murdock (1959). In his work following the mapping of African ethnicities, Murdock (1967) produced an *Ethnographic Atlas* measuring cultural, geographical and economic characteristics of 1,265 ethnicities around the world, of which 534 are in Africa.

Given the lack of a perfect match of ethnicities in the ethnolinguistic map and the *Ethnographic Atlas* sources, in some cases due to different naming, I employ the matching algorithm developed by Fenske (2013), which joins unmatched ethnic groups based on alternative name, supergroup or location.[2] The precolonial political centralisation measure I employ in the analysis is based on Murdock's (1967) Jurisdictional Hierarchy Beyond the Local Community Level index, or jurisdictional hierarchy for short, which is the standard variable referred to in the literature as a proxy for the institutional development of historical ethnic groups (Gennaioli and Rainer, 2006; Michalopoulos and Papaioannou, 2013). This ordered variable classifies ethnic groups into five categories, ranging from 0 to 4, indicating the number of political jurisdictions beyond the local level. Jurisdictional hierarchy is coded so that a score of 0 indicates precolonial stateless societies "lacking any form of centralised political organisation", a score of 1 indicates petty chiefdoms, a score of 2 indicates large chiefdoms/petty states, and scores of 3 or 4 indicate large states.

Figure 2(B) provides a visual representation of the ethnic groups' jurisdictional hierarchies according to the classification in the *Ethnographic Atlas*. While Murdock assembled this source by relying on the previous work of various ethnographers, which should prevent systematic bias arising from his own predispositions, the classification of political jurisdictions possibly suffers from some degree of subjectivity and measurement error. Following the literature, in particular Gennaioli and Rainer (2006, 2007), I consider an ethnic group to be politically centralised if the jurisdictional hierarchy variable has a score greater than 1. Therefore, I code precolonial political centralisation as a binary variable where a value of 1 indicates centralised ethnic groups and 0 indicates non-centralised, or stateless, ethnic groups.

[2] The file can be found in the web Appendix of Fenske (2013), available on his website.

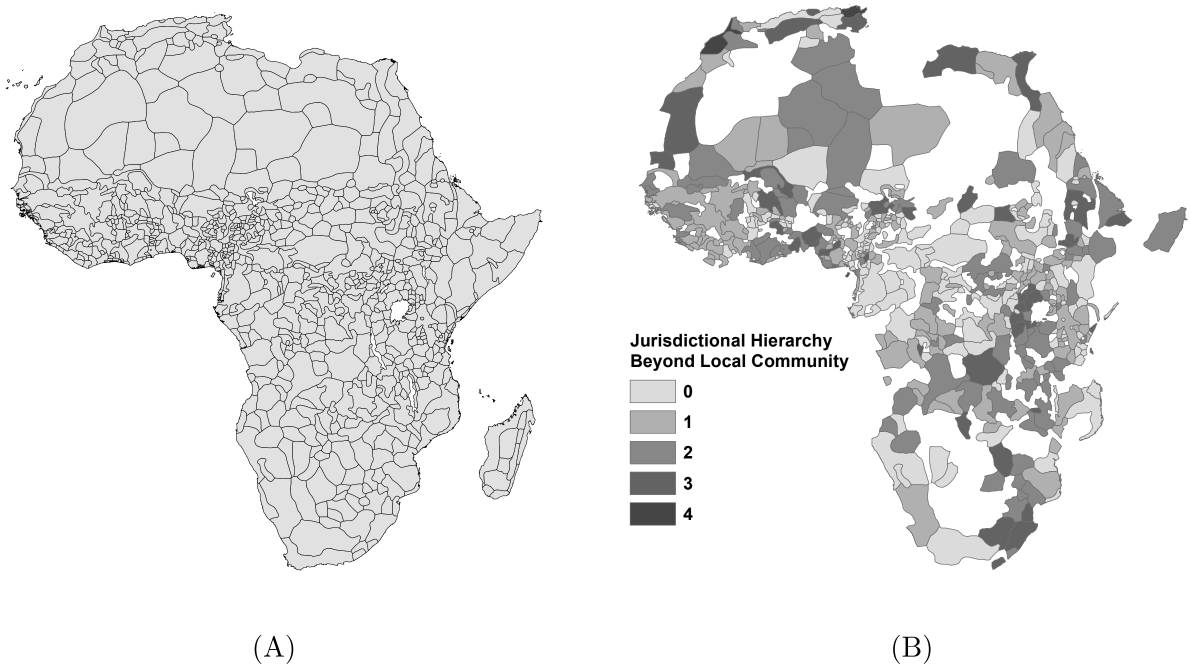


FIGURE 2. (A) Historical ethnic homelands from Murdock's (1959) ethnolinguistic map. (B) Precolonial centralisation from Murdock's (1967) *Ethnographic Atlas*.

2.3 Tsetse Suitability Index

As part of my identification strategy, I employ data on the historical disease environment, one of the factors through which the natural environment shaped historical institutions (Engerman and Sokoloff, 2000). In particular, I exploit data on the historical presence of the tsetse fly, a blood-sucking insect vector carrying a parasite that can infect humans and animals with trypanosomiasis (sleeping sickness). Such sickness is fatal if untreated, and it is particularly dangerous for animals. In a recent paper, Alsan (2015) shows how the presence of the tsetse fly, by limiting the use of domesticated animals in agriculture, affected African precolonial agricultural practices, patterns of subsistence, population density and the probability of ethnic groups being politically centralised. Alsan (2015) employs climate variables on temperature and humidity recreated for the late 19th century to generate a Tsetse Suitability Index (TSI) capturing the historical suitability of the environment to the tsetse fly.[3] By combining the TSI data with Murdock's (1959) ethnolinguistic map of Africa, I construct a suitability index for each ethnic homeland. Figure 3 provides a visual representation of the TSI across the historical ethnic homelands for which data from the *Ethnographic Atlas* is also available.

[3] Data and in-depth explanation of how the index was constructed can be found in the web Appendix of Alsan (2015), available at her website.

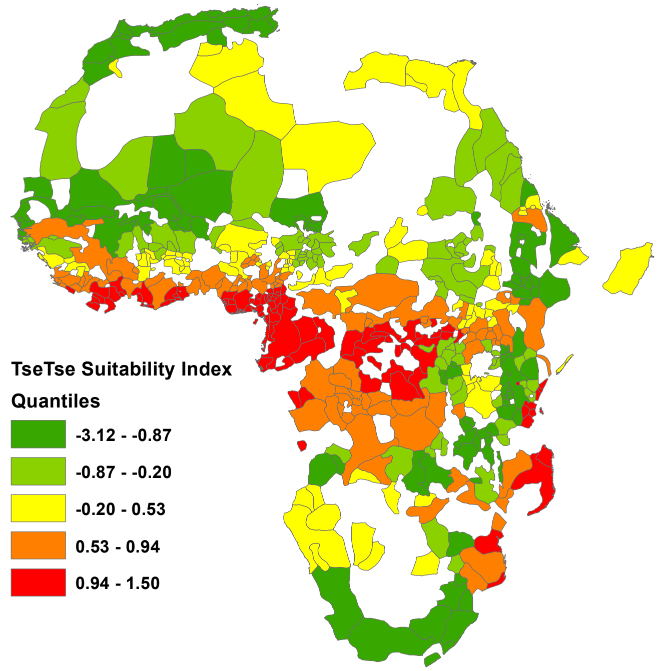


Figure 3. Tsetse Suitability Index from Alsan (2015)

3 Analytical Framework and Identification Strategy

This section presents the analytical framework employed to establish the impact of an individual's ethnic group's precolonial political centralisation on his or her current level of institutional trust, with sub-section 3.1 setting out the empirical specification and sub-section 3.2 describing the identification strategy.

To provide the intuition behind the hypothesised role played by culturally embodied beliefs transmitted intergenerationally within ethnic groups, I develop a theoretical model (see Appendix B) showing how reliance on tradition can cause beliefs to reflect experiences of past generations. The model predicts that, by relying on tradition, individuals with an ancestry of state history will be more likely to have positive attitudes towards the state. The idea builds on insights from the cultural anthropology literature, postulating that when information acquisition is either costly or imperfect, individuals employ heuristic decision-making strategies (Boyd and Richerson, 1985, 1995, 2005).

3.1 Main Empirical Model

The analysis is based on the model specification:

$$ITI_{i,e,a,d,c} = \beta_0 + \beta_1 \ PC_e + X'_{i,e,a,d,c} \ \Gamma + X'_{a,d,c} \ \Omega + X'_{d,c} \ \Phi + X'_e \ \Lambda + \delta_c + \varepsilon_{i,e,a,d,c} \qquad (1)$$

where i indexes individuals, e ethnic groups, a enumeration areas (village or town), d districts and c countries. The dependent variable $ITI_{i,e,a,d,c}$ denotes an individual's Institutional Trust Index, constructed from the four trust survey measures described in section 2. The main treatment variable, PC_e, measures the precolonial political centralisation of ethnic group e. The coefficient of interest is β_1, capturing the estimated relationship between the precolonial political centralisation of an individual's ethnic group and the current level of institutional trust of the individual.

The vector $X_{i,e,a,d,c}$ is a set of individual-level controls. The vector $X_{a,d,c}$ is a set of enumeration area-level controls. The vector $X_{d,c}$ is a set of district-level controls. The vector X_e is a set of ethnicity-level controls. Finally, δ_c denotes country fixed effects. The list of controls contained in each of these vectors, and their relevance, will be described in detail in 3.2. Summary statistics are reported in Appendix Table A.1.

Given that the main explanatory variable and many of the controls in Equation (1) vary at the ethnic group level and the potential for within-group residuals correlation, for the remainder of the analysis, my baseline approach will be to employ heteroskedasticity-robust standard errors clustered within ethnic groups.

3.2 Identification Strategy

To identify the causal effect of the individual's ethnic group's precolonial political centralisation on his current level of institutional trust, I employ an identification strategy addressing three main empirical challenges.

The first empirical challenge relates to the potential role played by contemporary and historical factors in driving the estimated relationship. To mitigate this concern, I account for a large set of confounding characteristics. Starting with the contemporary factors, I first account for a series of individual-level socio-economic characteristics from vector $X_{i,e,a,d,c}$, namely age, age squared, gender, education indicators, religion indicators and self-reported living conditions indicators. A number of previous studies established the correlation between generalised trust and income, and it is safe to assume that a similar relationship holds for institutional trust. Although the Afrobarometer surveys do not provide a direct measure of income, many of the individual-level controls are good proxies for the individual's economic situation, in particular the education and self-reported living conditions indicators. Next, I employ country fixed effects to capture nation-specific factors that may affect individuals' institutional trust, such as quality of institutions, government policies and overall level of economic development (Aghion, Algan, Cahuc, and Shleifer, 2010; Algan, Cahuc, and Shleifer, 2013; Tabellini, 2008b). I then turn to subnational factors that might correlate with differences in an individual's institutional trust through enumeration-area level controls from vector $X_{a,d,c}$ and district-level controls from vector $X_{d,c}$. First of all, I include an urban area indicator, which allows me to account for differences in economic conditions and attitudes that might exist between urban and non-urban areas. I then account for the presence of public goods (electricity, piped water, sewage system, school and health clinic) in the individual's town or village, as satisfaction with the provision of public goods can be expected to correlate with attitudes towards the state and institutional figures. Such a consideration can be thought as particularly relevant in this case, as Gennaioli and Rainer (2007) argues that a history of political centralisation affects the local provision of public goods. At the district level, I include distance from the capital to capture potentially limited state presence in the hinterland areas of African countries (Lewis, 1954; Herbst, 2000), which may lead to a different perception of the state and its representatives. I also include two control variables relating to the ethnic composition of the district. In particular, following Nunn and Wantchekon (2011), I employ ethnic fractionalisation, found to affect both income and trust (Easterly and Levine, 1997), and the proportion of the sample population belonging to the same ethnic group as the individual, accounting for the effect of being part of the majority group or an ethnic minority (Alesina and La Ferrara, 2002).

Turning to the historical factors that might confound the impact of precolonial political centralisation, the set of ethnicity-level controls from vector X_e includes historical characteristics of the individual's ethnic group and geographic features of the area the group historically inhabited. I first account for precolonial ethnic group features from the *Ethnographic Atlas*. More specifically, I include the ethnic groups' share of subsistence from agriculture and practice of intensive agriculture to capture the importance and type of agricultural practices, and thus the ability to generate an agricultural surplus. Furthermore, Michalopoulos, Putterman, and Weil (2019) shows that an ethnic group's ancestral subsistence practices are related to a variety of modern individual characteristics in African countries, providing an additional reason to control for agricultural practices. While the *Ethnographic Atlas* does not include information on the quality of precolonial states, Galor and Klemp (2017) provides evidence of indigenous slavery being an indicator of autocracies. For this reason, I employ an indigenous slavery indicator as a proxy of autocratic precolonial institutions, which may have historically eroded the view of institutional figures, possibly influencing current attitudes. I then employ settlement structure indicators, listing categories in order of increasing social and economic development, ranging from fully nomadic to complex settlements, as a proxy for historical prosperity and population density. I also control for the number of slaves taken from the ethnic group during the slave trade, given that the more advanced and densely populated groups were the most impacted and may therefore show overall lower levels of both generalised trust (Nunn and Wantchekon, 2011) and institutional trust. Next, I control for the potentially confounding effect of the geography of an ethnic group's historical homeland, namely absolute latitude, mean temperature, mean level of precipitation, ruggedness, log land area, distance to coast or navigable rivers and agricultural productivity, in order to account for historical differences in ethnic groups' development related to geographic characteristics. Of particular importance among these controls is the agricultural productivity of the homeland, potentially correlated with the likelihood of historical state formation, which I measure with Galor and Özak's (2016) Caloric Suitability Index, capturing the highest attainable potential caloric yields from cultivated crops.

While the vector of ethnicity-level controls allows me to account for a variety of potentially confounding factors, it is not necessarily exhaustive. In particular, I cannot exclude the existence of unobservable ethnicity-level factors that might correlate with both the ethnic group's precolonial political centralisation and the individual's institutional trust. Including ethnicity fixed effects would allow me to account for these factors. However, the fact that the main treatment variable is measured at the ethnicity level prevents me from doing so. Therefore, as a flexible alternative to ethnicity fixed effects, I employ cultural province fixed effects. Cultural provinces are groupings devised by Murdock (1967) to capture spatial, cultural and genealogical correlation of ethnic groups. Appendix Figure C.5 visually illustrates the division of ethnic groups in cultural provinces, as based on the *Ethnographic Atlas*, with colours used to represent the different provinces.

Finally, among the different historical factors that might drive the estimated relationship, one of the most important is the colonial rule. In particular, while I account for the country-level impact of colonialism through country fixed effects, I am unable to determine how colonisation interacted with the pre-existing political structures of the various ethnic groups, as captured by their precolonial political centralisation. While the *Ethnographic Atlas* is meant to describe living conditions and characteristics of ethnic groups prior to European contact, a problem would arise if these measures were to capture some form of colonial influence. This could be the case in the event of colonial rule already being in place by the time the ethnographers recorded the ethnic groups' features. Regarding this issue, I explore a variable in the *Ethnographic Atlas* indicating the "year of observation", that is when the ethnographic records were made. For the ethnic groups of interest, the earliest time of observation is 1830, with decade changes until the latest in 1960, and with the average being around 1916. The question is whether the ethnic groups which were recorded later are more likely to have been influenced by the colonisers, compared to the ones recorded earlier. Regarding the precolonial political centralisation measure I consider, the risk is that groups which were recorded later may result as politically developed due to colonial policies. An omitted variable bias would emerge if this particular colonial influence were to be correlated with some long-lasting effect that colonisation had on attitudes towards the state and institutional figures. To deal with this specific issue, I employ time of ethnographic record fixed effects, in order to conduct a within-period of observation comparison, isolating the unobserved factors that might be related with earlier or later ethnographic records. Appendix C.3 provides details on this measure and its relationship with the precolonial political centralisation variable.

The second empirical challenge I address relates to historical state formation also influencing current economic and institutional development (Michalopoulos and Papaioannou, 2013), which, in turn, influences institutional trust. While I am interested in the effect that historical state formation had on institutional trust through intergenerationally transmitted beliefs, I cannot exclude that there is an effect working through the current environment. To deal with this other potential channel, it is fundamental to employ country fixed effects. By comparing individuals with different ethnic backgrounds within the same country, I reduce the concern of capturing the effect of better institutions and higher levels of economic development for some of the countries in the sample. Nonetheless, the issue remains that even within the same country, ethnic homelands which historically had more developed institutions correspond to areas that are more developed today.

As a further step to address this challenge, I employ an epidemiological approach exploiting the fact that when individuals migrate their beliefs and values migrate with them, but their external environment remains behind (Fernandez, 2007; Fernandez and Fogli, 2009). Using Murdock's (1959) ethnolinguistic map of Africa to determine which area ethnic groups historically inhabited, as shown previously in Figure 2, and using data available from AidData.org (BenYishay, Rotberg, Wells, Lv, Goodman, Kovacevic, and Runfola, 2017) to georeference the location of respondents in the Afrobarometer surveys, I determine whether these individuals are still living in the historical homeland of the ethnic group they belong to (see Appendix Figure C.3). From this information, I establish that about 50% of the individuals in the sample are not living on the area historically associated with their ethnic group, a feature exploited in the analysis to examine the mechanism of intergenerationally transmitted beliefs, as the concern of the effect of historical state formation working through the current external environment will be reduced for these individuals. Comparing estimates employing the full sample of individuals with ones employing the subsample composed of individuals living outside of their historical ethnic homeland (hereafter referred to as "migrants") also allows a determination of the extent the estimates change because of the alternative channel. A possible concern of this strategy is that individuals in the migrants subsample would not be comparable to the rest of the sample, possibly introducing a selection problem which would make this strategy ineffective. In Appendix Table C.2, I compare migrants with non-migrants, that is the individuals who result as living in the area that was historically inhabited by their ancestors. The two subsamples are comparatively similar in terms of both modern individual-level and historical ethnic group-level characteristics. However, as expected, migrants are more likely to reside in urban areas and live in locations that are more ethnically fractionalised with fewer co-ethnics present. Given the overall comparability of the two groups of individuals, I exploit the migrants subsample to capture the culturally embodied, intergenerationally transmitted impact of precolonial political centralisation whose influence expands even outside of the ethnic group's historical homeland (Michalopoulos, Putterman, and Weil, 2019; Nunn and Wantchekon, 2011).

The third empirical challenge relates to precolonial political centralisation, as based on the *Ethnographic Atlas*, being an imprecise measure of the ethnic groups' state history. This is due to Murdock's (1967) variable capturing the presence of a centralised precolonial government, but not how long it existed. More specifically, a centralised government may have existed for centuries during the precolonial period or may have emerged only a couple of decades before colonisation, but Murdock's variable would still classify it as the same. Further, while by the time the ethnographic records took place, no centralised government existed, one such government may have been in place for a long time in a preceding period. This measurement error may therefore cause an attenuation bias (Greene, 2003) and underestimate the importance of state history in the results.

To deal with the attenuation bias, I exploit one of the factors that potentially shaped historical state formation in Africa, namely the presence of the tsetse fly. As argued by Alsan (2015), the presence of the tsetse fly affected African precolonial agricultural practices, patterns of subsistence, population density and the probability of ethnic groups being politically centralised. By circumscribing the use of domesticated animals as a source of draft power and the adoption of technologies complementary to draft power, the tsetse fly has been hypothesised as having hindered the generation of an agricultural surplus and the transportation of goods overland, negatively affecting the likelihood of groups adopting sedentary lifestyles and more complex organisational structures. I therefore estimate a two stage least squares version of Equation (1), employing the ethnic group's precolonial political centralisation predicted from the first stage:

$$PC_e = \gamma_0 + \gamma_1 \; TSI_e + X'_{i,e,a,d,c} \; \Gamma + X'_{a,d,c} \; \Omega + X'_{d,c} \; \Phi + X'_e \; \Lambda + \delta_c + \epsilon_{i,e,a,d,c} \tag{2}$$

where TSI_e denotes the Tsetse Suitability Index of the area historically inhabited by ethnic group e. By using one of the factors determining whether the historical homeland of an ethnic group had favourable conditions for early state formation, I aim to retrieve the true effect of the ethnic group's state history. An important point to stress is that the presence of the tsetse fly affected different characteristics of the ethnic groups apart from their precolonial political centralisation. As previously mentioned, Alsan (2015) argues that the presence of the fly had an impact on various ethnic groups' features. The set of ethnicity-level variables I include in Equation (1) are not only important controls meant to capture historical characteristics of the ethnic groups that might directly affect individuals' current institutional trust; they are also employed to properly isolate the effect of tsetse presence working through precolonial political centralisation, therefore ensuring the exclusion restriction holds.

Finally, even though I control for a variety of contemporary and historical confounding characteristics, the presence of hard-to-account-for factors is highly likely given the long time span considered by the analysis. Therefore, the presence of unobservables correlated with both historical political centralisation and modern-day institutional trust may further attenuate the estimated effect. Assuming the validity conditions are satisfied by Tsetse Suitability Index, the IV approach allows me to deal with residual omitted factors not captured by the control function. Nonetheless, I recognise that it is implausible for perfect exogeneity to hold exactly. It is possible that some unaccounted factor which is correlated with tsetse presence may have affected certain economic and institutional conditions of ethnic groups, which may in turn impact modern-day institutional trust. While keeping these limitations in mind, it is worth noting that the strategy of exploiting individuals living outside of their historical ethnic homeland allows me to mitigate any threat to identification introduced by alternative channels working through the ethnic group's environment.

4 Results

4.1 Main Results

Estimates of Equation (1) are reported in Table 1, with results from OLS in columns (1)-(2) and results from IV in columns (3)-(4). Odd-numbered columns employ the full sample of individuals, while even-numbered columns employ the subsample of migrants, defined as individuals not living on the territory of their ethnic group's historical homeland.

First focusing on the OLS estimates, starting from column (1), the coefficient of interest on the ethnic group's precolonial political centralisation indicates a statistically significant and positive effect on the individual's institutional trust.[4] To assess the magnitude of the

[4]Appendix Table A.2 reports the full set of coefficients.

coefficient, consider that the Institutional Trust Index is normalised to be in [0,1] and that precolonial political centralisation is a binary variable taking on the value of 1 if the ethnic group was politically centralised. Therefore, the reported coefficient can be interpreted as a percentage change in the ITI.

Table 1. OLS and IV Estimates of the Effect of State History on Institutional Trust

Dependent variable:	**Institutional Trust Index**			
	OLS		IV	
Sample:	Full	Migrants	Full	Migrants
	(1)	(2)	(3)	(4)
Precolonial political centralisation	0.0443***	0.0390***	0.0760***	0.0865***
	(0.0113)	(0.0118)	(0.0220)	(0.0227)
First stage:				
Tsetse suitability index			-0.2408***	-0.2388***
			(0.0381)	(0.0365)
Observations	29,546	15,595	29,546	15,595
Number of clusters	147	145	147	145
Adjusted-R^2	0.244	0.243	0.244	0.242
KP F-stat of excluded instrument			40.05	42.73
Individual-level controls	✓	✓	✓	✓
Location controls	✓	✓	✓	✓
Ethnicity-level controls	✓	✓	✓	✓
Country FE	✓	✓	✓	✓

Notes: The unit of observation is an individual. The dependent variable is an individual's institutional trust index. The main explanatory variable is the ethnic group's precolonial political centralisation, based on the *Ethnographic Atlas*. Regressions in columns (1)-(2) are estimated with OLS. Regressions in columns (3)-(4) are estimated with two stage least squares, using the tsetse suitability index of the area historically inhabited by the ethnic group as the instrument. Odd-numbered columns employ the full sample of individuals, while even-numbered columns employ the subsample of migrants, defined as individuals not living on the territory of their ethnic group's historical homeland. Individual controls include age, age squared, a gender indicator, self-reported living conditions indicators, education indicators and religion indicators. Location controls include an urban area indicator, a set of 5 dummies for the presence of public goods (electricity, piped water, sewage system, school and health clinic) in the enumeration area, the district's ethnic fractionalization, the proportion of the district's population belonging to the same ethnic group as the individual and distance to the country capital. Ethnicity-level controls include precolonial era settlement structure, intensity of agriculture, share of subsistence from agriculture and relevance of indigenous slavery, cultural province indicators, time of ethnographic record indicators, number of slaves taken from the ethnic group during the slave trade, and absolute latitude, mean temperature, mean level of precipitation, ruggedness, log land area, distance to waterways and caloric suitability index of the ethnic group's historical homeland. All specifications include country fixed effects. Heteroskedasticity-robust standard errors, clustered at the ethnic group level, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

When it comes to the coefficient reported in column (1), the estimated effect implies that belonging to an ethnic group that was politically centralised in the precolonial period is associated with a 4.43% higher present-day institutional trust.

As anticipated in the analytical framework section, these estimates may suffer from both attenuation and omitted variable biases. To mitigate these concerns, first I account for a variety of factors that may potentially confound the estimates. In particular, in all specifications, I include individual-level controls, location-level (enumeration area and district) controls, ethnicity-level historical and geographic controls, and country fixed effects. The identification strategy I employ uses instrumental variables and exploits individuals living outside of their ethnic group's historical homeland.

Starting with the IV strategy, column (3) of Table 1 reports IV estimates, where precolonial political centralisation of the ethnic group is instrumented with the Tsetse Suitability Index of the ethnic group's historical homeland. The first stage estimates report that a higher tsetse suitability is significantly and negatively correlated with precolonial political centralisation, reflecting the results from Alsan (2015). Consistent with the OLS estimates, the IV second stage estimates report a statistically significant and positive effect of precolonial political centralisation, with the estimated effect being a 7.6% higher individual's institutional trust. It must be noted that the IV coefficient on precolonial political centralisation in column (3) is significantly larger when compared with the OLS coefficient in column (1). If the IV strategy is valid, this would indicate that the OLS coefficient was indeed being biased downwards, underestimating the true effect of the ethnic group's precolonial political centralisation.

The next strategy I employ addresses the concern of an ethnic group's precolonial political centralisation also influencing its current economic and institutional environment. In columns (2) and (4) of Table 1, I replicate the estimates in columns (1) and (3) while only employing the sample of individuals not living on the area associated with their ethnic group's historical homeland, which results in a smaller sample of 15,595 observations. The reported estimates obtained from the migrants subsample show very similar results, with a statistically significant and positive effect of the ethnic group's precolonial political centralisation on the individual's institutional trust. In particular, the estimated increase of the individual's institutional trust is 3.9% for the OLS estimates in column (2) and 8.65% for the IV estimates in column (4), implying a culturally embodied, intergenerationally transmitted impact of the ethnic group's precolonial political centralisation that extends beyond the ethnic group's historical homeland. Furthermore, the alternative effect working through current institutional and economic development may not be as problematic as initially considered, given the remarkable similarity in both size and statistical significance between full sample and migrants subsample estimated effects.

As previously discussed, it is possible that the perfect exogeneity requirement of the instrument may not hold exactly, potentially producing imprecise IV point estimates. Therefore, while the OLS estimates are argued to be biased downwards by attenuation and omitted variable biases, it is also possible that the IV estimates are biased upwards as a result of the exclusion restriction being violated. For these reasons, a possible exercise is to take the OLS estimates as a lower bound and IV estimates as an upper bound, with the range produced by the two different estimation methods containing the true effect. Reassuringly, calculating the mean of OLS and IV coefficients, in columns (1) and (3) and columns (2) and (4), respectively, consistently yields a 6% increase in individual's institutional trust as a result of belonging to an ethnic group that was politically centralised in the precolonial period.

4.2 Alternative Mechanisms and Explanations

Thus far, I have established a significant and sizeable effect of the ethnic group's state history on the institutional trust of modern individuals, with results being consistent with a culturally embodied, intergenerationally transmitted explanation of the impact of the ethnic group's precolonial political centralisation. In this sub-section, I explore potential alternative mechanisms and explanations for the uncovered result.

First, I examine whether the effect of precolonial political centralisation is a function of the power held by an individual's ethnic group in the country. Second, I ascertain whether the uncovered effect is determined by horizontal transmission, both within-ethnic group and between-ethnic groups, by considering the presence of co-ethnics and the average state history of individuals living in same location. Third, I address the potential selection of individuals into more developed areas by introducing a series of current location fixed effects that allows me to conduct a within-location comparison of individuals.

4.2.1 Ethnic Group's Power

As an alternative explanation for the uncovered positive effect of the precolonial political centralisation of an ethnic group on modern individuals' institutional trust, it is possible that such a relation is due to the power that the individual's ethnic group holds in the country in the current day. In other words, if ethnic groups which were politically centralised in the precolonial period are more likely to be in a position of power in modern countries, a possibility is that individuals tend to be more trusting of institutional figures because those very positions are more likely to be occupied by co-ethnics, leading the individual to be inherently more trusting of them.

For this reason, to account for the power held by different ethnic groups in my sample of countries, I employ data from the Ethnic Power Relations (EPR) core dataset (Vogt, Bormann, Rüegger, Cederman, Hunziker, and Girardin, 2015).[5] This dataset provides country-level

[5]Interested readers can refer to the Ethnic Power Relations (EPR) website at icr.ethz.ch/data/epr/

time-varying data, for countries where ethnicity is politicised, on politically relevant ethnic groups and their access to executive state power. I link the ethnic groups in my sample of Afrobarometer survey respondents to the EPR dataset, and construct a power status variable based on the procedure described below. First, the EPR dataset provides time-varying entries of both the list of politically relevant ethnic groups and their power status classification. The number of time periods is country specific and depends on the presence of power shifts in the period under consideration (1946 to 2017 in the most recent version). In other words, if there were no changes in the power status of the ethnic groups in the country during the period under consideration, then that country will have only one time period. When multiple periods are available for a country, I employ the one that most closely overlaps with the years during which the Afrobarometer surveys were conducted.[6] The power status of ethnic groups is divided into three main categories, with each being divided in several subcategories. The first main category sees a certain group ruling alone in the country, in which case, the power status is "Monopoly" or "Dominance". The second category is for cases where power is shared, with the power status categories being "Senior Partner" or "Junior Partner". The final category is for groups excluded from power, with their possible statuses being "Powerless", "Discrimination" or "Self-Exclusion". Based on this classification, I construct a categorical power status variable taking a value of 0 if the ethnic group is classified as belonging to the "Powerless", "Discrimination" or "Self-Exclusion" categories; 1 for "Junior Partner"; 2 for "Senior Partner"; and 3 for "Monopoly" or "Dominance".[7]

For 12,366 individuals in my sample, the EPR dataset does not indicate the power held by their ethnic group in the country in which they live. According to the data codebook, ethnic groups are not considered for one of the following reasons: non-citizens, such as migrant workers; ethnic groups that never sought political representation at the national level or were never discriminated against; and ethnic groups that are too small to be considered. For these cases, I assign a value of 0 to the ethnic group in the power status variable.[8] Moreover, for two of the countries in the sample, Burkina Faso and Lesotho, ethnicity is not considered politicised. Therefore, survey respondents from these two countries are dropped for this part of the analysis.

[6] When multiple EPR periods overlap with the Afrobarometer survey years, I assign the mean power status across periods to the ethnic group.

[7] One can argue that being openly discriminated against is a worse status than being powerless or self-excluded. Nonetheless, assigning a lower value to the ethnic groups in the "Discrimination" category leaves the following results unchanged.

[8] An alternative would be to exclude individuals belonging to these ethnic groups from the sample. However, this approach might introduce a selection problem, as individuals belonging to ethnic groups with the least power in the country would be not be considered. For this reason, I follow the approach explained above.

Panel A of Table 2 reports estimates controlling, through the categorical power status variable, for the power held in the country by the ethnic group to which the individual belongs. For both OLS and IV regressions, the ethnic group's power bears a positive coefficient, although the effect is only weakly significant. On the other hand, across all specifications, the coefficient on the ethnic group's precolonial political centralisation remains remarkably similar to those in Table 1, retaining its positive effect and high statistical significance.

While I categorise the power held by the ethnic group closely following the classification provided by the EPR dataset, it is possible that what really matters is whether the ethnic group holds a significant amount of power in the country or not. Thus, I construct a new power status variable, this time as a binary indicator taking a value of 1 if the ethnic group is classified as belonging to the "Senior Partner", "Dominance" or "Monopoly" categories, and 0 otherwise. OLS and IV estimates employing this alternative measure are reported in Panel B of Table 2. Compared to the categorical measure, the binary measure of the ethnic group's power enters with a considerably larger and overall more significant positive coefficient.

Both sets of results in Panels A and B imply that individuals are indeed more likely to trust institutional figures when their ethnic group holds power in the country, although to a varying degree depending on how the measure of power is constructed. Nonetheless, the coefficient on precolonial political centralisation always remains stable when it comes to its size and significance, with point estimates that are even larger compared to the baseline ones, indicating that an ethnic group's state history and current power status, although both relevant and possibly related, each has their own separate impact on the individual's institutional trust.

4.2.2 Horizontal versus Vertical Transmission

The impact of ethnic groups' precolonial political centralisation on individuals' institutional trust may be vertically transmitted within ethnic groups, or it may be a result of horizontal transmission (Bisin and Verdier, 2001). In other words, it is possible that the established effect is due, in part or wholly, to cultural values horizontally transmitted from the individual's peers. I now consider two possible forms of horizontal transmission: (i) between-ethnic groups horizontal transmission, that is the impact of the average precolonial political centralisation of other ethnic groups; (ii) within-ethnic group horizontal transmission, that is the impact of the ethnic group's precolonial political centralisation through co-ethnics.

Regarding the between-ethnic groups horizontal transmission, it is possible that an individual's institutional trust is influenced by the average state history of other individuals or, more

Table 2. OLS and IV Estimates of the Effect of State History on Institutional Trust: Ethnic Group's Power

Dependent variable:	**Institutional Trust Index**			
	OLS		IV	
Sample:	Full	Migrants	Full	Migrants
	(1)	(2)	(3)	(4)
	Panel A: Categorical power status			
Precolonial political centralisation	0.0449***	0.0382***	0.0831***	0.0880***
	(0.0121)	(0.0118)	(0.0231)	(0.0261)
Ethnic group's power	0.0125*	0.0105*	0.0127*	0.0110*
	(0.0068)	(0.0062)	(0.0067)	(0.0061)
Adjusted-R^2	0.259	0.251	0.258	0.250
KP F-stat of excluded instrument			29.93	33.94
	Panel B: Binary power status			
Precolonial political centralisation	0.0457***	0.0396***	0.0912***	0.0934***
	(0.0119)	(0.0115)	(0.0231)	(0.0256)
Ethnic group's power	0.0324**	0.0239*	0.0333**	0.0267**
	(0.0140)	(0.0130)	(0.0138)	(0.0128)
Adjusted-R^2	0.259	0.251	0.258	0.250
KP F-stat of excluded instrument			31.23	34.74
Observations	26,795	14,733	26,795	14,733
Number of clusters	142	140	142	140
Individual-level controls	✓	✓	✓	✓
Location controls	✓	✓	✓	✓
Ethnicity-level controls	✓	✓	✓	✓
Country FE	✓	✓	✓	✓

Notes: The unit of observation is an individual. The dependent variable is an individual's institutional trust index. The main explanatory variables are the ethnic group's precolonial political centralisation, based on the *Ethnographic Atlas*, and the power held by the ethnic group in the respondent's country, based on the Ethnic Power Relations (EPR) core dataset. Regressions in columns (1)-(2) are estimated with OLS. Regressions in columns (3)-(4) are estimated with two stage least squares, using the tsetse suitability index of the area historically inhabited by the ethnic group as the instrument. Odd-numbered columns employ the full sample of individuals, while even-numbered columns employ the subsample of migrants, defined as individuals not living on the territory of their ethnic group's historical homeland. Individual controls include age, age squared, a gender indicator, self-reported living conditions indicators, education indicators and religion indicators. Location controls include an urban area indicator, a set of 5 dummies for the presence of public goods (electricity, piped water, sewage system, school and health clinic) in the enumeration area, the district's ethnic fractionalization, the proportion of the district's population belonging to the same ethnic group as the individual and distance to the country capital. Ethnicity-level controls include precolonial era settlement structure, intensity of agriculture, share of subsistence from agriculture and relevance of indigenous slavery, cultural province indicators, time of ethnographic record indicators, number of slaves taken from the ethnic group during the slave trade, and absolute latitude, mean temperature, mean level of precipitation, ruggedness, log land area, distance to waterways and caloric suitability index of the ethnic group's historical homeland. All specifications include country fixed effects. Heteroskedasticity-robust standard errors, clustered at the ethnic group level, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

generally, conforms to a common and possibly deep-rooted belief that is prevalent in a specific location. To examine this potential alternative mechanism, I consider, in addition to the precolonial political centralisation of the individual's ethnic group, the precolonial political centralisation of the other ethnic groups present in the same location as the individual. To this end, I calculate, for each individual, the population-weighted average precolonial political centralisation of survey respondents living in the same area but belonging to different ethnic groups.

Table 3 reports OLS estimates controlling for the average precolonial political centralisation of others, constructed at the district level in columns (1)-(2) and at the region level in columns (3)-(4). If there are no survey respondents from other ethnic groups in the area, then this variable is assigned a missing value. This results in a smaller sample, especially when the measure is constructed at the district level, whereas the sample reduction is limited when the measure is constructed at the region level because the geographical area is larger, and it is therefore more likely that other ethnicities are present. The estimated effect of the average precolonial political centralisation of other ethnic groups in the same location is statistically insignificant across all specifications, while the effect of the individual's ethnic group's centralisation is robust to controlling for the location-based measure, with point estimates that are comparable in size and significance to those in Table 1. Moreover, comparing the results obtained from employing the same reduced sample and specifications of Table 3 but without the location-based measure (i.e. 0.0359, 0.0296, 0.0433 and 0.0387, respectively), reveals that point estimates are only barely affected by the inclusion of this measure. Although it is possible that these results are due to a lack of precision in the location-based measure, being based on the sample population from the Afrobarometer surveys, they are also consistent with the insignificant role played by between-groups horizontal transmission.

Regarding the within-ethnic group horizontal transmission, I established through analysis that descending from an ethnic group with a history of political centralisation positively affects modern individuals' institutional trust. While the migrants subsample strategy provides evidence in support of an intergenerational transmission explanation, an important role may also be played by horizontal transmission from co-ethnics. In other words, the presence of co-ethnics may be a potential source of heterogeneity that may affect either the transmission or resilience of ethnicity-related beliefs and other ethnic traits across generations. I therefore examine whether being surrounded by people belonging to the same ethnic group or, more generally, living in areas predominantly inhabited by co-ethnics, is determinant in transmitting beliefs within the ethnic group.

For this purpose, I exploit a variable employed as a control up to this point, the proportion of co-ethnics in the district. This measure, based on the sample population of respondents from the Afrobarometer surveys, indicates the share of the district's population that is of the same ethnicity as the individual. Studying the distribution of this variable reveals that about 50% of the individuals in the sample live in a district with more than 75% of its population being composed of co-ethnics. This finding is consistent with what I previously determined regarding the number of individuals still living in the historical homeland of their ethnic group. To study the heterogeneous effect of within-ethnic group horizontal transmission, I interact the proportion of co-ethnics variable with the ethnic group's precolonial political centralisation measure in order to determine the existence of a differential effect based on the varying presence of co-ethnics.

Table 4 reports OLS and IV estimates employing the interaction with the proportion of co-ethnics. For both types of estimates I show the following: the coefficient on the ethnic group's precolonial political centralisation, the coefficient on its interaction with the (centred) proportion of co-ethnics in the district, and the sum of the previous two coefficients. OLS estimates in columns (1) and (2) show a positive and statistically significant coefficient on the interaction term, indicating that the impact of the ethnic group's precolonial political centralisation on the individual's institutional trust is increasing with co-ethnics presence. Moreover, after disentangling the heterogeneous effect of co-ethnics presence, the coefficient on the non-interacted term reports an even stronger effect of precolonial political centralisation when compared to the baseline estimates in Table 1. Finally, the sum of the coefficients on the ethnic group's precolonial political centralisation measure and its interaction with the proportion of co-ethnics shows an overall positive and highly significant estimated effect. The aforementioned results are very similar for the estimates employing the migrants subsample in column (2), implying that co-ethnics presence similarly matters for individuals not living on their ethnic group's historical homeland.

Turning to the IV estimates in columns (3) and (4), interacting the instrumented precolonial political centralisation measure requires an additional instrument. I use a commonly chosen approach, employing the interaction between the Tsetse Suitability Index and the proportion of co-ethnics as the second instrument. Unfortunately, because the two employed instruments do not emerge as being jointly excluded, estimates are prevented from attaining proper identification. Nonetheless, results from IV regressions present similar patterns to the OLS results, with the coefficient on the ethnic group's precolonial political centralisation showing an even stronger impact on the individual's institutional trust after including the interaction with the proportion of co-ethnics. Although the instrumented interaction term itself is not statistically significant, the sum of the coefficients on the interacted and non-interacted terms shows a highly significant and positive effect, in both columns (3) and (4), similar to the OLS specifications.

Table 3. OLS and IV Estimates of the Effect of State History on Institutional Trust: Between-Ethnic Groups Horizontal Transmission

Dependent variable:	**Institutional Trust Index**			
Location:	District		Region	
Sample:	Full	Migrants	Full	Migrants
	(1)	(2)	(3)	(4)
Precolonial political centralisation	0.0339***	0.0281**	0.0437***	0.0359***
	(0.0123)	(0.0120)	(0.0109)	(0.0114)
Average precolonial political central.	-0.0127	-0.0106	0.0048	-0.0172
among other ethnicities in same location	(0.0081)	(0.0077)	(0.0146)	(0.0157)
Observations	20,063	11,993	27,990	14,954
Number of ethnicity clusters	145	144	147	145
Number of location clusters	1,065	1,031	223	223
Adjusted-R^2	0.233	0.235	0.241	0.243
Individual-level controls	✓	✓	✓	✓
Location controls	✓	✓	✓	✓
Ethnicity-level controls	✓	✓	✓	✓
Country FE	✓	✓	✓	✓

Notes: The unit of observation is an individual. All regressions are estimated with OLS. The dependent variable is an individual's institutional trust index. The main explanatory variable is the ethnic group's precolonial political centralisation, based on the *Ethnographic Atlas*. The location-based measure is the population-weighted average precolonial political centralisation of survey respondents belonging to an ethnic group different from the individual and living in the same district or region. Odd-numbered columns employ the full sample of individuals, while even-numbered columns employ the subsample of migrants, defined as individuals not living on the territory of their ethnic group's historical homeland. Individual controls include age, age squared, a gender indicator, self-reported living conditions indicators, education indicators and religion indicators. Location controls include an urban area indicator, a set of 5 dummies for the presence of public goods (electricity, piped water, sewage system, school and health clinic) in the enumeration area, the district's ethnic fractionalization, the proportion of the district's population belonging to the same ethnic group as the individual and distance to the country capital. Ethnicity-level controls include precolonial era settlement structure, intensity of agriculture, share of subsistence from agriculture and relevance of indigenous slavery, cultural province indicators, time of ethnographic record indicators, number of slaves taken from the ethnic group during the slave trade, and absolute latitude, mean temperature, mean level of precipitation, ruggedness, log land area, distance to waterways and caloric suitability index of the ethnic group's historical homeland. All specifications include country fixed effects. Heteroskedasticity-robust standard errors, clustered within ethnic groups and districts in columns (1)-(2) and within ethnic groups and regions in columns (3)-(4), are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

4.2.3 Selection

As part of the identification strategy, I exploit individuals in the migrants subsample as a way to mitigate the potential for the ethnic group's precolonial political centralisation to affect the individual's institutional trust through the current economic and institutional environment. A potential issue with this strategy is that there may be a selection of where these individuals decided to live. For example, either the individual or a previous generation could have migrated to locations that are more developed or that have better institutions.

Alternatively, they could have migrated to a place that is similar to their homeland of origin or inhabited by co-ethnics. As shown in Appendix Table C.2, migrants are more likely to reside in urban areas, which is consistent with a general migration pattern of individuals moving to urban centres in search of better opportunities. The reported comparative statistics further show that, on average, fewer co-ethnics are present where migrants live. Although this work focuses on the precolonial political centralisation of the individual's ethnic group, with its impact hypothesised to be intergenerationally transmitted, there is no doubt that the current economic and institutional environment also exerts an influence on individuals' norms, beliefs and attitudes.

To deal with these potential concerns, based on Murdock's (1959) ethnolinguistic map and the georeferenced location of respondents in the Afrobarometer surveys, I introduce a series of current ethnic homeland fixed effects corresponding to the ethnic homeland where individuals are currently located. Adding these fixed effects will isolate time-invariant characteristics related to the geographic, economic and institutional environment of where individuals are currently located, allowing for a more controlled within-location comparison of individuals from different ethnic groups. Furthermore, corresponding to the location where ethnic groups historically resided, these fixed effects enable me to account for cultural and historical features of the location.

Table 5 reports OLS and IV estimates employing current homeland fixed effects. Although the estimates employing the migrants subsample see a general reduction in the size and statistical significance of the coefficient on the ethnic group's precolonial political centralisation, the estimated impact is overall consistent with the results of Table 1, even when conducting a more controlled within-location comparison.

4.3 Robustness

In this subsection, I describe a number of robustness checks to determine whether the results I document are valid and imply a causal relationship. First, I undertake a falsification exercise where I instrument the ethnic group's precolonial political centralisation with the ethnic homeland's Malaria Ecology Index instead of the Tsetse Suitability Index. Second, I employ the original classification of institutional development contained in the *Ethnographic Atlas* as an alternative measure of precolonial political centralisation. Third, I use data on the location of African cities in 1800 as an alternative proxy for precolonial state existence. Fourth, I use an alternative version of the Institutional Trust Index and run separate regressions employing the singular institutional trust questions. Fifth, to show there are no significant inference issues deriving from how I choose to treat standard errors, I employ different clustering techniques. Sixth, I use alternative sets of fixed effects.

Table 4. OLS and IV Estimates of the Effect of State History on Institutional Trust: Proportion of Co-Ethnics

Dependent variable:	**Institutional Trust Index**			
	OLS		IV	
Sample:	Full	Migrants	Full	Migrants
	(1)	(2)	(3)	(4)
Precolonial political centralisation	0.0469***	0.0480***	0.0799***	0.1026***
	(0.0110)	(0.0115)	(0.0213)	(0.0244)
Precolonial political centralisation ×	0.0112***	0.0121***	0.0256	0.0290
Proportion of co-ethnics	(0.0041)	(0.0039)	(0.0283)	(0.0199)
Sum of coefficients	0.0581***	0.0601***	0.1055***	0.1316***
	(0.0125)	(0.0130)	(0.0381)	(0.0375)
Observations	28,558	15,327	28,558	15,327
Number of ethnicity clusters	147	145	147	145
Number of district clusters	1,935	1,505	1,935	1,505
Adjusted-R^2	0.253	0.248	0.252	0.246
F-stats of excluded instruments:				
Tsetse suitability index			19.65	22.75
TSI × Proportion of co-ethnics			2.75	5.14
Joint			2.75	3.77
Individual-level controls	✓	✓	✓	✓
Location controls	✓	✓	✓	✓
Ethnicity-level controls	✓	✓	✓	✓
Country FE	✓	✓	✓	✓

Notes: The unit of observation is an individual. The dependent variable is an individual's institutional trust index. The main explanatory variables are the ethnic group's precolonial political centralisation, based on the *Ethnographic Atlas*, and the interaction between the precolonial political centralisation variable and the proportion of the district's population belonging to the same ethnic group as the individual, based on the Afrobarometer surveys' sample population. Regressions in columns (1)-(2) are estimated with OLS. Regressions in columns (3)-(4) are estimated with two stage least squares, using the tsetse suitability index of the area historically inhabited by the ethnic group, and its interaction with the proportion of co-ethnics, as instruments. Odd-numbered columns employ the full sample of individuals, while even-numbered columns employ the subsample of migrants, defined as individuals not living on the territory of their ethnic group's historical homeland. Individual controls include age, age squared, a gender indicator, self-reported living conditions indicators, education indicators and religion indicators. Location controls include an urban area indicator, a set of 5 dummies for the presence of public goods (electricity, piped water, sewage system, school and health clinic) in the enumeration area, the district's ethnic fractionalization and distance to the country capital. Ethnicity-level controls include precolonial era settlement structure, intensity of agriculture, share of subsistence from agriculture and relevance of indigenous slavery, cultural province indicators, time of ethnographic record indicators, number of slaves taken from the ethnic group during the slave trade, and absolute latitude, mean temperature, mean level of precipitation, ruggedness, log land area, distance to waterways and caloric suitability index of the ethnic group's historical homeland. All specifications include country fixed effects. Heteroskedasticity-robust standard errors, clustered within ethnic groups and districts, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

4.3.1 Malaria Ecology Falsification

My instrumental variable strategy rests on the assumption that historical presence of the tsetse fly affects modern institutional trust only through ethnic groups' precolonial political

Table 5. OLS and IV Estimates of the Effect of State History on Institutional Trust: Current Homeland Fixed Effects

Dependent variable:	**Institutional Trust Index**			
	OLS		IV	
Sample:	Full	Migrants	Full	Migrants
	(1)	(2)	(3)	(4)
Precolonial political centralisation	0.0415***	0.0208*	0.0723***	0.0403**
	(0.0090)	(0.0109)	(0.0173)	(0.0197)
Observations	29,196	15,245	29,196	15,245
Number of clusters	147	145	147	145
Adjusted-R^2	0.263	0.255	0.263	0.254
KP F-stat of excluded instrument			46.15	40.26
Individual-level controls	✓	✓	✓	✓
Location controls	✓	✓	✓	✓
Ethnicity-level controls	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Current ethnic homeland FE	✓	✓	✓	✓

Notes: The unit of observation is an individual. The dependent variable is an individual's institutional trust index. The main explanatory variable is the ethnic group's precolonial political centralisation, based on the *Ethnographic Atlas*. Regressions in columns (1)-(2) are estimated with OLS. Regressions in columns (3)-(4) are estimated with two stage least squares, using the tsetse suitability index of the area historically inhabited by the ethnic group as the instrument. Odd-numbered columns employ the full sample of individuals, while even-numbered columns employ the subsample of migrants, defined as individuals not living on the territory of their ethnic group's historical homeland. Individual controls include age, age squared, a gender indicator, self-reported living conditions indicators, education indicators and religion indicators. Location controls include an urban area indicator, a set of 5 dummies for the presence of public goods (electricity, piped water, sewage system, school and health clinic) in the enumeration area, the district's ethnic fractionalization, the proportion of the district's population belonging to the same ethnic group as the individual and distance to the country capital. Ethnicity-level controls include precolonial era settlement structure, intensity of agriculture, share of subsistence from agriculture and relevance of indigenous slavery, cultural province indicators, time of ethnographic record indicators, number of slaves taken from the ethnic group during the slave trade, and absolute latitude, mean temperature, mean level of precipitation, ruggedness, log land area, distance to waterways and caloric suitability index of the ethnic group's historical homeland. All specifications include country and current ethnic homeland fixed effects. Heteroskedasticity-robust standard errors, clustered at the ethnic group level, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

centralisation. To make sure the exclusion restriction holds, I employ a rich set of modern and historical controls, dedicating particular attention to the colonial rule, which, as highlighted in the analytical framework, is possibly one of the most important potential omitted factors. In the event that the colonisers' strategy was also affected by tsetse presence, there would be a concrete risk of the exclusion restriction not being satisfied. Specifically, the tsetse suitability instrument may affect the Institutional Trust Index not only through the ethnic group's precolonial political centralisation, but also through the colonisation strategy implemented by Europeans.

Given that malaria was found to be one of the major factors affecting the mortality, and consequently strategy, of European settlers (Acemoglu, Johnson, and Robinson, 2001, 2002), I undertake a falsification exercise where I employ as instrument a measure of the prevalence of malaria in the homelands historically inhabited by ethnic groups.

In this exercise, I compare results from IV regressions instrumenting with tsetse suitability, whose main impact is supposed to be via the local population's livelihood, with results obtained by instrumenting with an aspect of the disease environment that, via human mortality, has been found to have affected the attractiveness of a locality to colonisers. The approximation of malaria disease environment in the historical ethnic homelands is obtained by combining the Malaria Ecology Index by Kiszewski, Mellinger, Spielman, Malaney, Sachs, and Sachs (2004) with Murdock's (1959) ethnolinguistic map.

Appendix Table A.3 reports IV estimates comparing first-stage and second-stage results obtained from employing the Tsetse Suitability Index and Malaria Ecology Index, respectively. First-stage estimates report a negative and weakly significant effect of the Malaria Ecology Index on precolonial political centralisation in column (3), while the the effect is not statistically significant in column (4). Furthermore, first-stage coefficients obtained with the malaria instrument are more than ten times smaller than those obtained with the tsetse instrument, with the addition of the F-statistic of excluded instrument being considerably low. Finally, the second-stage coefficient on precolonial political centralisation is not statistically significant when the measure is instrumented with the Malaria Ecology Index.

While this falsification exercise does not provide direct support for the validity of the Tsetse Suitability Index instrument, the stark difference in results obtained while instrumenting with the Malaria Ecology Index shows how these two aspects of the disease environment are fundamentally different in the context of this analysis, offering suggestive evidence that is consistent with the functionality of the tsetse instrument and against it capturing omitted colonial rule factors.

4.3.2 Different Classification of Political Centralisation

Up to this point, the analysis employed a binary measure of precolonial political centralisation, based on Murdock's (1967) Jurisdictional Hierarchy Beyond the Local Community Level index. This variable, ranging from 0 to 4, indicates the number of jurisdictional levels above the local level, where 0 stands for stateless societies, 1 and 2 stand for petty and large chiefdoms, and 3 and 4 stand for states and large states. Following Gennaioli and Rainer (2006, 2007), I aggregated the index into a binary measure to account for measurement error potentially deriving from subjectivity in the classification of political jurisdictions by ethnographers. However, the baseline results may be a function of this aggregation. Furthermore, as argued by Marshall (2016), aggregating a treatment variable can in some cases result in IV estimates being biased upwards.

Appendix Table A.4 replicates the OLS and IV estimates of Table 1 while using the categorical Jurisdictional Hierarchy Beyond the Local Community Level index. Overall, results using the original index are similar in terms of both size and statistical significance to those obtained with the binary measure of precolonial political centralisation. Worth noticing, however, is that these estimates perform more poorly in terms of excluded instrument statistics.

4.3.3 Alternative Proxy for Precolonial States

To assess the robustness of results to alternative measures of precolonial state existence, I employ a proxy based on data from Chandler (1987) on the location of African cities in 1800. Under the assumption that states tended to have a large city as the political centre, I construct a binary indicator equalling 1 if there was a city in 1800 with more than 20,000 inhabitants on the area associated with an ethnic group's historical homeland.

Appendix Table A.5 reports OLS and IV estimates employing the city measure. OLS estimates in columns (1)-(2) are similar to those obtained with the precolonial political centralisation measure, in terms of both size and significance. Regarding the results from IV in columns (3)-(4), the first-stage coefficient on tsetse suitability is negative and highly significant, although the size is two to three times smaller compared to Table 1, coupled with a lower excluded instrument statistic. Second-stage estimates report a highly significant and positive coefficient on the presence of a city on the ethnic homeland territory, with an effect on the Institutional Trust Index, which is even stronger than the one obtained with the ethnic group's precolonial political centralisation. About this last point, the number of ethnic groups with a city on their homeland in 1800 is limited compared to the number that were politically centralised in the precolonial period, meaning it is possible that the estimates are biased upwards.

4.3.4 Institutional Trust Index Components

The Institutional Trust Index employed throughout the analysis is based on four survey questions, concerning trust with respect to president, parliament, police and courts of law. The Afrobarometer surveys contain additional questions on trust with respect to institutional figures and bodies: local government council, electoral commission, ruling party and opposition party. While these figures are not necessarily comparable to the ones I consider, as they are either related to a specific political party or not immediately referable to the state, a common underlying determinant affected by the ethnic group's precolonial political centralisation may still influence individuals' trust with respect to them. I construct an alternative version of the

Institutional Trust Index which is based both on the baseline questions and these additional trust questions.

Appendix Table A.6 reports results from regressions employing the alternative index, with OLS estimates in columns (1)-(2) and IV estimates in columns (3)-(4). Overall, the results are comparable to those in Table 1, albeit the size of the point estimates is slightly reduced. Additionally, Appendix Figure A.1 presents coefficient plots from regressions employing as dependent variable each institutional trust question separately, alongside the baseline Institutional Trust Index coefficient. Although similar patterns are present across the different trust questions, there is some degree of variation in the size and statistical significance of the point estimates.

4.3.5 Clustering

Given that the explanatory variables and controls I employ vary at the ethnicity level, my baseline approach has been to cluster standard errors at the ethnic group level. Nonetheless, some additional correlation of observed and unobserved features may exist either at the country level or at more disaggregated geographical level.

In Appendix Table A.7, I replicate the results of Table 1, while employing standard errors adjusted for two-way clustering within ethnic groups and districts, within ethnic groups and regions, and within ethnic groups and countries. These methods all produce standard errors that are similar to those obtained with clustering at the ethnic group level.

Additionally, I employ Conley's (1999) standard errors adjusting for two-way spatial correlation. In Appendix Table A.8, I replicate the OLS estimates of Table 1, applying spatially adjusted standard errors considering windows of 250, 500 and 1,000 km. In particular, in columns (1)-(2), the spatial adjustment uses a uniform kernel, assigning equal weight to all observations inside the considered window, and zero weight to observations further apart. In columns (3)-(4), the spatial adjustment uses a conical kernel, with weights decaying linearly with distance. These methods also yield similar standard errors.

4.3.6 Alternative Fixed Effects Specifications

The two rounds of Afrobarometer surveys I employ were conducted in 2005 and 2008, respectively. Up to this point, I assumed that the time difference was small enough to consider individuals surveyed in the same country in the two different time periods comparable. However, some country- and time-specific events and conditions may make it so that individuals interviewed in the two rounds are not readily comparable. For this reason, in Appendix Table A.9, I run regressions employing country-year fixed effects, instead of simple country fixed effects. OLS estimates in columns (1)-(2) and IV estimates in columns (3)-(4) are similar to the ones in Table 1, although coefficients are relatively smaller in size and come with a slightly reduced statistical significance.

Another alternative to a within-country comparison is to consider fixed effects specifications based on smaller administrative units, thus ensuring that the geographical area is small enough that only individuals subjected to closely similar economic and institutional conditions are compared. Going beyond countries, the other exploitable administrative units from the Afrobarometer surveys are, in decreasing order of size: regions, districts and enumeration areas (towns or villages). The surveyed individuals in my sample are divided into 239 regions, 2,071 districts and 5,255 enumeration areas. Therefore, employing fixed effects specifications at either the district or enumeration area level would introduce a very large number of fixed effects, in fact too large to provide any meaningful comparison of individuals belonging to different ethnic groups. For this reason, the only accessible alternative fixed effects specification is at the region level. Appendix Table A.10 reports OLS and IV estimates employing region fixed effects. Even through this demanding specification, the estimated effect of an ethnic group's precolonial political centralisation on the individual's institutional trust remains positive and significant, although both coefficient size and statistical significance are reduced compared to Table 1.

5 Conclusions

At the start of European colonisation, Africa's numerous ethnic groups were characterised by a remarkable heterogeneity in terms of state institutions. Generations later, descendants of those ethnic groups live together in the setting of colonially established modern countries. While current political, economic and institutional factors shape trust in the state and recognition of its legitimacy, various studies have documented how historical factors can shape cultural traits and how these cultural differences can persist for long periods of time. With mounting evidence of the long-run impact of precolonial states on economic development and the far-reaching consequences of the creation of artificial states during the Scramble for Africa, what is the legacy of the significant heterogeneity of ethnic groups' precolonial state institutions on the attitudes of individuals with respect to modern states?

In this paper, I advance the hypothesis that African ethnic groups' precolonial state institutions shaped individuals' attitudes towards the modern state through culturally embodied beliefs transmitted intergenerationally within ethnic groups. I test the hypothesis empirically by combining contemporary individual-level data on trust in state institutions from the Afrobarometer surveys with historical data on ethnic groups' precolonial political centralisation from the *Ethnographic Atlas*. To identify a causal effect, I employ a combination of identification strategies. First, to account for the potentially confounding effect of contemporary and historical factors, I employ an array of controls based on individual-level characteristics, subnational features, country fixed effects and ethnicity-level historical characteristics and geographic features. Second, to establish the mechanism of intergenerational transmission and determine the role of culturally embodied ethnicity-related traits, I exploit survey individuals living outside of their ethnic group's historical homeland. Third, I use data on the historical disease environment, namely tsetse fly presence, as an instrument for the ethnic groups' precolonial political centralisation in order to deal with issues related to measurement error and unobserved factors. Through the analysis, I establish that modern-day individuals whose ancestors were exposed to precolonial state institutions show overall higher confidence in current state institutions.

To confirm whether culturally embodied norms and beliefs transmitted intergenerationally within ethnic groups played a fundamental role in determining the impact of precolonial state institutions on individuals' institutional trust, I consider a number of alternative mechanisms and explanations. In particular, I examine the role played by the power status of the individual's ethnic group in the country; study the role played by horizontal transmission, both within and between ethnic groups; consider the impact of, and on, other trust dimensions; and address the selection of individuals into more developed areas.

I then perform a number of robustness checks to assess the validity of my identification strategy and results. Specifically, I consider potential inference issues deriving from the treatment of standard errors; perform a falsification exercise by instrumenting with a different measure of the disease environment; employ a different classification of precolonial political centralisation; use an alternative proxy of precolonial states based on data on the location of African cities in 1800; consider alternative indexes and variables to capture institutional trust; and employ alternative fixed effects specifications.

This paper adds to a vibrant new literature in economics studying how persistent cultural traits can act as a mechanism through which historical factors shape current outcomes. Results call for future research. Particularly, it highlights a promising research agenda on the intergenerational transmission of values and beliefs. Moreover, a next natural step would be to study how these ethnic legacies interact with national policies.

Appendix A - Tables and Figures

Table A.1. Summary Statistics

Variable	*Mean*	*St. Dev.*	*Min*	*Max*	*N*
Modern data					
Institutional Trust Index	0.59	0.29	0	1	33,419
Age	36	14.52	18	99	36,832
Male	0.50	0.50	0	1	37,287
Urban area	0.37	0.48	0	1	37,287
Education	3.12	2.03	0	9	37,186
Living conditions	2.57	1.21	1	5	37,130
School in enumeration area	0.83	0.38	0	1	37,075
Electricity in enumeration area	0.53	0.50	0	1	36,990
Piped water in enumeration area	0.48	0.50	0	1	36,876
Sewage in enumeration area	0.22	0.41	0	1	36,450
Health clinic in enumeration area	0.55	0.50	0	1	36,555
District-level ethnic fractionalization	0.38	0.30	0	0.91	37,287
Proportion of ethnic group in district	0.62	0.35	0	1	37,287
Distance to the capital (in 1,000 kms)	0.30	0.28	0	1.89	37,287
Ethnic group-level historical data					
Precolonial political centralisation	0.40	0.49	0	1	162
Jurisdictional hierarchies beyond local community	1.35	0.92	0	3	162
Share of subsistence from agriculture	5.78	1.48	0	9	170
Intensive agriculture	0.28	0.45	0	1	162
Indigenous slavery	0.84	0.37	0	1	158
Settlement structure	5.97	1.47	1	8	162
Year of ethnographic record	1916	23	1830	1960	170
Slave exports (ln(1+exports/area))	0.31	0.70	0	3.66	170
Ethnic group-level geographic data					
Tsetse suitability index	-0.01	0.99	-3.12	1.45	170
Calorie suitability index	0.76	0.19	0	1	170
Absolute latitude	9.48	6.65	0	33	170
Log land area	23.72	1.13	21.09	26.48	170
Distance to sea coast (in 1,000 kms)	0.50	0.37	0.00	1.39	170
Distance to major rivers (in 1,000 kms)	0.07	0.06	0.00	0.29	170
Average temperature (°C)	25.22	2.99	16.22	29.88	170
Average precipitation (m)	1.02	0.58	0.06	3.16	170
Ruggedness	0.13	0.15	0	1	170

Table A.2. OLS and IV Estimates of the Effect of State History on Institutional Trust: Baseline Results with all Coefficients

Dependent variable:	**Institutional Trust Index**			
	OLS		IV	
Sample:	Full	Migrants	Full	Migrants
	(1)	(2)	(3)	(4)
Precolonial political centralisation	0.0443***	0.0390***	0.0760***	0.0865***
	(0.0113)	(0.0118)	(0.0220)	(0.0227)
Age	0.0006	0.0006	0.0006	0.0006
	(0.0006)	(0.0008)	(0.0006)	(0.0008)
Age^2	0.0000	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Gender = male	-0.0036	-0.0002	-0.0036	0.0001
	(0.0029)	(0.0039)	(0.0029)	(0.0039)
Education (baseline: No formal schooling)				
Informal schooling only	-0.0116	-0.0045	-0.0121	-0.0057
	(0.0108)	(0.0164)	(0.0107)	(0.0164)
Some primary schooling	0.0017	0.0143	0.0017	0.0142
	(0.0066)	(0.0088)	(0.0066)	(0.0087)
Primary school completed	-0.0221***	-0.0148	-0.0220***	-0.0151*
	(0.0077)	(0.0090)	(0.0077)	(0.0088)
Some secondary school/high school	-0.0436***	-0.0362***	-0.0434***	-0.0358***
	(0.0088)	(0.0108)	(0.0088)	(0.0107)
Secondary school/high school completed	-0.0594***	-0.0443***	-0.0593***	-0.0442***
	(0.0116)	(0.0127)	(0.0115)	(0.0126)
Post-secondary qualifications, not university	-0.0681***	-0.0592***	-0.0677***	-0.0588***
	(0.0123)	(0.0148)	(0.0122)	(0.0147)
Some university	-0.0982***	-0.0834***	-0.0982***	-0.0837***
	(0.0170)	(0.0169)	(0.0169)	(0.0169)
University completed	-0.0679***	-0.0645***	-0.0676***	-0.0641***
	(0.0222)	(0.0218)	(0.0221)	(0.0216)
Post-graduate	-0.0523**	-0.0355	-0.0511**	-0.0334
	(0.0252)	(0.0290)	(0.0250)	(0.0288)
Living conditions (baseline: Much worse)				
Worse	0.0412***	0.0410***	0.0412***	0.0414***
	(0.0074)	(0.0098)	(0.0074)	(0.0097)
Same	0.0528***	0.0499***	0.0528***	0.0504***
	(0.0072)	(0.0084)	(0.0072)	(0.0083)
Better	0.0967***	0.0945***	0.0965***	0.0946***
	(0.0072)	(0.0093)	(0.0072)	(0.0092)
Much better	0.0995***	0.1045***	0.0993***	0.1047***
	(0.0130)	(0.0149)	(0.0130)	(0.0148)
Urban area	-0.0197***	-0.0210***	-0.0195***	-0.0208***
	(0.0053)	(0.0063)	(0.0053)	(0.0062)
School in enumeration area	0.0009	0.0054	0.0010	0.0053
	(0.0077)	(0.0088)	(0.0076)	(0.0087)
Electricity in enumeration area	-0.0220***	-0.0261***	-0.0223***	-0.0263***
	(0.0053)	(0.0071)	(0.0053)	(0.0070)
Pipep water in enumeration area	-0.0030	0.0016	-0.0035	0.0007
	(0.0039)	(0.0051)	(0.0039)	(0.0050)
Sewage in enumeration area	-0.0112*	-0.0115*	-0.0110*	-0.0110*
	(0.0064)	(0.0062)	(0.0063)	(0.0062)
Health clinic in enumeration area	0.0058	0.0056	0.0057	0.0056
	(0.0041)	(0.0048)	(0.0041)	(0.0048)
District-level ethnic fractionalization	0.0251*	0.0393***	0.0249*	0.0385***
	(0.0150)	(0.0144)	(0.0150)	(0.0143)
Proportion of ethnic group in district	0.0250***	0.0254**	0.0247***	0.0241**
	(0.0092)	(0.0113)	(0.0092)	(0.0114)
Distance to the capital	0.0052***	0.0074***	0.0052***	0.0077***
	(0.0015)	(0.0017)	(0.0015)	(0.0017)

Table A.2 (continued)

Dependent variable:	**Institutional Trust Index**			
	OLS		IV	
Sample:	Full	Migrants	Full	Migrants
	(1)	(2)	(3)	(4)
Settlement structure (baseline: Nomadic)				
Seminomadic	0.0069	0.0847**	-0.0044	0.0756**
	(0.0419)	(0.0390)	(0.0388)	(0.0358)
Semisedentary	-0.1157***	-0.1346***	-0.1410***	-0.1676***
	(0.0437)	(0.0440)	(0.0465)	(0.0487)
Compact but impermenant settlements	-0.1275**	-0.1573***	-0.1781***	-0.2290***
	(0.0574)	(0.0594)	(0.0643)	(0.0660)
Neighborhoods of dispersed family homesteads	-0.0478	-0.0595	-0.0792*	-0.0984**
	(0.0442)	(0.0392)	(0.0481)	(0.0424)
Separated hamlets, forming a single community	-0.0691	-0.0730*	-0.0909*	-0.1012**
	(0.0509)	(0.0429)	(0.0524)	(0.0459)
Compact and relatively permanent	-0.0751	-0.0763*	-0.1103**	-0.1244**
	(0.0481)	(0.0442)	(0.0530)	(0.0489)
Complex settlements	-0.1827***	-0.1403***	-0.2219***	-0.1952***
	(0.0562)	(0.0473)	(0.0590)	(0.0522)
Intensive agriculture	-0.0143	-0.0053	-0.0208*	-0.0160
	(0.0119)	(0.0116)	(0.0112)	(0.0110)
Share of subsistence from agriculture	-0.0042	0.0052	-0.0045	0.0041
	(0.0060)	(0.0049)	(0.0061)	(0.0053)
Indigenous slavery	0.0285	0.0091	0.0299	0.0100
	(0.0245)	(0.0187)	(0.0252)	(0.0209)
Slave exports	-0.0039	-0.0070	-0.0051	-0.0105
	(0.0065)	(0.0062)	(0.0069)	(0.0071)
Absolute latitude	0.0013	0.0014	0.0021	0.0028
	(0.0021)	(0.0022)	(0.0023)	(0.0024)
Average temperature	-0.0035	0.0028	-0.0051	0.0001
	(0.0037)	(0.0032)	(0.0037)	(0.0036)
Average precipitation	-0.0000	-0.0000*	-0.0000	-0.0000*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Ruggedness	0.0000	-0.0000	0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Log land area	-0.0132**	-0.0075	-0.0160**	-0.0118*
	(0.0064)	(0.0053)	(0.0068)	(0.0062)
Distance to sea coast	0.0044	0.0025	0.0012	-0.0026
	(0.0086)	(0.0069)	(0.0085)	(0.0077)
Distance to major rivers	-0.0011	-0.0051	-0.0013	-0.0058
	(0.0051)	(0.0044)	(0.0051)	(0.0048)
Calorie suitability index	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Observations	29,546	15,595	29,546	15,595
Number of clusters	147	145	147	145
Adjusted-R^2	0.244	0.243	0.244	0.242
KP F-stat of excluded instrument			40.05	42.73
Individual-level controls	✓	✓	✓	✓
Location controls	✓	✓	✓	✓
Ethnicity-level controls	✓	✓	✓	✓
Country FE	✓	✓	✓	✓

Notes: The unit of observation is an individual. The dependent variable is an individual's institutional trust index. The main explanatory variable is the ethnic group's precolonial political centralisation, based on the *Ethnographic Atlas*. Regressions in columns (1)-(2) are estimated with OLS. Regressions in columns (3)-(4) are estimated with two stage least squares, using the tsetse suitability index of the area historically inhabited by the ethnic group as the instrument. Odd-numbered columns employ the full sample of individuals, while even-numbered columns employ the subsample of migrants, defined as individuals not living on the territory of their ethnic group's historical homeland. Regressions are conditional on individuals controls, location controls, ethnicity-level controls and country fixed effects, as introduced in Table 1. Heteroskedasticity-robust standard errors, clustered at the ethnic group level, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table A.3. IV Estimates of the Effect of State History on Institutional Trust: Malaria Ecology Falsification

Dependent variable:	**Institutional Trust Index**			
Instrument:	Tsetse suitability		Malaria ecology	
Sample:	Full	Migrants	Full	Migrants
	(1)	(2)	(3)	(4)
Second stage estimates:				
Precolonial political centralisation	0.0760***	0.0865***	0.0371	-0.0707
	(0.0220)	(0.0227)	(0.0808)	(0.1220)
First stage estimates:				
Tsetse suitability index	-0.2408***	-0.2388***		
	(0.0381)	(0.0365)		
Malaria ecology index			-0.0185*	-0.0141
			(0.0105)	(0.0099)
Observations	29,546	15,595	29,546	15,595
Number of clusters	147	145	147	145
Adjusted-R^2	0.244	0.242	0.244	0.238
KP F-stat of excluded instrument	40.05	42.73	3.10	2.02
Individual-level controls	✓	✓	✓	✓
Location controls	✓	✓	✓	✓
Ethnicity-level controls	✓	✓	✓	✓
Country FE	✓	✓	✓	✓

Notes: The unit of observation is an individual. The dependent variable is an individual's institutional trust index. The main explanatory variable is the ethnic group's precolonial political centralisation, based on the *Ethnographic Atlas*. All regressions are estimated with two stage least squares, instrumenting with the tsetse suitability index of the area historically inhabited by the ethnic group in columns (1)-(2) and the malaria ecology index in columns (3)-(4). Odd-numbered columns employ the full sample of individuals, while even-numbered columns employ the subsample of migrants, defined as individuals not living on the territory of their ethnic group's historical homeland. Individual controls include age, age squared, a gender indicator, self-reported living conditions indicators, education indicators and religion indicators. Location controls include an urban area indicator, a set of 5 dummies for the presence of public goods (electricity, piped water, sewage system, school and health clinic) in the enumeration area, the district's ethnic fractionalization, the proportion of the district's population belonging to the same ethnic group as the individual and distance to the country capital. Ethnicity-level controls include precolonial era settlement structure, intensity of agriculture, share of subsistence from agriculture and relevance of indigenous slavery, cultural province indicators, time of ethnographic record indicators, number of slaves taken from the ethnic group during the slave trade, and absolute latitude, mean temperature, mean level of precipitation, ruggedness, log land area, distance to waterways and caloric suitability index of the ethnic group's historical homeland. All specifications include country fixed effects. Heteroskedasticity-robust standard errors, clustered at the ethnic group level, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table A.4. OLS and IV Estimates of the Effect of State History on Institutional Trust: Jurisdictional Hierarchies Beyond Local Community

Dependent variable:	**Institutional Trust Index**			
	OLS		IV	
Sample:	Full	Migrants	Full	Migrants
	(1)	(2)	(3)	(4)
Jurisdictional hierarchies beyond local community index	0.0177***	0.0184***	0.0549***	0.0698***
	(0.0056)	(0.0063)	(0.0168)	(0.0205)
First stage:				
Tsetse suitability index			-0.3331***	-0.2960***
			(0.0918)	(0.0872)
Observations	29,546	15,595	29,546	15,595
Number of clusters	147	145	147	145
Adjusted-R^2	0.244	0.243	0.013	0.002
KP F-stat of excluded instrument			13.18	11.52
Observations	29,546	29,546	15,595	15,595
Number of clusters	147	147	145	145
Individual-level controls	✓	✓	✓	✓
Location controls	✓	✓	✓	✓
Ethnicity-level controls	✓	✓	✓	✓
Country FE	✓	✓	✓	✓

Notes: The unit of observation is an individual. The dependent variable is an individual's institutional trust index. The main explanatory variable is the ethnic group's precolonial jurisdictional hierarchies beyond local community from the *Ethnographic Atlas*. Regressions in columns (1)-(2) are estimated with OLS. Regressions in columns (3)-(4) are estimated with two stage least squares, using the tsetse suitability index of the area historically inhabited by the ethnic group as the instrument. Odd-numbered columns employ the full sample of individuals, while even-numbered columns employ the subsample of migrants, defined as individuals not living on the territory of their ethnic group's historical homeland. Individual controls include age, age squared, a gender indicator, self-reported living conditions indicators, education indicators and religion indicators. Location controls include an urban area indicator, a set of 5 dummies for the presence of public goods (electricity, piped water, sewage system, school and health clinic) in the enumeration area, the district's ethnic fractionalization, the proportion of the district's population belonging to the same ethnic group as the individual and distance to the country capital. Ethnicity-level controls include precolonial era settlement structure, intensity of agriculture, share of subsistence from agriculture and relevance of indigenous slavery, cultural province indicators, time of ethnographic record indicators, number of slaves taken from the ethnic group during the slave trade, and absolute latitude, mean temperature, mean level of precipitation, ruggedness, log land area, distance to waterways and caloric suitability index of the ethnic group's historical homeland. All specifications include country fixed effects. Heteroskedasticity-robust standard errors, clustered at the ethnic group level, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table A.5. OLS and IV Estimates of the Effect of State History on Institutional Trust: City in 1800

Dependent variable:	**Institutional Trust Index**			
	OLS		IV	
Sample:	Full	Migrants	Full	Migrants
	(1)	(2)	(3)	(4)
City on ethnic homeland in 1800	0.0370**	0.0488***	0.1693***	0.1695***
	(0.0145)	(0.0144)	(0.0614)	(0.0555)
First stage:				
Tsetse suitability index			-0.1075***	-0.1217***
			(0.0289)	(0.0303)
Observations	29,566	15,615	29,566	15,615
Number of clusters	148	146	148	146
Adjusted-R^2	0.244	0.243	0.011	0.003
KP F-stat of excluded instrument			13.80	16.12
Individual-level controls	✓	✓	✓	✓
Location controls	✓	✓	✓	✓
Ethnicity-level controls	✓	✓	✓	✓
Country FE	✓	✓	✓	✓

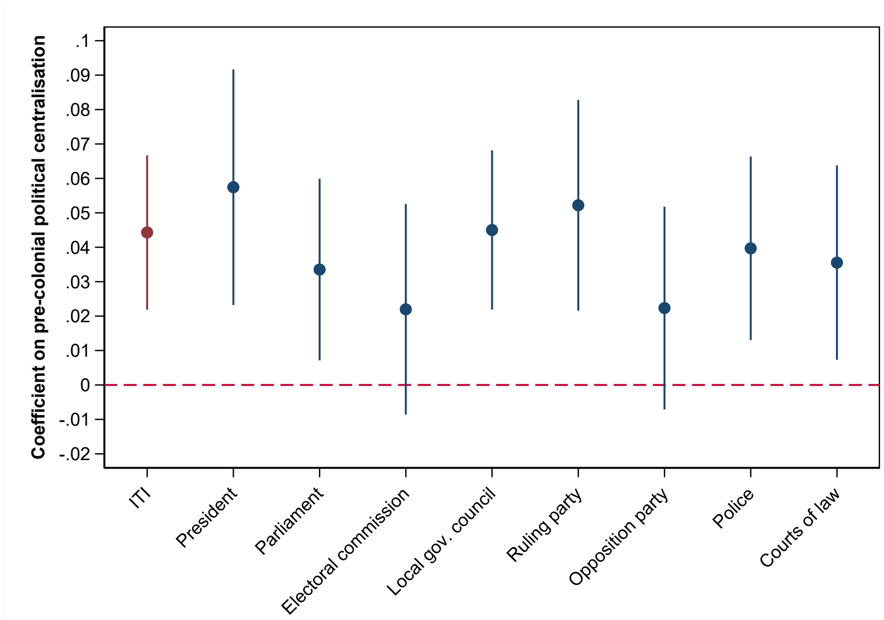
Notes: The unit of observation is an individual. The dependent variable is an individual's institutional trust index. The main explanatory variable is a dummy equalling one if there was a city with more than 20,000 inhabitants located on the land inhabited by the ethnic group in 1800. Regressions in columns (1)-(2) are estimated with OLS. Regressions in columns (3)-(4) are estimated with two stage least squares, using the tsetse suitability index of the area historically inhabited by the ethnic group as the instrument. Odd-numbered columns employ the full sample of individuals, while even-numbered columns employ the subsample of migrants, defined as individuals not living on the territory of their ethnic group's historical homeland. Individual controls include age, age squared, a gender indicator, self-reported living conditions indicators, education indicators and religion indicators. Location controls include an urban area indicator, a set of 5 dummies for the presence of public goods (electricity, piped water, sewage system, school and health clinic) in the enumeration area, the district's ethnic fractionalization, the proportion of the district's population belonging to the same ethnic group as the individual and distance to the country capital. Ethnicity-level controls include precolonial era settlement structure, intensity of agriculture, share of subsistence from agriculture and relevance of indigenous slavery, cultural province indicators, time of ethnographic record indicators, number of slaves taken from the ethnic group during the slave trade, and absolute latitude, mean temperature, mean level of precipitation, ruggedness, log land area, distance to waterways and caloric suitability index of the ethnic group's historical homeland. All specifications include country fixed effects. Heteroskedasticity-robust standard errors, clustered at the ethnic group level, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table A.6. OLS and IV Estimates of the Effect of State History on Institutional Trust: Alternative Institutional Trust Index

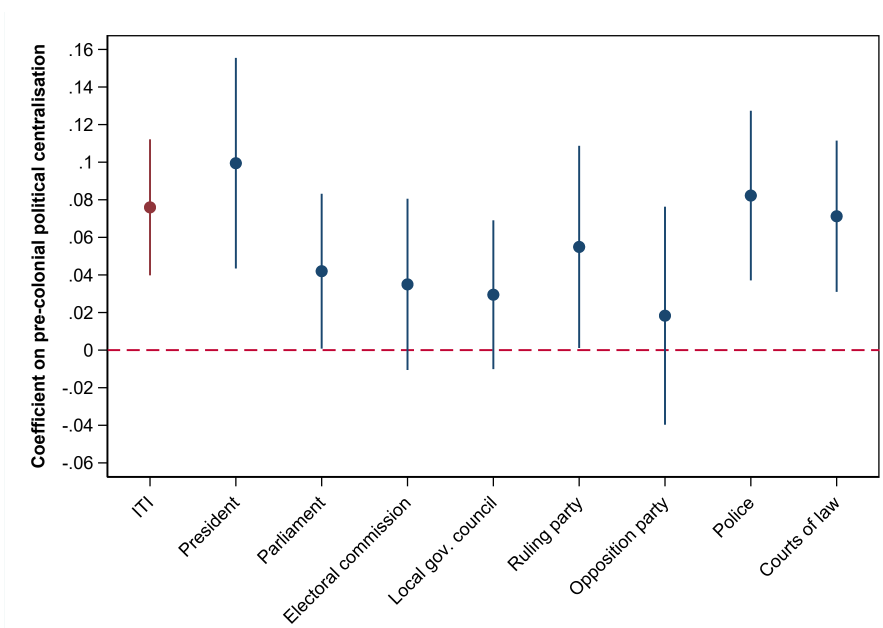
Dependent variable:	**Institutional Trust Index (alternative version)**			
	OLS		IV	
Sample:	Full	Migrants	Full	Migrants
	(1)	(2)	(3)	(4)
Precolonial political centralisation	0.0440***	0.0365***	0.0591***	0.0638***
	(0.0112)	(0.0117)	(0.0208)	(0.0204)
Observations	26,763	14,151	26,763	14,151
Number of clusters	147	145	147	145
Adjusted-R^2	0.264	0.265	0.017	0.009
KP F-stat of excluded instrument			38.39	41.02
Individual-level controls	✓	✓	✓	✓
Location controls	✓	✓	✓	✓
Ethnicity-level controls	✓	✓	✓	✓
Country FE	✓	✓	✓	✓

Notes: The unit of observation is an individual. The dependent variable is an individual's institutional trust index. The main explanatory variable is the ethnic group's precolonial political centralisation, based on the *Ethnographic Atlas*. Regressions in columns (1)-(2) are estimated with OLS. Regressions in columns (3)-(4) are estimated with two stage least squares, using the tsetse suitability index of the area historically inhabited by the ethnic group as the instrument. Odd-numbered columns employ the full sample of individuals, while even-numbered columns employ the subsample of migrants, defined as individuals not living on the territory of their ethnic group's historical homeland. Individual controls include age, age squared, a gender indicator, self-reported living conditions indicators, education indicators and religion indicators. Location controls include an urban area indicator, a set of 5 dummies for the presence of public goods (electricity, piped water, sewage system, school and health clinic) in the enumeration area, the district's ethnic fractionalization, the proportion of the district's population belonging to the same ethnic group as the individual and distance to the country capital. Ethnicity-level controls include precolonial era settlement structure, intensity of agriculture, share of subsistence from agriculture and relevance of indigenous slavery, cultural province indicators, time of ethnographic record indicators, number of slaves taken from the ethnic group during the slave trade, and absolute latitude, mean temperature, mean level of precipitation, ruggedness, log land area, distance to waterways and caloric suitability index of the ethnic group's historical homeland. All specifications include country fixed effects. Heteroskedasticity-robust standard errors, clustered at the ethnic group level, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Figure A.1. OLS and IV Estimates of the Effect of State History on Institutional Trust: Individual Trust Questions - Coefficient Plots



(A) OLS estimates



(B) IV estimates

Notes: Coefficient plots from separate regressions employing as dependent variable an individual's institutional trust index and responses to Afrobarometer survey questions on trust with respect to institutional figures. All dependent variables are normalised to be in [0,1]. Plotted coefficients refer to the estimated effect of the ethnic group's precolonial political centralisation, with 95% confidence interval bars from heteroskedasticity-robust standard errors, clustered at the ethnic group level. Regressions in sub-figure (A) are estimated with OLS. Regressions in sub-figure (B) are estimated with two stage least squares, using the tsetse suitability index of the area historically inhabited by the ethnic group as the instrument. All regressions are conditional on individuals controls, location controls, ethnicity-level controls and country fixed effects, as introduced in Table 1.

Table A.7. OLS and IV Estimates of the Effect of State History on Institutional Trust: Double-Clustered Standard Errors

Dependent variable:	**Institutional Trust Index**			
	OLS		IV	
Sample:	Full	Migrants	Full	Migrants
	(1)	(2)	(3)	(4)
Precolonial political centralisation	0.0443	0.0390	0.0760	0.0865
Ethnicity-District double-cluster s.e.	(0.0109)***	(0.0113)***	(0.0212)***	(0.0227)***
Ethnicity-Region double-cluster s.e.	[0.0106]***	[0.0109]***	[0.0169]***	[0.0191]***
Ethnicity-Country double-cluster s.e.	{0.0102}***	{0.0100}***	{0.0237}***	{0.0288}***
Observations	29,546	29,546	15,595	15,595
Number of ethnicity clusters	147	145	147	145
Number of district clusters	1,935	1,505	1,935	1,505
Number of region clusters	239	238	239	238
Number of country clusters	18	18	18	18
Adjusted-R^2	0.244	0.243	0.244	0.242
Individual-level controls	✓	✓	✓	✓
Location controls	✓	✓	✓	✓
Ethnicity-level controls	✓	✓	✓	✓
Country FE	✓	✓	✓	✓

Notes: The unit of observation is an individual. The dependent variable is an individual's institutional trust index. The main explanatory variable is the ethnic group's precolonial political centralisation, based on the *Ethnographic Atlas*. Regressions in columns (1)-(2) are estimated with OLS. Regressions in columns (3)-(4) are estimated with two stage least squares, using the tsetse suitability index of the area historically inhabited by the ethnic group as the instrument. Odd-numbered columns employ the full sample of individuals, while even-numbered columns employ the subsample of migrants, defined as individuals not living on the territory of their ethnic group's historical homeland. Individual controls include age, age squared, a gender indicator, self-reported living conditions indicators, education indicators and religion indicators. Location controls include an urban area indicator, a set of 5 dummies for the presence of public goods (electricity, piped water, sewage system, school and health clinic) in the enumeration area, the district's ethnic fractionalization, the proportion of the district's population belonging to the same ethnic group as the individual and distance to the country capital. Ethnicity-level controls include precolonial era settlement structure, intensity of agriculture, share of subsistence from agriculture and relevance of indigenous slavery, cultural province indicators, time of ethnographic record indicators, number of slaves taken from the ethnic group during the slave trade, and absolute latitude, mean temperature, mean level of precipitation, ruggedness, log land area, distance to waterways and caloric suitability index of the ethnic group's historical homeland. All specifications include country fixed effects. In parentheses are heteroskedasticity-robust standard errors adjusted for two-way clustering within ethnic groups and within districts. In square brackets are heteroskedasticity-robust standard errors adjusted for two-way clustering within ethnic groups and within regions. In curly brackets are heteroskedasticity-robust standard errors adjusted for two-way clustering within ethnic groups and within countries. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table A.8. OLS Estimates of the Effect of State History on Institutional Trust: Spatially-Adjusted Standard Errors

Dependent variable:	**Institutional Trust Index**			
Kernel:	Uniform		Conical	
Sample:	Full	Migrants	Full	Migrants
	(1)	(2)	(3)	(4)
Precolonial political centralisation	0.0443	0.0390	0.0443	0.0390
Conley s.e. - 250 km	(0.0096)***	(0.0110)***	(0.0103)***	(0.0113)***
Conley s.e. - 500 km	[0.0102]***	[0.0088]***	[0.0099]***	[0.0109]***
Conley s.e. - 1,000 km	{0.0081}***	{0.0076}***	{0.0093}***	{0.0093}***
Observations	29,546	15,595	29,546	15,595
Adjusted-R^2	0.244	0.243	0.244	0.242
Individual-level controls	✓	✓	✓	✓
Location controls	✓	✓	✓	✓
Ethnicity-level controls	✓	✓	✓	✓
Country FE	✓	✓	✓	✓

Notes: The unit of observation is an individual. All regressions are estimated with OLS. The dependent variable is an individual's institutional trust index. The main explanatory variable is the ethnic group's precolonial political centralisation, based on the *Ethnographic Atlas*. Odd-numbered columns employ the full sample of individuals, while even-numbered columns employ the subsample of migrants, defined as individuals not living on the territory of their ethnic group's historical homeland. Individual controls include age, age squared, a gender indicator, self-reported living conditions indicators, education indicators and religion indicators. Location controls include an urban area indicator, a set of 5 dummies for the presence of public goods (electricity, piped water, sewage system, school and health clinic) in the enumeration area, the district's ethnic fractionalization, the proportion of the district's population belonging to the same ethnic group as the individual and distance to the country capital. Ethnicity-level controls include precolonial era settlement structure, intensity of agriculture, share of subsistence from agriculture and relevance of indigenous slavery, cultural province indicators, time of ethnographic record indicators, number of slaves taken from the ethnic group during the slave trade, and absolute latitude, mean temperature, mean level of precipitation, ruggedness, log land area, distance to waterways and caloric suitability index of the ethnic group's historical homeland. All specifications include country fixed effects. Conley (1999) standard errors adjusted for two-dimensional spatial correlation are employed, considering windows of 250 km (in parentheses), 500 km (in square brackets) and 1,000 km (in curly brackets). In columns (1)-(2), the adjustment for spatial correlation uses an uniform kernel, assigning equal weight to all observations inside the window, and zero weight to observations further apart. In columns (3)-(4), the adjustment for spatial correlation uses a conical kernel, with weights decaying linearly with distance. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table A.9. OLS and IV Estimates of the Effect of State History on Institutional Trust: Country-Year Fixed Effects

Dependent variable:	**Institutional Trust Index**			
	OLS		IV	
Sample:	Full	Migrants	Full	Migrants
	(1)	(2)	(3)	(4)
Precolonial political centralisation	0.0301***	0.0270**	0.0530**	0.0626***
	(0.0115)	(0.0116)	(0.0219)	(0.0208)
Observations	29,546	15,595	29,546	15,595
Number of clusters	147	145	147	145
Adjusted-R^2	0.264	0.258	0.264	0.257
KP F-stat of excluded instrument			40.50	43.67
Individual-level controls	✓	✓	✓	✓
Location controls	✓	✓	✓	✓
Ethnicity-level controls	✓	✓	✓	✓
Country-year FE	✓	✓	✓	✓

Notes: The unit of observation is an individual. The dependent variable is an individual's institutional trust index. The main explanatory variable is the ethnic group's precolonial political centralisation, based on the *Ethnographic Atlas*. Regressions in columns (1)-(2) are estimated with OLS. Regressions in columns (3)-(4) are estimated with two stage least squares, using the tsetse suitability index of the area historically inhabited by the ethnic group as the instrument. Odd-numbered columns employ the full sample of individuals, while even-numbered columns employ the subsample of migrants, defined as individuals not living on the territory of their ethnic group's historical homeland. Individual controls include age, age squared, a gender indicator, self-reported living conditions indicators, education indicators and religion indicators. Location controls include an urban area indicator, a set of 5 dummies for the presence of public goods (electricity, piped water, sewage system, school and health clinic) in the enumeration area, the district's ethnic fractionalization, the proportion of the district's population belonging to the same ethnic group as the individual and distance to the country capital. Ethnicity-level controls include precolonial era settlement structure, intensity of agriculture, share of subsistence from agriculture and relevance of indigenous slavery, cultural province indicators, time of ethnographic record indicators, number of slaves taken from the ethnic group during the slave trade, and absolute latitude, mean temperature, mean level of precipitation, ruggedness, log land area, distance to waterways and caloric suitability index of the ethnic group's historical homeland. All specifications include country-year fixed effects. Heteroskedasticity-robust standard errors, clustered at the ethnic group level, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table A.10. OLS and IV Estimates of the Effect of State History on Institutional Trust: Region Fixed Effects

Dependent variable:	**Institutional Trust Index**			
	OLS		IV	
Sample:	Full	Migrants	Full	Migrants
	(1)	(2)	(3)	(4)
Precolonial political centralisation	0.0342***	0.0230**	0.0635***	0.0539***
	(0.0093)	(0.0099)	(0.0175)	(0.0198)
Observations	29,546	15,595	29,546	15,595
Number of clusters	147	145	147	145
Adjusted-R^2	0.271	0.272	0.270	0.272
KP F-stat of excluded instrument			38.94	36.84
Individual-level controls	✓	✓	✓	✓
Location controls	✓	✓	✓	✓
Ethnicity-level controls	✓	✓	✓	✓
Country FE	✓	✓	✓	✓
Region FE	✓	✓	✓	✓

Notes: The unit of observation is an individual. The dependent variable is an individual's institutional trust index. The main explanatory variable is the ethnic group's precolonial political centralisation, based on the *Ethnographic Atlas*. Regressions in columns (1)-(2) are estimated with OLS. Regressions in columns (3)-(4) are estimated with two stage least squares, using the tsetse suitability index of the area historically inhabited by the ethnic group as the instrument. Odd-numbered columns employ the full sample of individuals, while even-numbered columns employ the subsample of migrants, defined as individuals not living on the territory of their ethnic group's historical homeland. Individual controls include age, age squared, a gender indicator, self-reported living conditions indicators, education indicators and religion indicators. Location controls include an urban area indicator, a set of 5 dummies for the presence of public goods (electricity, piped water, sewage system, school and health clinic) in the enumeration area, the district's ethnic fractionalization, the proportion of the district's population belonging to the same ethnic group as the individual and distance to the country capital. Ethnicity-level controls include precolonial era settlement structure, intensity of agriculture, share of subsistence from agriculture and relevance of indigenous slavery, cultural province indicators, time of ethnographic record indicators, number of slaves taken from the ethnic group during the slave trade, and absolute latitude, mean temperature, mean level of precipitation, ruggedness, log land area, distance to waterways and caloric suitability index of the ethnic group's historical homeland. All specifications include country and region fixed effects. Heteroskedasticity-robust standard errors, clustered at the ethnic group level, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Appendix B - Model

To guide my analysis, I develop a model, showing how reliance on tradition can cause beliefs to reflect experiences of past generations. The model predicts that, by relying on tradition, individuals with an ancestry of state history will be more likely to have positive attitudes towards the state. The idea builds on insights from the cultural anthropology literature, postulating that when information acquisition is either costly or imperfect, individuals employ heuristic decision-making strategies (Boyd and Richerson, 1985, 1995, 2005).

There are three periods of time: period 0 (pre-past), period 1 (past) and period 2 (present). The model is constituted by Individuals living in Societies, with Nature randomly drawing Societies' initial conditions and Individuals' types. There is a continuum of Individuals in each Society, and each period a new generation is born while the old one dies.

B.1 Nature

Nature exogenously determines the initial conditions e (environment) of each Society, with e_s being the environment of Society s. The environment e is chosen from a given distribution with PDF $g(.)$ and support $[\underline{\omega}, \overline{\omega}]$, with values of e tending to $\underline{\omega}$ indicating an environment more favourable to state formation and values of e tending to $\overline{\omega}$ indicating an environment less favourable to state formation. Once e_s is chosen by Nature, it stays the same for all periods. Nature also exogenously decides the type θ of Individuals, according to a binomial distribution with PMF $b(.)$. Individuals can be of two types, traditionalists (T) and information-seekers (I), with both types having an equal probability of being born. Individuals' types are not correlated across generations.

B.2 Societies

Statehood. A Society can have two statehood types μ. It can either be stateless (0) or have a state (1), such that:

$$\mu \in \{0; 1\}$$

During period 0, all Societies are stateless, such that $\mu_0 = \{0\}$ for every s. Societies in period 1 can either have a state (1) or be stateless (0), such that $\mu_1 \in \{0; 1\}$. The Society's environment e_s determines of which type Society s is going to be at the beginning of period 1. There is a threshold Δ such that if the value of e_s is below (above) this threshold, then Society s is going to have a state (be stateless), therefore:

$$\mu_1 = \begin{cases} 1 & if \quad e_s < \Delta \\ 0 & if \quad e_s \geq \Delta \end{cases}$$

In period 2, there is a shock happening for each Society (one possible interpretation being colonialism), such that every Society receives a state, i.e. $\mu_2 = \{1\}$ for every s. An example of the possible evolution of two Societies' statehood μ is given by:

	μ of Society i	*μ of Society j*
Period 0	0	0
Period 1	1	0
Period 2	1	1

Goodness of state. There are two types η of state, good (G) and bad (B):

$$\eta \in \ \{G; B\}$$

At the beginning of period 1 both good and bad states can emerge, however only good states survive through the rest of the period. This is due to the fact that, outside of period 2, bad states are unsustainable and experience a collapse. In case a bad state emerges, the Society will therefore revert back to being stateless, and stay like that for the rest of period 1. Thus, in period 1 Societies can either be stateless ($\mu_1 = 0$) or have a good state ($\mu_1 = 1, \ \eta_1 = G$).[9]

During period 2 all Societies become states regardless of their type in the previous period. Differently from period 1, both good and bad states can exist in period 2, i.e. $\eta_2 \in \{G; B\}$.

[9]The period 1 assumption of bad states always collapsing and good states always surviving could be relaxed, for example with both types of state having a positive probability to collapse and bad states collapsing with a higher probability than good states. Given that, in the context of this model, relaxing the assumption would yield the same results intuition-wise, the current extreme assumption is chosen for the sake of simplicity.

	Period 0	*Period 1*		*Period 2*
Statehood (μ)	0	0	1	1
Goodness of state (η)	-	-	G	G/B

In case a Society was already a state in period 1 ($\mu_1 = 1$), then the goodness of the state in period 2 (η_2) is correlated with the goodness of the state in period 1 (η_1) by a factor $\rho > 0$. In other words, since only good states can survive in period 1, a Society that is a state in both periods has a higher chance to be a good state in period 2, compared to a Society that was stateless in period 1, therefore:

$$Pr(\eta_2 = G \mid \mu_1 = 1, \mu_2 = 1) > Pr(\eta_2 = G \mid \mu_1 = 0, \mu_2 = 1)$$

B.3 Individuals

There are three generations of Individuals, with each generation living in one of the three periods. Individuals can have two kinds of beliefs σ, either that the state is good (P) or that the state is bad (N), such that:

$$\sigma \in \{P; N\}$$

There are two types θ of Individuals, traditionalists (T) and information-seekers (I), such that:

$$\theta \in \{T; I\}$$

The type θ of an Individual is independent of his parents' type in the previous generation, therefore it is possible for a traditionalist to descend from an info-seeker and vice-versa. Info-seekers form their beliefs based on observation and judgement, therefore based on the actual goodness of the state η. If no information is available about the goodness of the state, info-seeker will form a belief that the state is bad.[10] The different cases are:

i. Info-seekers ($\theta = I$) living in a good state ($\eta = G$) will form a belief that the state is good ($\sigma = P$);

[10] Similarly to the assumption regarding the collapse of states in period 1, one could relax the assumption on the Individuals' prior belief of the state, without changing the model's results. A more generalised version would have the prior belief being that the state is bad only with some positive probability.

ii. Info-seekers ($\theta = I$) living in a bad state ($\eta = B$) or stateless Society ($\mu = 0$) will form a belief that the state is bad ($\sigma = N$);

Traditionalists, on the other hand, are unable to acquire information. Instead, they employ a heuristic decision-making strategy where they imitate the belief formed by their previous generation. Their belief is independent of the existence or goodness of the state. Some examples are:

i. A traditionalist ($\theta = T$) living in a good state ($\eta = G$) but descending from someone with a bad belief of the state ($\sigma_{t-1} = N$) will also believe that the state is bad ($\sigma_t = N$);

ii. A traditionalist ($\theta = T$) living in a bad state ($\eta = B$) but descending from someone with a good belief of the state ($\sigma_{t-1} = P$) will also believe that the state is good ($\sigma_t = P$);

Individuals from the first generation, living in period 0, can be both info-seekers or traditionalists. However, regardless of their type θ all Individuals will have a belief that the state is bad ($\sigma = N$), given that during period 0 all Societies are stateless ($\mu_0 = 0$) and no information is available about states and their goodness. After the emergence of states is determined by the environment e at the beginning of period 1, Individuals from the second generation form their beliefs. If a Society remains stateless in period 1, info-seekers from that generation will believe that the state is bad. In the case a good state emerges and survives, info-seekers will update and form a belief that the state is good. Traditionalists, on the other hand, will not update their beliefs from the previous generation and keep believing that the state is bad, regardless of what kind of Society turns out in period 1, since they can only imitate bad beliefs about the state. In period 2, when all Societies become states, which can be either good or bad, Individuals from the third generation form their beliefs. Info-seekers will have a good (bad) belief if the Society they live in has a good (bad) state. Traditionalist, similarly to the ones from the previous period, form good or bad beliefs by imitating their parents.

B.4 Mechanism and Prediction

The outcomes for each period are:

Period 0: Both info-seekers ($\theta = I$) and traditionalists ($\theta = T$) believe that the state is bad ($\sigma = N$).

Period 1: Info-seekers ($\theta = I$) can either have good ($\sigma = P$) or bad ($\sigma = N$) beliefs about the state, depending on whether they live in a state ($\mu_1 = 1$) or not ($\mu_1 = 0$). Traditionalists ($\theta = T$), on the other hand, can only have bad beliefs ($\sigma = N$) about the state in this period.

Period 2: Info-seekers ($\theta = I$) can either have good ($\sigma = P$) or bad ($\sigma = N$) beliefs about the state, depending on whether they live in a good state ($\eta_2 = G$) or in a bad state ($\eta_2 = B$). Traditionalists ($\theta = T$) can either have good ($\sigma = P$) or bad ($\sigma = N$) beliefs about the state, depending on who they descend from ($\sigma_t = \sigma_{t-1}$). A period 2 traditionalist will believe the state is bad if descending from a period 1 traditionalist or from a period 1 info-seeker who lived in a stateless Society. A period 2 traditionalist will believe the state is good if and only if descending from a period 1 info-seeker who lived in a good state.

Based on the two assumptions that traditionalists have beliefs that reflect those of the previous generation and that the cross-period correlation in goodness of state ρ is greater than zero, the model predicts that Individuals from Societies that were characterised by a state for longer will be more likely to have a good belief of the state. In other words, Individuals from Societies with a "state history", i.e. Societies characterised by a state in both periods 1 and 2, will be more likely to believe that the state is good in period 2, because of these two reasons:

1. Due to the cross-period correlation in goodness of state ρ, societies having states in both periods 1 and 2 will be more likely to be good states in period 2, making info-seekers more likely to have a good belief.
2. Due to the traditionalists' heuristic decision-making, period 2 traditionalists in societies with a "state history" will have a chance of descending from a period 1 info-seeker with a good belief, therefore making traditionalists more likely to have a good belief.

Based on the mechanism of tradition, I formulate a general hypothesis that I bring to the data.

Hypothesis *Reliance on tradition causes beliefs to reflect experiences of past generations. The presence of traditionalists makes descendants from societies with a "state history" more likely to believe that the state is good.*

The other mechanism explained above also allows me to highlight a challenge I will address in the analysis, i.e. the need to distinguish the effect working through the traditionalists' beliefs from the effect working through the cross-period correlation ρ.

Appendix C - Variables

C.1 Institutional Trust Index

This section describes the construction of the Institutional Trust Index in detail. The index is constructed by aggregating via PCA the answers to four questions from the Afrobarometer surveys, being the reported trust with respect to: the president, the parliament, the police and courts of law. For each of these variables, the answer categories were (i) "Not at all", (ii) "Just a little", (iii) "Somewhat", and (iv) "A lot". An overview of the responses to these trust questions is provided in Table C.1.

Table C.1. Distribution of Responses to the Institutional Trust Questions

	How much do you trust each of the following:			
Response	President	Parliament	Police	Courts of law
Not at all	15.97%	17.77%	22.62%	13.74%
Just a little	20.36%	25.06%	23.56%	23.67%
Somewhat	20.54%	26.37%	23.66%	27.7%
A lot	43.13%	30.81%	30.17%	34.89%
Total	35,932	35,007	36,232	35,386

Notes: The table reports the distribution of responses to four measures of institutional trust from the 2005 and 2008 rounds of Afrobarometer surveys. The unit of observation is an individual.

Given that the answers are categorical, I convert them into a variable assigning a number to each category. The resulting variables take on values going from 0 to 3, where 0 corresponds to "Not at all", 1 to "Just a little", 2 to "Somewhat" and 3 to "A lot".

The Institutional Trust Index is the first principal component of these four trust questions, normalised to be in $[0, 1]$. The first principal component explains 0.62 of total variance and has an eigenvalue of 2.49, while the second component has an eigenvalue of 0.75. The Institutional Trust Index loads positively on trust of the president (0.5006), positively on trust of the parliament (0.5107), positively on trust of the police (0.4956) and positively on trust of courts of law (0.4929). Figure C.2 illustrates the distribution of the index across the countries in the sample.

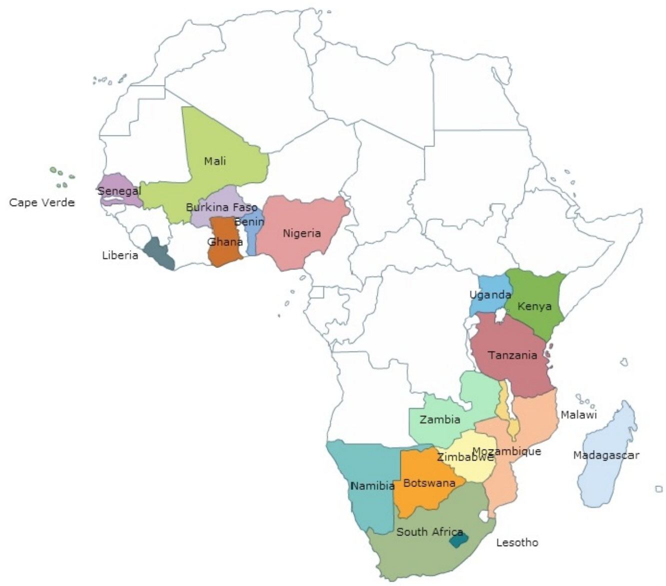


Figure C.1. Countries included in the 2008 round of the Afrobarometer surveys

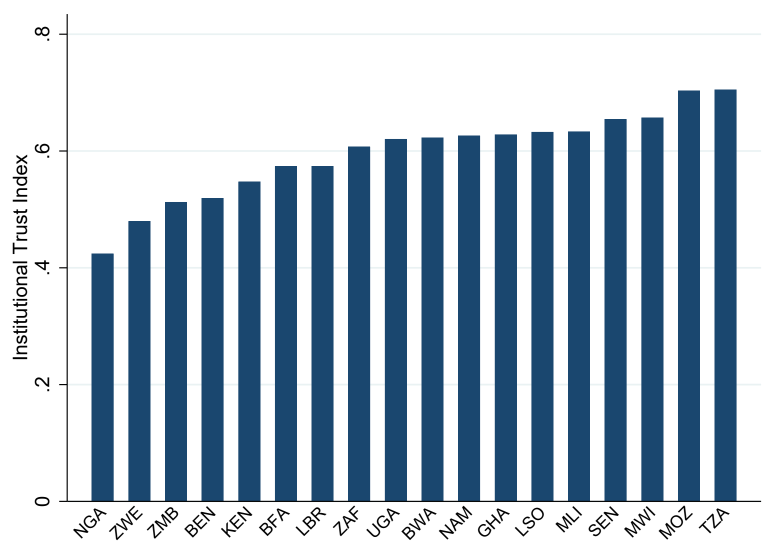


Figure C.2. Institutional Trust Index by country

C.2 Migrants

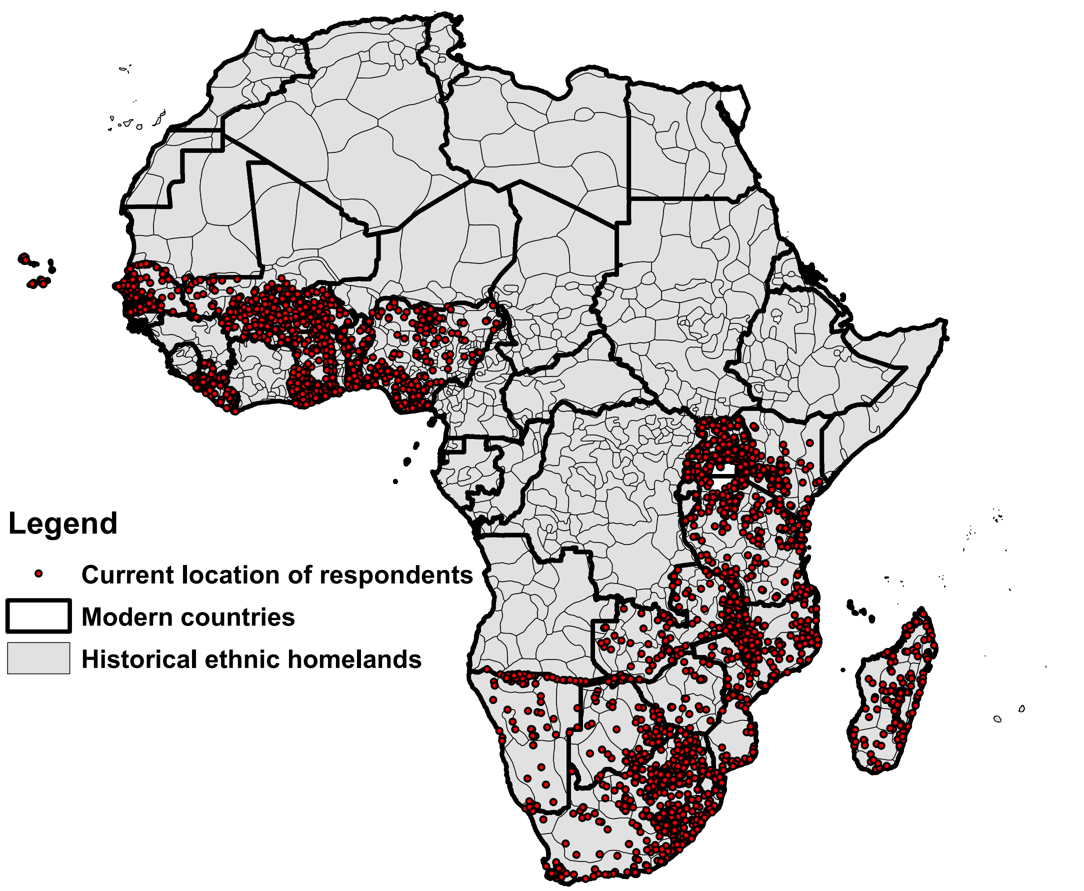


FIGURE C.3. Historical ethnic homelands from Murdock's (1959) ethnolinguistic map and current location of respondents in the Afrobarometer surveys.

Table C.2. Migrants Strategy: Comparative Statistics

Sample:	**Migrants**		**Non-Migrants**	
Variable	*Mean*	*(SD)*	*Mean*	*(SD)*
Modern data				
Institutional Trust Index	0.58	(0.29)	0.60	(0.29)
Age	36	(14.23)	37	(14.84)
Male	0.50	(0.50)	0.50	(0.50)
Urban area	0.42	(0.49)	0.31	(0.46)
Education	3.26	(2.04)	2.96	(2.00)
Living conditions	2.59	(1.22)	2.56	(1.21)
School in enumeration area	0.83	(0.38)	0.83	(0.38)
Electricity in enumeration area	0.58	(0.49)	0.47	(0.50)
Piped water in enumeration area	0.51	(0.50)	0.45	(0.50)
Sewage in enumeration area	0.26	(0.44)	0.17	(0.37)
Health clinic in enumeration area	0.56	(0.50)	0.54	(0.50)
District-level ethnic fractionalization	0.45	(0.29)	0.31	(0.29)
Proportion of ethnic group in district	0.52	(0.35)	0.74	(0.30)
Distance to the capital (in 1,000 kms)	0.30	(0.28)	0.29	(0.27)
Ethnic group-level data				
Precolonial political centralisation	0.41	(0.49)	0.45	(0.50)
Share of subsistence from agriculture	5.78	(1.49)	5.74	(1.45)
Intensive agriculture	0.28	(0.45)	0.27	(0.45)
Indigenous slavery	0.85	(0.36)	0.83	(0.38)
Settlement structure	5.79	(1.78)	5.83	(1.72)
Year of ethnographic record	1916	(22.68)	1915	(23.51)
Slave exports (ln(1+exports/area))	0.31	(0.71)	0.31	(0.67)
Tsetse suitability index	0.00	(0.98)	-0.06	(1.02)
Calorie suitability index	0.76	(0.19)	0.76	(0.19)
Absolute latitude	9.54	(6.65)	9.49	(7.08)
Log land area	23.74	(1.13)	23.80	(1.06)
Distance to sea coast (in 1,000 kms)	0.50	(0.37)	0.49	(0.36)
Distance to major rivers (in 1,000 kms)	0.07	(0.06)	0.07	(0.06)
Average temperature (°C)	25.19	(2.99)	25.04	(3.06)
Average precipitation (m)	1.02	(0.58)	1.04	(0.61)
Ruggedness	0.13	(0.15)	0.14	(0.16)

Notes: The table compares mean and standard deviation of variables from the migrants and non-migrants subsamples. The former subsample is composed of individuals who are not living on the territory of their ethnic group's historical homeland. The latter subsample is composed of individuals who are living on the territory of their ethnic group's historical homeland.

C.3 Further Ethnographic Atlas Controls

C.3.1 Year of Ethnographic Record

The year of ethnographic record, one of the variables from the *Ethnographic Atlas*, provides for each ethnic group the time of sampling. Figure C.4 provides an illustration of the distribution of the ethnographic records over the years.

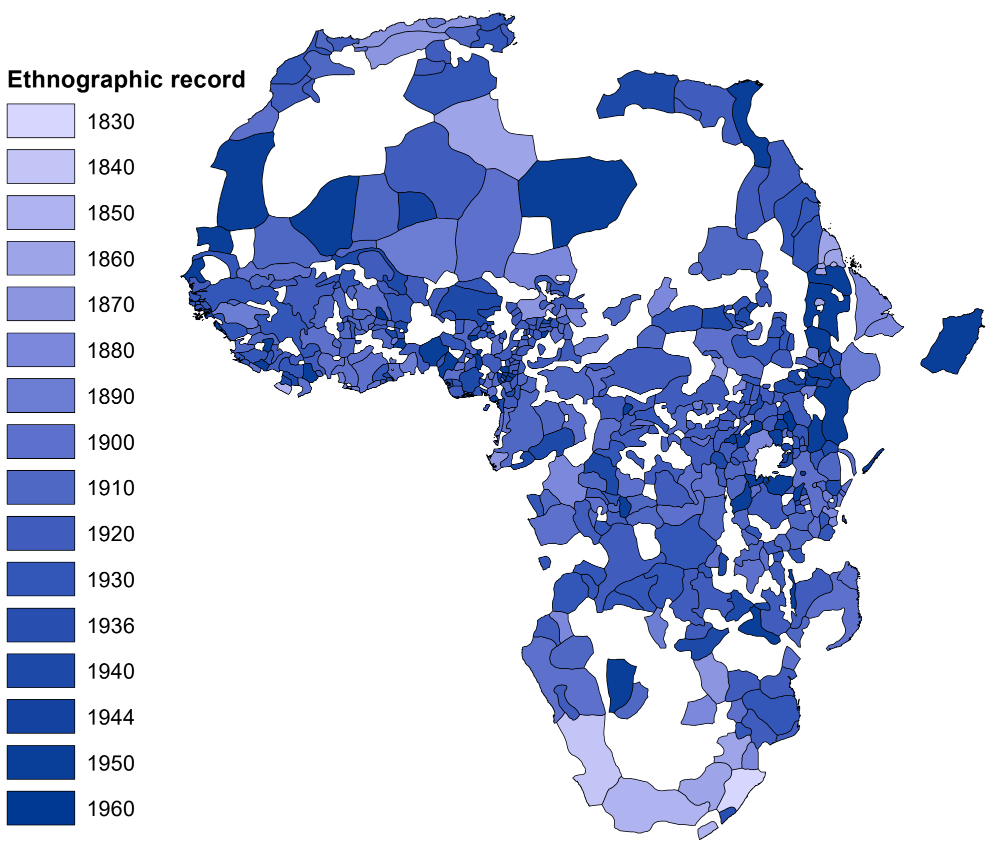


FIGURE C.4. Year of ethnographic record from Murdock's (1967) *Ethnographic Atlas*.

A concern is whether the groups that were recorded later resulted as being politically developed because they were more likely to have been influenced by the colonisers by the time the record took place. Regarding this concern, it would be useful to have information on how much time passed between the time an ethnic group first came in contact with Europeans and the time the ethnographic record took place. While this information is not available, looking at the correlation between the year of observation and the measure of precolonial political centralisation reveals a statistically significant and negative relationship ($\rho = -0.2390$). While this negative correlation does not confirm a lack of colonial influence, it implies that the more institutionally developed groups tended to be recorded earlier. Moreover, the year of observation is significantly and negatively correlated with the presence of a coast on the ethnic homeland's boundaries ($\rho = -0.1703$), coherently with the historical narrative. On the other hand, precolonial political centralisation is not significantly correlated with the presence of a coast($\rho = 0.0479$).

C.3.2 Cultural Provinces

Cultural provinces are groupings used in Murdock (1967), capturing spatial, cultural and genealogical correlation of ethnic groups.

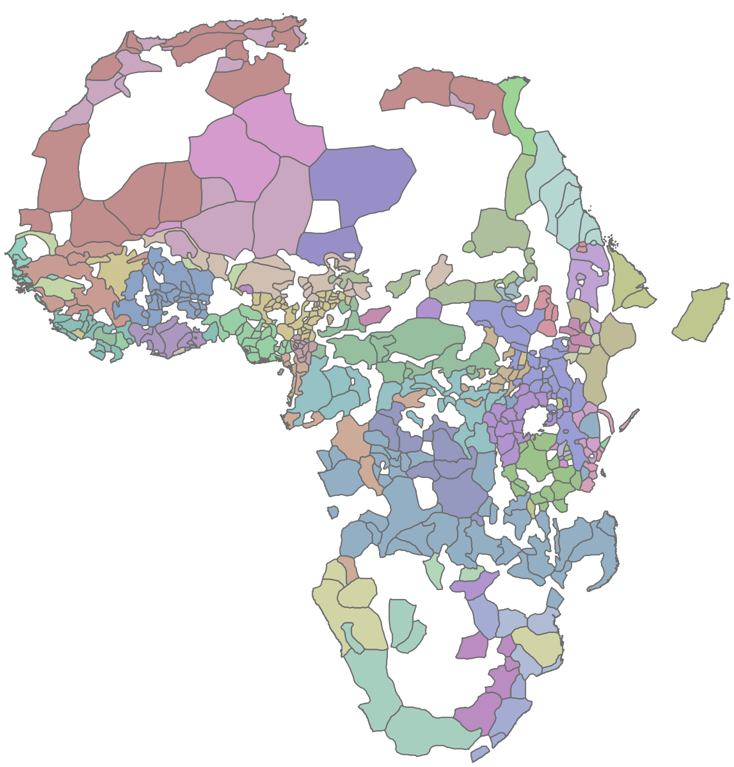


FIGURE C.5. Cultural provinces in Murdock's (1967) *Ethnographic Atlas*, with shading used to represent the 44 provinces.

CHAPTER II

THE POLITICAL ECONOMY OF THE ITALIAN RAILWAY EXPANSION, 1879-1890*

*Co-authored with Roberto Bonfatti, Giovanni Facchini, Alexander Tarasov and Cecilia Testa

1 Introduction

There is a widespread belief that a well-designed transport network is at the heart of a flourishing economy, and indeed transport infrastructure takes up a large share of public investments around the world. Motivated by these facts, a large literature has attempted to evaluate the impact of transport infrastructure. One typical approach has been to exploit some source of quasi-experimental variation in its placement, allowing the researcher to identify its impact on various economic outcomes. This literature has typically found that (a) infrastructure has large positive effects on economic activities; (b) the actual process of allocating infrastructure is not random, suggesting that political factors could play an important role in the choice of its location (for an overview, see Redding and Turner 2015). Yet, work on whether and how politics shapes infra-structural development remains scant.[1] Importantly, since investment in infrastructure has been shown to have both short- and long-run economic consequences, then if political factors affect investment decisions, short-run political circumstances may shape long-term local development in a significant way.[2] In this paper we make progress on these important questions by analysing the political determinants of a major Italian railway expansion between 1879 and 1890, and their long-term effects on economic growth.

This is an ideal setting for a number of reasons. First, the network expansion set out by the so-called "Baccarini Law" of 1879 sparkled the second Italian railroad "boom" (Fenoaltea 1983), almost doubling the size of the country's network in a short period of time. The program had very significant implications for the finances of the new state. Its total cost for the government was initially estimated at around 1.26 billion Liras (i.e. roughly comparable to the total revenues collected in that year) but eventually came to almost twice that amount (Schram 1997, page 111, Ferrucci 1898).[3] Second, most of the railways built in this period were designed to connect smaller cities to the long-distance lines built in the previous decades. The local railways realized under the Baccarini Law were often built on difficult terrain and were of great practical utility, as on many of those routes no alternative means of transportation existed (Schram 1997). This gives us rich, within-country variation in the decision to build, or not, a new line. Finally, historians have long argued that in the allocation of these lines electoral considerations played an important role (e.g. see Tajani 1944, Mercurio

[1] For an early contribution, see Knight (2005). More recent work include Curto–Grau, Herranz–Loncan, and Sole–Olle (2012), Burgess, Jedwab, Miguel, Morjaria, and i Miquel (2015), Alder and Kondo (2019) and Selod and Soumahoro (2019).

[2] For an overview of the short run effects, see Redding and Turner (2015). As for the long-run consequences of railroad development, see, for instance, Hornung (2015), Jedwab and Moradi (2016), Berger and Enflo (2017), Banerjee, Duflo, and Qian (2020) and Büchel and Kyburz (2020).

[3] Total revenues in 1879 were estimated at 1,228 billion Liras, see Houghton (1889).

1994, Schram 1997).

Political economy models emphasize that political actors shape outcomes taking into account both their own preferences and those of their constituencies (Besley 2006). A large literature points out that those in a position of power are able to bring the 'bacon' home (see Larcinese, Snyder, and Testa 2013 for an overview). Thus, one key implication of models of pork-barrel politics is that constituencies represented by elected officials in government are more likely to be recipients of distributive goods – such as for example transport infrastructure – and if elected government officials reward their partizan supporters, then areas that voted more for government candidates are more likely to attract investment spending. Importantly, if the latter affects economic growth, then short run pork-barrel politics may have persistent economic effects. In this paper we bring these predictions to the data to shed light on whether (a) support for pro-government candidates affects the location of railway infrastructure and (b) political factors prevailing right after the Italian unification, by shaping infrastructural development, have long-term effects on economic growth.

To this end, we proceed in two steps. First, we study how competition among representatives affects their ability to bring infrastructure to their constituency; second, we analyse how rivalry among municipalities shapes the actual path followed by a railway line. Our key hypothesis is that alignment with powerful politicians will increase the likelihood of obtaining a railroad. To identify the causal effect of alignment, we use a regression discontinuity design focusing on close elections, i.e. on districts in which a pro-government candidate won or lost by a narrow margin. The railway construction plans on which we focus were laid out in the 1879 Baccarini Law. The legislation determined the lines to be built by including a list of city-pairs that would have to be connected, but it did not specify the exact route that each line would have to follow. In the first step, we focus on district-level electoral outcomes to determine whether municipalities belonging to districts that narrowly elected a pro-government candidate in 1876 (the election shaping the composition of parliament when the Baccarini Law was approved) were in closer geographic proximity to a Baccarini Law line, compared to those narrowly won by a pro-opposition candidate. In the second step, we focus instead on the actual construction of the railways over the subsequent three legislatures (1880-82, 82-86, 86-90), and on the exact route that was chosen for them. For this analysis, we exploit detailed information on voting in approximately 7,000 municipalities. We attain identification by restricting the sample to municipalities included in marginal districts, and in close proximity to the Baccarini Law railways. Within this restricted sample, we ask whether municipalities were more likely to receive a railway depending on the extent to which they voted for the district's winning candidate, and whether or not the latter belonged to the government party.

Our results support the hypothesis that politics played an important role both in terms of which lines were included in the state-planned railway expansion and in determining the path followed by the railways that were built. In particular, regression discontinuity estimates comparing municipalities belonging to electoral districts where the government marginally lost to ones where the government marginally won reveal that Baccarini Law lines are significantly closer to municipalities in districts carried by the government. Results are robust to the inclusion of a rich set of municipality-level controls, fixed effects and to varying RD specifications. Furthermore, studying the actual construction of the planned railways through the subsequent period reveals that a municipality's alignment with national politicians played a very important role: first, we show that municipalities strongly supporting members of parliament elected in their district are significantly more likely to see a railway crossing their territory; second, we find that such support is particularly relevant if the elected candidate was affiliated with the government party. These findings hold across the three different elections included in the analysis and two electoral systems.

In the second part of the paper, building on a rich literature studying the contribution of railways, or more in general transport infrastructure, to economic development, we study the long-run consequences of the railway expansion promoted by the Baccarini Law. In particular, we ask whether the impact of politics on the shape of the railway network expansion that we study in the first part of the paper had long-lasting consequences on the subsequent growth path of municipalities. To address this question, we exploit detailed municipality-level population data covering the period 1861-1991 to study whether long-run population dynamics were impacted by railway access.

We deploy a variety of strategies to attain identification. First, we use a rich set of pre-determined municipality characteristics and demanding fixed effects specifications. Second, to further improve our comparison of municipalities with similar characteristics and subject to similar conditions, we focus our analysis on the sample of municipalities that are in close proximity to the railway lines. Third, we exploit the political conditions that we argue had affected construction of the railways by further restricting our sample to municipalities that belonged to electoral district interested by marginal elections at the time when construction was being decided, and therefore where the role of pork-barrelling was more salient. Fourth, building on the previous strategy, we exploit municipality-level voting outcomes when the path to be followed by railways was decided to instrument for municipalities gaining railway access.

Through our results we show that municipalities that gained railway access thanks to one of the lines belonging to the Baccarini Law expansion experienced significantly higher population growth in the century following their construction. In particular, the analysis reports that significant, although contained, differences in population growth were already present at the beginning of the century, with growth differentials around 5%. A larger impact of railway access can instead be detected in the second half of the twentieth century, with differences in population growth reaching up to 50% by 1991.

This paper is related to two strands of research. First, it speaks to the literature on distributive politics and in particular to the comparatively small body of work that has analysed the allocation of transport infrastructure.[4] Most of the studies carried out in this tradition use panel fixed effect models, whereas we deploy a regression discontinuity design, enabling us to more convincingly uncover the causal effect of politics. In an early contribution, Knight (2005) shows that politicians enjoying positions of power - i.e. members of the US House Committee on Transportation and Infrastructure – are able to secure more infrastructure funding for their districts in the 1991-1998 period. Similar patterns were also uncovered by Curto–Grau, Herranz–Loncan, and Sole–Olle (2012) in their analysis of Spanish road infrastructure projects in the late XIX- early XX century. Burgess, Jedwab, Miguel, Morjaria, and i Miquel (2015) study instead the role played by ethnic favoritism on the allocation of roads in Kenya after independence, finding that districts sharing the same ethnicity as the president receive twice as much expenditure on roads and have five times the length of paved roads built. Interestingly, this effect disappears during periods of democracy. Alder and Kondo (2019) analyse the effect of political connections on the design of the National Trunk Highway System in China. They uncover that the birthplaces of the top officials in power at the time of the network implementation were closer to the actual network, compared to the optimal network predicted by a quantitative spatial equilibrium model. The paper in the literature that is closer to ours is Selod and Soumahoro (2019), focusing on road construction in Mexico between 1993 and 2012. Their main result is that in municipalities narrowly won by the president's party, the length of the federal highways constructed more than doubled compared to locations lost to the opposition. Similarly to our paper the authors use a regression discontinuity design, but since the Mexican electoral system relies on a combination of single member districts and proportional representation, they cannot precisely define a close race. As a result, they rely on voting outcomes at the municipality level to proxy for that. One advantage of our analysis is that we can instead define precisely a close race at the district level, providing quasi-random variation in the representative elected to Parliament, and then study whether municipalities which sided with a winner were more likely to see a railway cross their territory. Under the assumption that a municipality cannot single handedly affect election outcomes at the district level, this allows us to causally estimate the effect of politics on railway allocation. Importantly,

[4] For a recent survey, see Golden and Min (2013).

taking advantage of the historical nature of our data, we can also investigate the long-run repercussions of the political factors that shaped access to the network. Our paper is thus related also to the recent growing literature on the impact of infrastructure on long-run development. In particular, Hornung (2015) explores the long-run effects of access of Prussian cities to the German railroad network on their population growth rates in 1840-1871. He finds that having a railroad station results in a 1-2% higher city annual growth rate. Berger and Enflo (2017) document a substantial long-run effect of having early access to the railroad network on population growth. They show that cities that gained access to the network in Sweden during its first wave of construction in 1840-1870 are around 62% larger in present times. Büchel and Kyburz (2020) report that gaining early access to the railroad network in Switzerland in the 19th century results in around 0.4% higher regional annual growth rate. A positive long-run impact of early access to transport infrastructure is also documented for developing and less developed countries. Jedwab and Moradi (2016) find persistent positive long-run effects of investing in the railroad network in Africa on city growth. Banerjee, Duflo, and Qian (2020) examine whether in the long run better access to transportation infrastructure in China leads to better regional economic outcomes. Specifically, they find that better proximity to the railroad network in the late 19th and early 20th century causes a moderately higher per capita income level and has not impact on per capita income growth in 1986-2003.

The paper is organised as follows. Section 2 provides a historical overview of railway expansion and elections in Italy in this period. Section 3 describes the data. Section 4 discusses the results linking electoral outcomes and railway development. Section 5 discusses the results analysing the long-run consequences of railway access on economic development. Finally, Section 6 concludes.

2 Historical Background

In this section we provide background information on the development of railroads in Italy in this period, and on the evolution of the Italian electoral system.

2.1 Railway constructions in Italy, 1879-1913

In this paper we focus on the railway expansion overseen by the governments of the "Historical Left" (1876-1896), which dealt mainly with the realization of the secondary lines required to connect smaller towns and cities to pre-existing main lines. Demand for these type of infrastructure had been strong since the 1860s, as the unconnected towns and cities complained that the main lines drew economic activity away from them (Maggi 2003, p.

63). However, construction only began with the advent of the Left and responded to this political group's need to consolidate its power base (see e.g. Tajani 1944 and Mercurio 1994). The key legislative milestone was the so-called "Baccarini Law" of 1879 which set out the plan of railway expansion for the next two decades.[5] This was a controversial initiative – the transcripts of the parliamentary debates around this act exceed 1,500 pages – and over 600 amendments were proposed and defended in many speeches, one of them lasting two days. Crucially, as Schram (1997) points out (page 111) "Clearly,... the members of parliament could not resist the temptation to advance their electoral interests by promoting the construction of a railway line in their own constituency." Based on this plan, 6,794 km of new lines were added between 1879 and 1913, of which 89% were already in place by the time the Left was ousted from power (1896) (see Figure 1 and Appendix Table A.1).

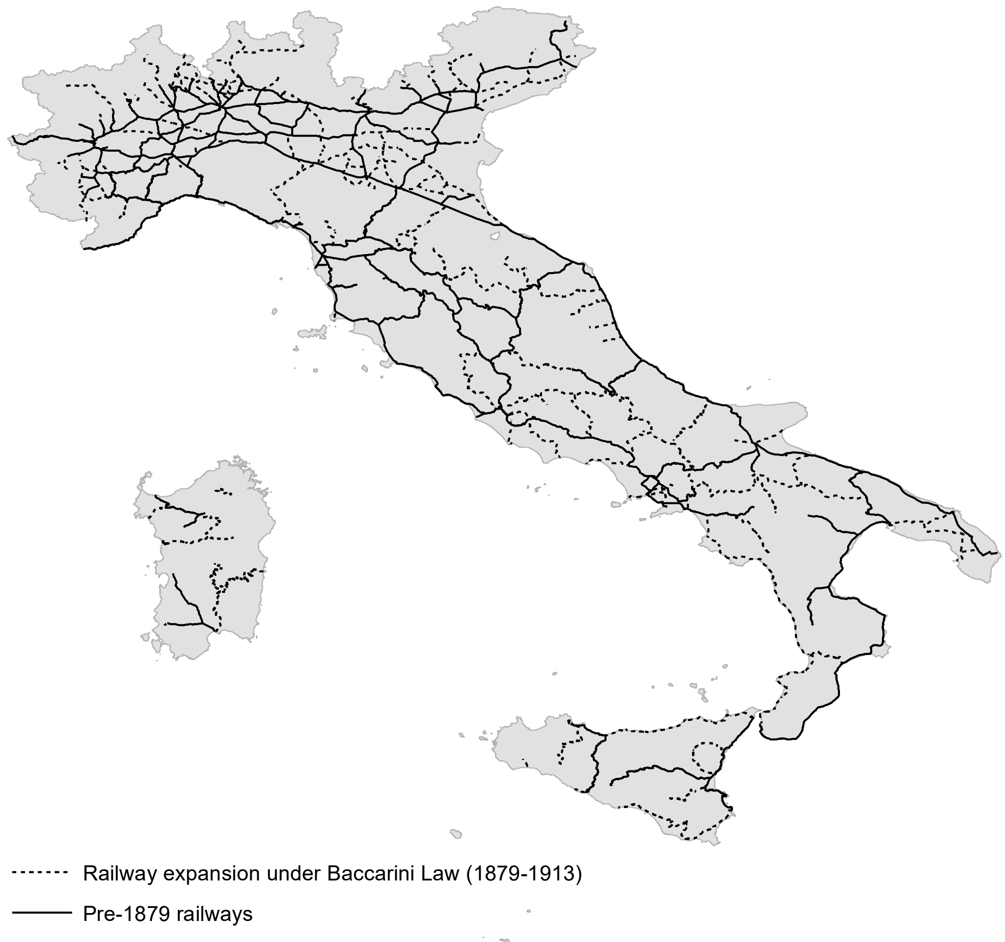
The Baccarini Law set aside a large sum to be spent by the national government on the new railways: 1,260 million lire, or 12.3% of 1879 GDP (Istat 1957), and the actual cost of the network realization turned out to be significantly larger (Ferrucci 1898; Tajani 1944). It envisaged that the building costs should be shouldered also by local governments. The constructions were initially slowed down by financial troubles. There were two reasons behind this poor start. On the one hand, many provinces and municipalities proved unable to contribute to the costs of construction. On the other hand, the cost estimates on which the initial funding assignments relied proved too low, based as they were on very rough technical studies. In many cases, actual construction costs turned out to be more than twice as high as initially expected (Tajani 1944, pp. 94-95). As a result, only 1,491 km of new lines were completed by 1886, or about 22% of the total (see Appendix Table A.1). To speed up the constructions, the government passed new laws in 1885-1888, setting aside additional funding and increasing the role of the private sector.[6] This led to an acceleration in the constructions between 1886-1890. However by the early 1890s, the Italian State was in serious financial troubles, to the point that it could not afford to run the 1891 census. Many blamed the railways for this situation (Ferrucci 1898, p. 7).

In the first part of this paper, we study whether the railway constructions of this period were shaped by pork-barrel politics. It has long been argued that they were. Railways were at the forefront of Italian politics in the 1880s. Members of parliament often centered their campaigns on the promise to bring railways to their constituency (Maggi 2003, p. 67). Contemporary commentators accused the Left of using the railways to help the political and economic fortunes of its own people (Tajani 1944, p. 94-95). Perhaps not surprisingly,

[5] We exclude from our sample a few lines which were built in this period, despite being approved by earlier laws.

[6] The State outsourced 1,000 km of railway constructions (amongst which some category 1-3 lines) to the three private companies which, from 1885 onwards, operated the Italian railways.

Figure 1. Railways Expansion in Italy, Pre- and Post-1879



Notes: Extent of the Italian railway network before 1879 and lines constructed between 1879 and 1913 under the Baccarini Law. Authors' elaborations based on the data by Ciccarelli and Groote (2017).

Fenoaltea (1983) finds that many of the lines built in this period generated little operating revenues.[7]

To investigate these claims, we exploit a peculiar feature of the Baccarini Law. The law grouped the railways into five categories, in decreasing order of national importance and financial contribution by the State.[8] There was an important distinction between category

[7]According to Fenoaltea, the Left's lines did decrease transport costs more than the ones constructed by the Right. This was largely due to the fact that the North-South lines built by the Right, while fulfilling the goal of facilitating defence, were largely unable to compete with sea transport.

[8]The construction of category 1 lines was entirely financed by the State, while category 2-4 lines only attracted a 90%-60% state contribution (the rest being shouldered by the affected provinces and municipalities). Category 5 lines, to which here we add 581 km of Sardinian lines approved by a 1885 law, were entirely financed by the relevant provinces or municipalities, or by private companies. Profits from operating category

1-3 lines, on the one hand, and category 4-5 lines on the other. In the former case, the Baccarini Law determined exactly which lines should be built, by providing lists of city pairs that should serve as start and end points. The law, however, did not describe the exact route to be followed.[9] There were 4,490 km of category 1-3 lines. As for category 4-5 lines, the law merely authorized the government to construct up to 2,530 km of category 4 lines, and to allow the construction of an unspecified amount of category 5 lines.[10] Appendix Figure A.2 draws the two groups of railways in two separate maps.[11]

We focus on category 1-3 lines. There are two different stages at which politics may have affected the allocation of these railways. The first was in the shaping of the Baccarini Law itself. As explained above, the Law provided lists of city pairs that should serve as start and end point of a new line. The politics of the 1876-1879 period may have shaped decisions on which city pairs to include in these lists. We study whether this was the case by analysing whether electoral districts that were aligned with the government in the 1876 election were more likely to be crossed by the straight lines connecting the city pairs listed in the Baccarini Law. The second was the implementation stage. The legislation did not specify the exact route to be followed by the new lines, or the order of construction. Before a line connecting cities A and D could be built, a feasibility study would identify the technical challenges to be overcome, and the exact route to be followed between A and D (e.g. passing through city B as opposed to city C); the State bureaucracy would then have to approve this study and the government would finally contract out the construction of the line. The latter was typically done by trunks: in the example above, the A-B trunk could be contracted out at a different date than the B-D trunk. The governments which were in power after 1879 presumably had a say in the determination of the exact routes to be followed by each line. Thus, for a trunk of a line contracted out after 1879, our hypothesis is that political factors at play during the legislature immediately predating that trunk's realization may have shaped the decision to contract it out, and affected its exact location.

2-4 lines would be shared between the State and the other investors according to their initial contributions, while profits from category 5 lines would be entirely appropriated by non-State investors.

[9] For example, the law would say that a category 1 line connecting the cities of Faenza and Pontassieve should be built, a category 2 line connecting the cities of Bassano and Primolano, and so on. The lists remained essentially unaltered in the following two decades.

[10] The limit for category 4 lines includes an additional 1,000 km that were added by a 1885 law. Maggi (2003), p. 65, argues that demand for category 4 lines largely exceeded supply. According to this author, this category was introduced in order to give members of parliament, whose district would not be connected by category-1-3 lines, the hope (but no firm commitment) that they could still be connected by a category 4 line. The government's assumption was that, without such a provision, the law would never have passed.

[11] All category 1-3 lines would be built "by the State" (though the actual constructions was contracted out to private companies), while the construction of category 4-5 lines could be outsourced to the relevant provinces or municipalities, or to the private sector.

Table 1. Km of Railways Contracted Out and Completed, as Compared to Predicted by Baccarini Law (1879)

	$1^{st} - 3^{rd}$ category	4^{th} category	5^{th} category	Total
Km. contracted out during legislature:				
1876 - 1880	275	0	0	275
1880 - 1882	1,041	531	94	1,666
1882 - 1886	1,166	538	180	1,885
1886 - 1890	1,604	207	882	2,693
Total (1876-1890)	4,086	1,276	1156	6,519
Total km. completed (1876-1913):	4,360	1,276	1,156	6,794
Total km. predicted by Baccarini Law:	4,490	2,530	-	6,620

Notes: The 2530 km of predicted category-4 lines include the 1000 km added to this category by the Legge 27 aprile 1885 n. 3048. Authors' elaborations based on the data by Ciccarelli and Groote (2017).

We investigate this question by asking whether municipalities located close to one of the straight lines connecting the city pairs listed in the Baccarini Law were more likely to receive the railway during a post-1879 legislature, if they had aligned with the government at the start of the legislature. By "receiving" the railway, we mean that construction of a trunk cutting through the municipality was started during the legislature. As start of construction date, we take the date in which the trunk was contracted out. Table 1 reports the cumulative km of lines contracted out during each post-1879 legislature. The focus of our analysis will be on the three legislature spanning the period between 1880 and 1890. We disregard the 1876-80 legislature since very few lines were contracted out during it. We also disregard the post-1890 legislatures, for two reasons. First, the near-bankruptcy of the Italian State meant that the official publications on which we rely become much more patchy after 1890, calling into question the reliability of the data. Second, the vast majority of the Baccarini Law lines were contracted out before 1890.[12]

In summary, we will be studying how election results at the 1876, 1880, 1882 and 1886 elections determined the allocation of the secondary railways of the Baccarini Law. To better understand the politics of this period, we now turn to reviewing the existing electoral rules.

[12]The actual constructions lagged behind, as can be gauged by comparing Table 1 with Appendix Table A.1

2.2 Electoral systems

The Italian Constitution during the Kingdom of Italy envisioned a bicameral system, with a House or Representatives, elected by the population and a Senate whose members were appointed for life by the King. The two chambers had similar powers, and legislation had to clear both of them in order to be enacted by the King. Our focus will be on the House of Representatives whose electoral system evolved quite significantly across the four elections we consider (1876, 1880, 1882 and 1886), in terms of both electoral rules and the extent of the franchise. In the 1876 and 1880 elections, representatives were chosen according to a standard majoritarian rule, with single-member constituencies and runoff voting. There were 508 constituencies and in each of them the candidate who obtained the most votes was elected, as long as he had been supported by at least one third of the eligible voters, and one half of the votes cast (excluding invalid votes). If no candidate satisfied these requirements, then the two candidates with the most votes proceeded to a second round, where the winner was selected by simple majority. Second rounds were frequent under this electoral rule: they occurred in 30% of the electoral districts on average in the 1876 and 1880 elections.

In 1882, a major reform introduced a majoritarian rule with *multi*-member constituencies and runoff voting. Constituencies were made fewer and larger: there were 135 of them, electing a number of representatives varying between 2 and 5.[13] In each n-seat constituency, voters could cast n preferences (at most one per each candidate), with the exception of 5-seat constituencies where voters could only cast 4 preferences.[14] The n candidates with the most votes were elected, provided they had obtained at least as many votes as one eighth of the eligible voters. Where some seats had remained unfilled, the unelected candidates with the most votes (in a number equal to twice the number of seats to be filled) proceeded to a second and final round.

The extent of the franchise was very low throughout the period, but increased significantly with the reform of 1882. Only males could vote. Before the reform, the franchise was strictly based on income: only individuals older than 25 years and declaring an annual income of at least 40 lire were allowed to vote (with a few exceptions made for high-skills professions). As a result, only about 2% of the population was allowed to vote in 1876 and 1880. The 1882 reform reduced the age requirement to 21 years and the income requirement to 19.8 lire, and abolished the income requirement for individuals holding the mandatory elementary school

[13] The law specified the number of seats to be attributed to each province. Such number was proportional to the province's population, and was to be updated in the first parliamentary session after the publication of each population census. The law went on to subdivide provinces into constituencies, in a number varying between 1 and 5.

[14] The rationale for 5-seat constituencies was to make it easier for important minority group to elect at least one representative (see the discussion in Brunialti 1882, pp. 250 onwards).

degree (2 years of schooling), as well as for an extended number of high-skills professions. As a result, the franchise rose to more than 7% of the population (see Appendix Table A.2).

3 Data

We assembled a novel dataset covering approximately 7,000 Italian municipalities, including data on railway development, electoral outcomes and a wealth of socio-economic characteristics observed at various points in time between 1863 and 1991. One issue that we face is that municipalities did not remain identically defined over this period. For example, some of them merged to form larger municipalities, others split to form smaller ones, some acquired territory from others, and there were denomination changes and transfer of municipalities across larger administrative units (e.g. provinces, regions).

To address this issue, we used information from *Sistema Informativo Storico delle Amministrazioni Territoriali* (SISTAT) to adjust the territories of the municipalities to be at constant 1991 borders. SISTAT is a database maintained by the Italian National Institute of Statistics (ISTAT) reporting any territorial or administrative changes that municipalities have gone through since 1861. Its data allows us to keep track of denomination changes, and any territorial changes involving municipalities as a whole. The most frequent case is mergers, e.g. municipalities A and B merging to form municipality C at time $t < 1991$. In this case, when collecting historical data for time $t_1 < t$, we separately collected data for A and B, and then aggregated up so that only C appears in our dataset at time t_1. In a much smaller number of instances, municipalities split, e.g. C splitting into A and B at time $t < 1991$.[15] In this case, for any $t_1 < t$ for which we have access to data, we first collect the observation referring to C, and we then use population weights in 1991 to split this observation into A and B, so that these municipalities also exist in our dataset in $t_1 < t$. This procedure does not allow us to track territorial changes at the level below the municipality. For example, if municipality A incorporated 10% of the former territory of municipality B, then we can not take care of this, since our minimal unit of observation is the municipality. This kind of changes in territory were minor and infrequent, however.

[15] Municipalities were mostly aggregated during this period, as evidenced by the fact that there were 8,400 municipalities in 1871, but only 7,700 in 1991, at constant 1871 national territory. The decades immediately after World War II were exceptional in this respect, since many mergers that had occurred during the Fascist era were reversed afterwards. For this period, we only use population data which was independently adjusted to be at constant 1991 municipal boundaries by Sistema Statistico Nazionale (1994).

3.1 Railways

Our analysis focuses on the railway expansion plan laid out by Baccarini Law of 1879. The vast majority of these railways were contracted out for construction in 1880-1890, and opened by 1895 (see Table 1 and Appendix Table A.1). Our starting point is the database by Ciccarelli and Groote (2017), which is based on historical sources and maps and provides a georeferenced reconstruction of the development of the Italian railway network spanning the period 1839-1913. In this GIS dataset, the authors digitise railway segments as line features, and provide additional information such as year of opening, main/secondary line classification and a distinction between standard and narrow gauges. Since our municipalities are defined at constant 1991 boundaries, we combined this database with the 1991 shapefile of Italian municipalities, which is also the earliest provided by ISTAT.[16] This procedure allowed us to determine the list of municipalities that each new railway trunk cut through.

We begin by constructing measures of the stock of railways (if any at all) present on each municipality's territory in 1876, the date of the first election in our sample. We next identified the Baccarini Law lines in our GIS dataset, and proceeded to determine the date in which the construction of each trunk was started, using information from Ministero dei Lavori Pubblici (1885), Ministero dei Lavori Pubblici (1889) and Ministero dei Lavori Pubblici (1891).[17,18] See Appendix Figure A.3 for an example of the information available in the construction reports we have employed.

3.2 Electoral outcomes

Data on parliamentary election outcomes have been obtained from Corbetta and Piretti (2009). This source provides digitised data on Italian parliamentary elections from 1861 to modern times, at the district and municipality level. Our elections of interest are 1876, 1880, 1882 and 1886.

The district-level data contains the name of the district, the number of registered voters, the number of voters who cast a ballot and the name and political affiliation of the elected candidate(s). For the 1876 and 1880 elections, in which representatives were chosen in single-member districts, we also have information on the number of votes obtained in the

[16] We exclude from the analysis the North-Eastern parts of the country that were only annexed after World War I

[17] We decided not to use the completion date by Ciccarelli and Groote (2017) for this purpose, since completion could lag the start of construction by up to several years.

[18] For category 1-3 lines and some category 4 lines, the State was directly involved in signing the construction contracts. However for the remaining category 4 lines, and for all category 5 lines, the State signed concessions to the interested provinces or municipalities, or to private firms. This meant that it was the concessionaires to sign the construction contracts. The concessions were signed after the State bureaucracy had approved the technical study for the line, and before the construction contracts were signed.

district by the winning candidate. For the 1882 and 1886 elections, in which a variable number of representatives (between 2 and 5) were elected in districts at large, we have information on the number of preferences obtained by each elected candidate. Since Corbetta and Piretti (2009) do not provide data on the number of votes or preferences obtained by the losing candidates in the district, we obtained this information digitising data from Nuvoloni (1898). These sources allowed us to collect the same information for run-off elections, which took place on average in 30% of the districts in the 1876 and 1880 elections, and in 2% of the districts in the 1882 and 1886 elections. Appendix Table A.2 provides summary statistics on district-level population, franchise and turnout in the four elections in our sample.

Corbetta and Piretti (2009) also provide municipality-level electoral outcomes.[19] For the 1876 and 1880 elections, for all municipalities located in a district, they report the votes obtained by all candidates in the district. This gives us the total number of votes cast in the municipality, and hence the share of votes obtained by candidates who were elected in the district as well as those who were not. For the 1882 and 1886 elections, Corbetta and Piretti (2009)'s municipality-level data only reports the number of preferences obtained by the candidates who were elected in the district. To complement this data, we collect and digitise information on registered voters and number of voters who cast a ballot at the municipality level for the 1882 and 1886 elections from *Archivio della Camera Regia*. Based on the number of voters who cast a ballot and the number of expressible preferences in the district to which the municipality belongs, we estimate the share of preferences obtained in the municipality by each elected candidate.[20]

3.3 Additional socio-economic characteristics

The existing literature has highlighted the role played by a number of additional socio–economic factors in determining whether a locality is connected by a railroad. To account for them, we have assembled a rich set of municipality-level characteristics from a variety of historical sources.

We obtain population figures spanning the period 1861–1991 by digitizing data from Sistema Statistico Nazionale (1994).[21] Importantly, this data refer to municipalities boundaries as

[19]More precisely, the electoral outcomes for municipalities reported by Corbetta and Piretti (2009) are based on data at the level of *sezione*, an electoral division whose relation to the municipality depended on the latter's size and on the extent of the franchise. For small municipalities in the early part of our sample a "sezione" would often encompass multiple municipalities, in which case we assign to each of municipality the electoral result of the "sezione" they belong to.

[20]Specifically, we compute the share of preferences obtained by each elected candidate as the ratio between the number of preferences cast in favour of the candidate in the municipality and the number of total number of voters who cast a ballot in the municipality multiplied by the number of expressible preferences.

[21]Censuses were held every ten years except for 1891 and 1941, when they were not held, and 1936, when

defined in 1991 and thus matches the information we have on the presence of railroads and political outcomes. We additionally employ this population data to construct a municipality-level measure of market access.[22]

We have then digitized information on other indicators of initial economic development, e.g. the presence of a post office, telegraph office, railway station or sea port in 1871 (from Ministero dell'Interno 1874), and the number of secondary schools and libraries of municipalities in 1863 (from Ministero dell'Educazione Nazionale 1866 and Ministero dell'Educazione Nazionale 1865). For our long-run analysis we have also collected information on a series of additional municipality-level characteristics, which have been digitized from a series of inquires carried out in the 1880's. In particular, data on hygienic and sanitary conditions in 1885 is provided by Direzione Generale della Statistica (1886). We observe quality and quantity of water available, percentage of roads with sewage system, percentage of houses with toilets, number of doctors, number of pharmacies, number of hospital beds and number of years with registered cases of cholera. We have supplemented this information with data on public revenues and expenditures in 1884 obtained by digitizing reported in Ministero di Agricoltura, Industria e Commercio (1887). In particular, we collected total municipality revenues as well as revenues obtained by taxing terrains and buildings – to proxy for overall extent of economic activity – and expenditures on public education, to capture local investment in human capital acquisition.

We additionally collected information on a series of time-invariant geographic characteristics. Using the FAO-GAEZ database, we have constructed ten separate municipality-level indexes on suitability for growing agricultural crops.[23] To account for the role played by difficult terrains, particularly in obstructing railway passage, we used data on terrain ruggedness from Nunn and Puga (2012) to construct a municipality-level Terrain Ruggedness Index (TRI). Finally, we compute measures of land area, elevation and an indicator for municipalities situated on the sea coast.

Table 2 and Appendix Table A.3 provide summary statistics on the various municipality-level characteristics we collected.

---

an additional census was held.

[22]We follow the standard approach in the literature to obtain our measure of market access, where for each municipality i we compute $MA_i = \sum_{i \neq j} P_j D_{ij}^{-1}$, with P being the population of municipality j, and D the geodesic distance between municipalities i and j. In the calculation of MA_i we employ the population of all municipalities, with the exception of municipality i's own population.

[23]In particular, we have information on barley, bean, cereals, citrus, cotton, oat, olive, rice, rye and wheat – with suitability varying between 0 (lowest) and 1 (highest).

TABLE 2. SUMMARY STATISTICS ON MUNICIPALITY-LEVEL CHARACTERISTICS

	Mean	*S.D.*	*Min*	*Max*
Pre-1879 and geographic characteristics				
Log population density in 1871	4.5751	0.8106	0.6530	9.9889
Post office in 1871	0.3551	0.4786	0	1
Telegraph office in 1871	0.1491	0.3562	0	1
Railway station in 1871	0.0922	0.2893	0	1
Sea port in 1871	0.0059	0.0765	0	1
Number of secondary schools and libraries in 1863	0.1030	0.6234	0	22
Market access	0.3757	0.1326	0	1
Wheat suitability	0.3594	0.2081	0	1
Cereals suitability	0.2943	0.2528	0	1
Rice suitability	0.0707	0.1805	0	1
Cotton suitability	0.1755	0.2162	0	1
Barley suitability	0.3634	0.2089	0	1
Rye suitability	0.3713	0.1989	0	1
Olive suitability	0.3047	0.1899	0	1
Citrus suitability	0.1751	0.2521	0	1
Oat suitability	0.3605	0.2089	0	1
Bean suitability	0.3397	0.1926	0	1
Terrain ruggedness	0.1823	0.1799	0	1
Log land area	3.0807	1.0064	-2.1295	6.3868
Elevation (m)	429	422	0	2,699
Coast	0.0806	0.2723	0	1
Additional characteristics (post-1879)				
Quantity of water	2.6924	0.5960	1	3
Quality of water	2.7811	0.7765	1	4
% of roads with sewage	0.1292	0.2693	0	1
% of houses with toilets	0.4452	0.2952	0	1
Number of farmacies (per capita)	0.0003	0.0004	0	0.0045
Number of medics (per capita)	0.0005	0.0005	0	0.0112
Number of years with colera epidemics	1.4847	1.5089	0	11
Number of hospital beds (per capita)	0.0008	0.0063	0	0.4491
Revenue tax on terrains (log per capita)	1.4675	0.5017	0	4.0726
Revenue tax on buildings (log per capita)	0.4847	0.2911	0	1.9581
Municipal surtax (log per capita)	1.4492	0.6224	0	4.1623
Total tax revenues (log per capita)	2.2397	0.4235	0.4686	5.2679
Education ordinary expenses (log per capita)	0.7649	0.2013	0	3.0933
Education extraordinary expenses (log per capita)	0.0903	0.3189	0	4.0723
Education optional expenses (log per capita)	0.1161	0.2074	0	2.4050

Notes: Authors' elaborations. Data sources described in Section 3.

4 Elections and Railway Development

In this section we study the role of electoral competition in shaping the development of the Baccarini Law lines. We start by analysing the decision to include a line in the bill, and then turn to explain the actual path followed in its construction.

4.1 Identification

The decision to build the railroads considered in our analysis can be articulated in two steps. The first involves the choice of including a railway line in the Baccarini Law. As already mentioned, the law specified the initial and end (focal) points, but no additional detail was provided on the route to be followed. The second step involves instead the choice of the path of the various trunks which were subsequently built, and whose construction was contracted out between 1880-1890. We are interested in understanding the role of politics in shaping both these decisions.

In both steps we focus on close races, i.e. elections where the choice of one candidate over another can be thought of as being as good as random. One important caveat applies though. As pointed out in Section 2, while in the 1876 and 1880 elections representatives were chosen in single member districts (SMD), by 1882 a new electoral law prescribed the creation of districts at large (AL) in which a variable number of representatives (between 2 and 5) were elected. In the SMD system the definition of a close race will be based on the gap between the vote shares received by the winner and the runner up, whereas in the AL system, we will focus on the (normalized) gap in the vote share received by the last of the elected candidates and the first of the non-elected ones.[24]

In the first step of the process, representatives of different districts compete for the allocation of a railway line to their constituency, and if alignment matters, we expect those supporting the existing government to be in a better position to exert influence. To assess whether this is the case, we focus on electoral outcomes at the district level and implement a sharp regression discontinuity design where we compare districts in which the government candidate marginally won to those in which he marginally lost. More formally, we estimate the following specification:

$$y_{md} = \beta_0 + \beta_1 GovWin_d + f(Margin_d) + \beta_2 X_{md} + e_{md} \tag{1}$$

where m denotes the municipality and d the electoral district. y_{md} is the outcome of interest, which is defined in two alternative ways. First, we consider the log of the distance in km of the municipality's centroid from the straight line connecting the focal points of the planned railway; second, we define two indicator variables, taking a value of one if the municipality falls respectively within a distance of 10 km or 5 km from the planned line.[25] $GovWin_d$ is

[24] In particular, let W_d be the vote share received by the least popular of the elected candidates, and let L_d be the vote share received by the most popular of the non–elected candidates. The normalized gap will then be given by $\frac{W_d - L_d}{W_d + L_d}$.

[25] See Appendix Figure A.4 for an illustration of the straight-line railway plans.

an indicator taking a value of one if the government carried the district, and zero otherwise. $f(Margin_d)$ is the RD polynomial, which allows us to flexibly control for the impact of the distance between the share of the government party and the 50 percent threshold. Finally, X_{md} is a vector of pre–determined municipality level characteristics. Our coefficient of interest is β_1, capturing the effect of a government's party victory at the discontinuity point (i.e. the 50 percent threshold) on the relevant outcome variable.

Turning now to the actual realization of the railroads, our hypothesis is that once a line has been authorized by the Baccarini Law, different municipalities compete to shape the actual path that will be followed. We thus focus on municipalities located in close proximity to the authorized railway lines (i.e. at a distance of either 10 km or 5 km from the straight lines connecting start to end) and consider the role played by local – i.e. municipality level – political outcomes. We posit that municipalities which are aligned with candidates that ended up being elected are in a better position to see railroad lines constructed on their territory. To identify this effect we focus once again on close races, i.e. districts in which the elected candidate won by a small margin. Within these districts, we examine whether the railway was more likely to be assigned to municipalities that gave a large share of their votes to the winning candidate, as compared to those who did not. Because the district is marginal, alignment with the winning candidate is as good as random. Our identifying assumption is that municipality electoral outcomes – on average – did not systematically affect district-level ones, and as a result, a municipality might well have overwhelmingly supported a candidate, who in turn was not elected. We thus estimate the following model:

$$Rail_{md} = \gamma_0 + \gamma_1 WinShare_{cd} + \gamma_2 X_{md} + e_{md} \tag{2}$$

where $Rail_{md}$ is an indicator variable taking a value of one if a railroad whose construction was contracted out during the legislature crosses the territory of the municipality and zero otherwise; $WinShare_{cd}$ is the share of votes received in municipality m by the candidate who carried district d and X_{md} is a vector of pre–determined municipality level characteristics. Our coefficient of interest is γ_1 and captures the effect of greater support in the municipality for the candidate who marginally won the election as opposed to those who lost.

4.2 Explaining which lines will be built

The first step of our analysis consists in determining whether the inclusion of a railway line in the Baccarini Law of 1879 was driven by the lobbying of members of parliament on behalf of their own constituents. In particular, we focus on district-level electoral outcomes for the 1876 election, which determined the composition of the parliament that passed the aforementioned law in 1879.

Our estimates are based on equation (1), where we regress a series of municipality-level outcomes – capturing the geographic proximity to railways planned by the Baccarini Law – on an indicator for whether the government carried the district to which the municipality belongs. Through this part of the analysis, we adopt a sharp regression discontinuity design where we study our outcomes of interest at the 50% vote share threshold, comparing municipalities belonging to district where the government party marginally lost to those where the government party marginally won. To account for potential correlation in the estimated relationship across nearby municipalities, we employ spatially-adjusted standard errors, following the methodology of Conley (1999), based on a window of 20 km around the municipality's centroid.

Table 3 reports the results from the RD estimates. In odd-numbered columns we employ the full sample of almost 7,000 municipalities across 458 electoral districts.[26] These estimates include a third-order RD polynomial on the distance of the government party vote share from the 50% threshold, fitted separately on each side of the threshold. Furthermore, the estimates control, in addition to province fixed effects, for a rich set of municipality-level pre-determined and geographic characteristics, including log population density, presence of post office, telegraph, port, railways or railway station on the territory, total number of secondary schools and libraries, market access, terrain ruggedness, agricultural suitability, log land area, elevation and location on the coast.[27]

Across all three dependent variables, the estimates present a clear discontinuity at the 50% threshold. In particular, column (1) shows a negative discontinuity in the distance of the municipality from the planned railways (represented as a straight line connecting the destination points indicated on the Baccarini Law) in districts secured by the government party. Similarly, columns (3) and (5) show a positive discontinuity in the likelihood of a municipality being in a 10 or 5 km proximity, respectively, to the planned railways, when comparing municipalities in districts where the government party won with ones in districts where the government party lost.

While the results reported in odd-numbered columns of Table 3 are helpful in providing a general overview of the patterns in the data, the estimated effects might not be properly identified. This is because employing the whole sample of municipalities, even those in

[26]The number of 458 electoral districts is lower than the total of 508 present in the 1876 election due to the group of very large municipalities incorporating multiple electoral districts. In this case, the electoral district is therefore smaller than the municipality, our unit of analysis. Furthermore, these very large municipalities were already reached by the main railway network by 1879. For the above reasons, we exclude these municipalities and their electoral districts.

[27]While in Table 3 we only show the coefficient on the main explanatory variable for reasons of space, in Appendix Table A.4 we also report the coefficients on the control variables.

Table 3. Baccarini Law's Railway Projects and 1876 District-Level Elections: Regression Discontinuity Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	*log(km) dist. from projects*		*10 km proximity to projects*		*5 km proximity to projects*	
RD polynomial order:	3rd	2nd	3rd	2nd	3rd	2nd
Margin:	-	± 5%	-	± 5%	-	± 5%
Government-party win in district	-0.3081***	-0.3587**	0.1892**	0.2494**	0.1898***	0.3508***
	(0.1180)	(0.1722)	(0.0789)	(0.1154)	(0.0659)	(0.0952)
Adjusted R^2	0.515	0.733	0.260	0.528	0.196	0.370
Observations	6,957	1,471	6,957	1,471	6,957	1,471
Number of electoral district	458	102	458	102	458	102
Municipality-level controls	✓	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓

Notes: The table reports reports RD estimates. The unit of observation is a municipality. The dependent variable is the distance in log(km) from the railway projects in columns (1)-(2); a binary indicator for being in a 10 km proximity to the railway projects in columns (3)-(4); a binary indicator for being in a 5 km proximity to the railway projects in columns (5)-(6). The distance and proximity measures are based on straight-line overall projects, connecting the focal points of the line as indicated on the Baccarini Law. The explanatory variable indicates a government-party win in the electoral district, with 3rd-order polynomials on the margin of victory being employed in odd-numbered columns and 2nd order polynomials in even-numbered columns. In even-numbered columns, the sample is limited to electoral districts where the government-party candidate's vote share was within a 5% distance of the 50% threshold needed for election. Municipality-level controls include: log population density in 1871, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation and presence of a coast. All specifications employ province fixed effects. Standard errors adjusted for spatial correlation, based on a 20 km window, in parenthesis. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

electoral districts where the winning candidate's vote share is far above the 50% threshold, might cause our estimates to capture some unobservable variation in the data, potentially biasing the results. For this reason, we repeat our analysis focusing on municipalities belonging to districts where the election was determined by a close margin. To determine the optimal bandwith we follow the selection procedure proposed by Calonico, Cattaneo, and Titiunik (2014), yielding margins between 5% and 10%. While these might appear "large", it is worth noting that the restricted extent of the franchise in 1876 resulted in only 800 votes cast in the average district in that election, implying that a 5% margin involved on average a distance of only 40 votes between the candidate and the election threshold. Following the most conservative approach (e.g. 5% margin) reduces the sample size to 1,471 municipalities across 102 electoral districts. Our results are reported in the even-numbered columns of Table 3 and include municipality-level controls, province fixed effects and a second-order polynomial on the margin of victory. As it can be observed, the district's alignment with the government has a statistically significant and economically meaningful effect across all three dependent variables, namely a lower distance of the municipality from a planned railway in columns (2) and a higher likelihood of the municipality being in a 10 or 5 km proximity to a planned railway in columns (4) and (6), respectively. To give an idea of the size of the estimated effect, the results in column (6) imply a 35% higher likelihood of the municipality being in a 5 km proximity to a railway planned by the Baccarini Law if the district elected a representative aligned with the government rather than one belonging to the opposition.

4.2.1 Validity of the RD design

To establish the validity of our RD design, we carry out a number of tests and checks. First, a required condition is the absence of selective sorting around the threshold. In the context of our study, this condition would be violated if candidates were able to manipulate the margin of victory, with a resulting discontinuity at the 50% vote share cutoff. To evaluate the presence of sorting around the threshold, we implement a McCrary test (McCrary, 2008), collapsing the data into government party vote share bins and plotting the density of observations in each bin. Appendix Figure A.5 shows that there is no discontinuous change in the number of observations in each bin around the 50% vote share threshold, suggesting that government party candidates were not able to systematically manipulate their vote share to achieve victory.

Another RD identification assumption that needs to be satisfied is that all other relevant features aside from the treatment do not systematically differ at the 50% vote share threshold. This assumption ensures that observations just below the threshold serve as an appropriate counterfactual for those just above it. To assess the plausibility of this assumption, we examine whether socio-economic and geographic characteristics of municipalities are balanced across electoral districts that were marginally won or lost by government party candidates. We report the results of this balancing test in Appendix Table A.5, where we compute the mean value of our various municipality-level characteristics across the two groups of districts. In particular, columns (1) and (2) report mean values for districts marginally won or lost, respectively, by a government party candidate with a vote share margin below 5%. Finally, column (3) reports the difference across the two groups and results from a t-test on the equality of means. As one can establish, there are no systematic differences of the covariates around the 50% vote share threshold.

An additional exercise is to examine the sensitivity of results to alternative bandwidth choices. In Appendix Figure A.6 we repeat the RD regressions focusing on marginal electoral districts, reported in columns (2)-(4)-(6) of Table 3, while employing different vote share margins to restrict the sample. The plotted coefficients show a consistent estimated effect of a government-party win across all three dependent variable while employing different bandwidths. Moreover, in line with the margins suggested by the optimal bandwidth selection procedure proposed by Calonico, Cattaneo, and Titiunik (2014), we can observe a disappearing effect once the margins increase above 10%, and unstable estimates, due to the resulting greatly reduced sample size, once the margins fall below 5%.

4.2.2 Robustness

To further establish the resilience of our regression discontinuity results, we conduct a series of robustness checks. First, we repeat the estimates of equation (1) while employing the full set of municipality-level controls, including the additional set of contemporary socio-economic characteristics consisting of measures on hygienic and sanitary conditions as well as public revenues and expenditures. The results are presented in Appendix Table A.6 and are very similar to those reported in Table 3.

Next, we turn our attention to the treatment of standard errors. While our baseline approach has been to employ spatially-adjusted standard errors based on a window of 20 km around the municipality centroid, other types of correlation in observed and unobserved characteristics may potentially affect our results. Fort this reason in Appendix Table A.7, we assess the robustness of our findings to alternative clustering structures, namely clustering by electoral district, double-clustering by electoral district and railway line and spatially-clustering based on a bigger window of 50 km around the municipality centroid. All these different methodologies yield comparable standard errors, which do not impact the significance of our results.

4.3 Explaining the path to be followed by each line

The next step in our analysis studies the implementation of the railway expansion set out by the Baccarini Law. In other words, after studying whether politics influenced the inclusion of railway lines in this piece of legislation, we turn to the analysis of pork-barrel in shaping the actual path followed by the realized railway lines.

4.3.1 Single member districts

We start with the second legislature in our timeline, which was elected in 1880 under a majoritarian system with single-member districts. In the following set of regressions, our dependent variable indicates whether a railway – contracted out for construction during the 1880 legislature – crossed the municipality's territory.

Studying the political influence on the construction path of the railway expansion predetermined by the Baccarini Law begs the question of what level of electoral outcomes were crucial in this context. When studying the inclusion of a railway line in the planned expansion it is natural to focus on electoral results at the district level, given that the planned railway lines could span long distances and were therefore likely to be affected by electoral results only at an aggregate level. On the other hand, when it comes to the implementation of the plans, i.e. the construction of the railways, it is not immediately obvious whether this kind of outcome can be affected by electoral results at the district level or only at local level.

In Appendix Table A.8 we study the implementation of the plans by first associating it with district-level electoral outcomes. Like before, we focus on "close" districts, where the elected candidate won by a margin within 5% of the vote share. Additionally, knowing the location of the planned railways, we further restrict our analysis on the sample of municipalities that are in close proximity to the plans, i.e. that lie within 10 km (columns (1)-(2)) and 5 km (columns (3)-(4)) from them. All specifications include our set of municipality-level pre-determined and geographic controls, province fixed effects and a polynomial on the margin of victory. All reported coefficients show a statistically insignificant effect for a government-party win in the electoral district, implying that electoral outcomes at this level did not play a relevant role when it comes to the implementation of the railway plans.

In Table 4 we turn instead our attention to electoral results at the municipality level. These estimates, based on equation (2), employ the share of votes obtained in the municipality by the candidate elected in the district. In particular, we study whether *within a district* municipalities aligning with the elected candidate saw a new railway crossing their territory. Following our earlier approach, the analysis restricts the sample to municipalities in a 10 or 5 km proximity to the railway plans, and focus on districts where the elected candidate won by a vote share margin below 5%. All estimates come from regressions employing the set of municipality-level pre-determined and geographic controls, together with province fixed effects. Additionally, we employ district fixed effects, allowing us to compare only municipalities belonging to the same electoral district and employ spatially-adjusted standard errors based on a window of 20 km around the municipality's centroid. The results reported in Panel A indicate a positive effect for the elected candidate's vote share in the municipality on the probability of the municipality obtaining a railway. In particular, results in column (1) indicate that a standard deviation increase in the share of votes received by elected candidate in the municipality is associated with a 2.9% increase in the likelihood of obtaining a railway. On the other hand, results in column (2), while positive, are not statistically significant.

The findings in Panel A do not consider the political alignment of the elected candidate, concealing a potentially important source of heterogeneity in the estimates. In fact, an elected candidate's ability to reward municipalities for their support may crucially depend on his relationship with the government. To study this question we decompose the effect of the share of votes obtained by the winner into two separate effects, distinguishing between candidates affiliated with the government and those affiliated with the opposition. The results

Table 4. Implementation of Baccarini Law's Railway Projects and 1880 Municipality-Level Electoral Outcomes

	(1)	(2)
Dependent variable:	*Municipality obtains rail during legislature*	
Proximity to projects:	10 km	5 km
Margin of victory:	$\leq$ 5%	$\leq$ 5%
	Panel A	
Winner share in municipality	0.0292**	0.0255
	(0.0138)	(0.0175)
Adjusted R^2	0.309	0.354
	Panel B	
Winner share in municipality	-0.0179	-0.0505
$\times$ Opposition-party win in district	(0.0241)	(0.0336)
Winner share in municipality	0.0470***	0.0486**
$\times$ Government-party win in district	(0.0157)	(0.0192)
Adjusted R^2	0.313	0.362
Observations	619	420
Number of electoral district	75	61
Municipality-level controls	✓	✓
Province FE	✓	✓
District FE	✓	✓

Notes: The table reports reports OLS estimates. The unit of observation is a municipality. The dependent variable indicates the passage of a railway on the municipality territory, which construction was contracted out during the 1880 legislature. In Panel A the explanatory variable indicates the share of votes obtained in the municipality by the candidate elected in the electoral district. In Panel B the share of votes obtained in the municipality is interacted with the party of the elected candidate. Column (1) employs the sample of municipalities in a 10 km proximity to the railway plans, while column (2) employs the sample of municipalities a 5 km proximity to the railway plans. The sample is restricted to municipalities belonging to districts where the candidate was elected with a margin of victory below 5%. Municipality-level controls include: log population density in 1871, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation and presence of a coast. All specifications employ province and district fixed effects. Standard errors adjusted for spatial correlation, based on a 20 km window, in parenthesis. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

in Panel B reveal a positive and significant effect for aligning with the winning candidate if he belongs to the government, while a negative and not significant effect if he belongs instead to the opposition.[28] Specifically, our results indicate that a standard deviation increase in the share of votes given in the municipality to an elected candidate supporting the government is associated with an increase in the likelihood of the municipality obtaining a railway between 4.7% and 4.9%. The opposing sign of the two coefficients distinguishing based on the party of the elected candidate also provides an indication as to the reason why the estimated effect is not significant in column (2) of Panel A, when a single variable is employed.[29]

[28]For the coefficients on the control variables, see Appendix Table A.9.

[29]These results also help dispel a potential endogeneity concern. It could be that municipalities whose economy was, for whatever reason, doing well, were both supportive of the government, and particularly

4.3.2 Districts at large

We continue our analysis by studying the elections of 1882 and 1886. By 1882 a new electoral system introduced districts at large in which a number of representatives, varying between 2 and 5, were elected. This requires a change in our definition of marginal districts to the normalized gap in the vote share received by the last of the elected candidates and the first of the non elected ones.[30] For consistency and ease of comparability with our previous analysis under the single-member district system, our exercise of restricting the sample of municipalities to those in marginal districts will be based on the top quintile of our normalised marginality measure, closely corresponding to the 5% victory margin employed with the single-member constituencies electoral system.

Estimates are still based on equation (2), with the dependent variable represented by an indicator variable capturing whether the municipality receives a railway regressed on the vote share obtained in the municipality by the last of the candidates elected in the district. Controls include the set of municipality-level pre-determined and geographic controls previously deployed, together with province and electoral district fixed effects.

Table 5 reports estimation results employing electoral outcomes from the 1882 and 1886 elections. In Panel A we initially focus on the effect of the vote share obtained in the municipality by the marginal winner. The estimated coefficient reports that a standard deviation increase in the vote share obtained by the last elected candidate is associated with an increase in the likelihood of the municipality obtaining a railway varying between 4% and 3.5%.

Similar to our previous set of results focusing on the 1880 elections, in Panel B we decompose the impact of the vote share obtained in the municipality by the marginal winner based on whether the elected candidate is affiliated with the government or the opposition. Despite a different electoral system and our focus on the last of the elected candidates, the results paint a picture consistent with that of the 1880 elections, with an effect that is positive and significant for municipalities supporting a winning candidate affiliated with the government, while an effect that is not statistically significant for support for municipalities supporting winning candidates affiliated with the opposition.[31] Specifically, estimation results in Panel B indicate that a standard deviation increase in the share of votes given in the municipality to a winning candidate affiliated with the government is associated with an increase in

likely to receive a railway line. If this was driving our results, however, we would expect alignment with the government to increase the likelihood of receiving the railway, independently on whether the government won in the district. In contrast, we see that alignment with government only mattered in districts in which the government actually won.

[30] For more details see the discussion in footnote 24.

[31] For the coefficients on the control variables, see Appendix Table A.10.

the likelihood of the municipality obtaining a railway between 3.8% and 3.5%, an effect comparable to that uncovered for the 1880 elections in Table 4.

Table 5. Implementation of Baccarini Law's Railway Projects and 1882-1886 Municipality-Level Electoral Outcomes

	(1)	(2)
Dependent variable:	*Municipality obtains rail during legislature*	
Proximity to projects:	10 km	5 km
Marginality percentile:	≤ 20th	≤ 20th
	Panel A	
Last winner's share in municipality	0.0395***	0.0349**
	(0.0116)	(0.0157)
Adjusted R^2	0.368	0.387
	Panel B	
Last winner's share in municipality	0.0494	0.0371
× Opposition party	(0.0318)	(0.0417)
Last winner's share in municipality	0.0378***	0.0346**
× Government party	(0.0121)	(0.0163)
Adjusted R^2	0.368	0.386
Observations	1,052	671
Number of electoral district	49	46
Municipality-level controls	✓	✓
Province FE	✓	✓
District FE	✓	✓

Notes: The table reports reports OLS estimates. The unit of observation is a municipality. The dependent variable indicates the passage of a railway on the municipality territory, which construction was contracted out during the 1882 and 1886 legislatures. Panel A employs the share of total votes obtained in the municipality by the candidate elected in the last seat of the district. In Panel B the share of votes obtained in the municipality is interacted with the party of the elected candidate. Column (1) employs the sample of municipalities in a 10 km proximity to the railway plans, while column (2) employs the sample of municipalities a 5 km proximity to the railway plans. The sample is restricted to municipalities in districts belonging to the top quintile of the marginality measure. Municipality-level controls include: log population density in 1881, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation and presence of a coast. All specifications employ province and district fixed effects. Standard errors adjusted for spatial correlation, based on a 20 km window, in parenthesis. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

4.3.3 Threats to identification and robustness

We now explore the robustness of our results and consider potential threats to identification while focusing on the implementation of the Baccarini Law railway expansion.

One of our key identifying assumptions is that individual municipalities' electoral outcomes do not systematically affect district-level ones. This might not be true for particularly large municipalities, which could represent a considerable share of the total votes in the electoral district. To address this concern, we estimate equation (2) excluding the largest municipalities, i.e. those in the top 5% of the distribution by population. Results are reported in Appendix

Tables A.11 (focusing on the 1880 election) and A.12 (1882 and 1886 elections). The findings indicate that large municipalities are not driving our previous results.

Our analysis so far has focused on whether alignment with the district's elected member of parliament made a municipality more likely to see a Baccarini Law line crossing its territory. We have instead abstracted away from explicitly considering the timing of the construction, which may also be affected by political factors. In other words, while in first instance elected representatives compete to bring a rail line to a given municipality, they may also compete to get the project implemented sooner rather than later. To address this concern we re-estimate equation (2) using railway line fixed effects instead of province fixed effects, noting that trunks of the same line were approved for construction within similar time frames. The results from this alternative specification are presented in Appendix Tables A.13 (1880 election) and A.14 (1882 and 1886 elections) and are very similar to our benchmark findings.

In our main empirical model we have considered the alignment between a municipality and the candidate elected in the local district. With experienced politicians being in a better position to both exert their influence and secure votes, not accounting for incumbency might bias our estimates. To address this concern, we use data from Corbetta and Piretti (2009) to match the names of candidates across different elections to identify incumbents. Starting with the 1880 election we find that while 283 out of 508 representatives were re-elected in the same district, only for 15 of them this occurred in close races. Turning to the 1882 election, the electoral reform completely changed the design of districts and introduced multi-member constituencies, making it hard to exploit political rents from a previous election. Finally, considering the 1886 election there were 26 candidates re-elected in the last seat of the district, but only for 6 of them the election was close. Overall, these numbers indicate that incumbents were much more likely to be re-elected with a large margin of victory than in close races, and as a result they tend to be excluded by construction in our analysis focusing on marginal elections. Nonetheless, as evident from Appendix Tables A.15 (1880 election) and A.16 (1882 and 1886 elections), excluding districts where the elected candidate of interest is an incumbent leads to very similar results.

Next, to further establish that our outcome of interest, i.e. obtaining a railway, is not uniquely determined by the level of economic development of municipalities, we also experiment deploying a richer set of controls, with the caveat that the relevant information is available only for the 1880's, i.e. after the enactment of the Baccarini Law. These additional socio-economic measures, capturing hygienic and sanitary conditions, and the amount of public revenues and expenditures, allow us to more precisely account for the municipal development level when the the railway construction plans were implemented. Our baseline results, as evidenced by Appendix Tables A.17 (for the 1880 election) and A.18 (for the 1882 and 1886 elections) are unaffected. Finally, allowing for different clustering structures (i.e. clustering by electoral district, double-clustering by electoral district and railway line, and spatial clustering based on a 50 km window around the municipality centroid) also does not affect the significance of our main results (see Appendix Tables A.19 and A.20).

5 Railways and Long-Run Development

A long-standing literature has highlighted the important contribution played by railroad construction on long-run development. Up to this point our analysis has shown that political alignment had an important effect on the pattern followed in the construction of the railroad network promoted by the Baccarini law. Hence, an important question is whether the political conditions prevailing in 1880-1890 had long-lasting consequences on the economic development of the country. To tackle this question we employ detailed municipality-level population figures for the period 1861-1991.

We start our empirical analysis by reporting some correlations patterns to illustrate the relationship between long-run population dynamics and access to the railways introduced by the Baccarini Law. Specifically, we consider the following baseline model:

$$ln(P_{mpdt}) = \delta Rail_{mp} + \beta X_{mp} + \sigma_{pt} + \phi_{dt} + \epsilon_{mpdt}. \quad (3)$$

In this model, P_{mpdt} indicates the population of municipality m, located in province p and district d, observed in census year t, with t=(1901, 1911, 1921, 1931, 1936, 1951, 1961, 1971, 1981, 1991). Given the structure of our population panel, 1901 is the first census year by which the effect of railway access can be expected to appear, as the earliest available census before then is the one from 1881, when the railway constructions were just getting underway.[32] As it is apparent from Appendix Table A.1, by 1901 the bulk of railway construction was completed. $Rail_{mp}$ is an indicator variable taking a value of one for municipalities reached by the Baccarini Law railways, and zero otherwise. σ_{pt} and ϕ_{dt} indicate, respectively, province-year and district-year fixed effects. X_{mp} is a vector of municipality-level controls. In the above specification, δ represents the effect of the Baccarini Law railways on the average level of population over the period 1901-1991. We employ robust standard errors clustered by circondario and year, with a circondario being an administrative unit, grouping multiple municipalities but smaller than provinces, that allows us to account for spatial correlation in patterns of economic development while still having a sizeable number of clusters.

Table 6 reports the results. We start with a parsimonious specification including only year fixed effects in column (1). In column (2), we augment it by accounting for a rich set of

[32]Due to budgetary problems, the 1891 census was not carried out.

municipality-level characteristics.[33] Adding province-year fixed effects in column (3) allows us to further improve the comparison of municipalities with similar characteristics and subject to similar conditions. Results indicate a strong positive correlation between population and railway access: municipalities reached by the Baccarini Law railways were on average 9.3% larger, according to the estimates in column (3). In column (4), we focus our analysis on the sample of municipalities that are in a 10 km proximity of the railways.[34]

Next, we further restrict the sample to municipalities located in electorally close districts at the time of approval of the railway. As we have shown in the previous section, quasi-random political events had a great deal of influence on the allocation of the railways across municipalities in those districts. Thus, for these municipalities, the arrival of a railway was particularly likely to be exogenous. To construct our restricted sample, we proceed trunk by trunk. For each trunk of each line, we identify the legislature in which it was approved, and include in the sample municipalities that were located in a 10 km proximity of the trunk, and belonged to districts that were marginal in that legislature. We repeat this procedure for all trunks of the Baccarini Law lines. Our favorite specification using this restricted sample is reported in column (6), where we also include district-year fixed effects. Here, we are comparing municipalities with similar socio-economic characteristics, located within 10 km of a line, in the same province and district, and only within districts that were electorally close at the time of approval of the railway. The larger coefficient in this column, compared to column (3), suggests that our earlier estimates were downward biased due to the non-random placement of the railways: in particular, these were more likely to be awarded to municipalities whose population would grow *less*.

The specification in column (6) is very demanding, since it only compares similar municipalities which have obtained a railway due to quasi-random political factors. Nevertheless, we cannot completely rule out that the allocation of the railways was also driven by endogenous factors. To the extent that the award of a railway depended on specific local population dynamics, our results would still be biased upward or downward. To address this issue, we next instrument for the placement of the railway using our measure of random alignment with the winning candidate in the legislature when the railway was approved. As a reminder, our

[33]These include log population in 1871, presence of post office, telegraph, port, railway or railway station on the territory, total number of secondary schools and libraries, terrain ruggedness, agricultural suitability, log land area, elevation, presence of a coast, quality and quantity of water available, percentage of roads with sewage system, percentage of houses with toilets, per capita number of doctors, per capita number of farmacies, per capita number of hospital beds, number of years with registered cases of cholera, total tax revenues (log per capita), revenues obtained by taxing terrains and buildings (log per capita) and expenditure on public education (log per capita).

[34]Note that here we use distance from the actual railway lines, and not the straight lines approximations as in the previous section.

Table 6. Long-Run Impact of Baccarini Law's Railway Expansion on Municipalities' Population: Panel Estimates

	Dependent variable: *Log population*						
	OLS						IV
Proximity to railways:	-	-	-	10 km	10 km	10 km	10 km
Marginal districts	-	-	-	-	Yes	Yes	Yes
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Railway access	0.6492***	0.1024***	0.0931***	0.0850***	0.1486***	0.1543***	0.2281**
	(0.0701)	(0.0266)	(0.0215)	(0.0228)	(0.0344)	(0.0328)	(0.1139)
Observations	68,159	63,309	63,309	27,010	5,450	5,450	5,370
Adjusted R^2	0.048	0.838	0.883	0.899	0.890	0.909	0.909
KP F-stat							11.80
Year FE	✓	✓					
Municipality-level controls		✓	✓	✓	✓	✓	✓
Province × year FE			✓	✓	✓	✓	✓
District × year FE						✓	✓

Notes: Table presents panel estimates regressing population on railway access. The unit of observation is a municipality. The dependent variable is the log of municipality population for periods between 1901 and 1991 (census years are 1901, 1911, 1921, 1931, 1936, 1951, 1961, 1971, 1981 and 1991). The explanatory variable is a dummy denoting municipalities reached by one of the Baccarini Law railways. Columns (1) and (2) employ year fixed effects. Column (2) adds municipality-level pre-determined controls. Column (3) adds province-by-year fixed effects. Column (4) restricts the sample to municipalities in a 10 km proximity of the railways. Column (5) further restricts the sample to municipalities that belonged to marginal electoral districts when the railway construction was contracted out. Column (6) adds district-by-year fixed effects. IV estimates in column (7) instrument railway access with the share of votes obtained in the municipality by the candidate elected in the electoral district if construction of the railway was contracted out in the 1880 parliament, and with the share of votes obtained in the municipality by the candidate elected in the last seat of the district if construction of the railway was contracted out in the 1882 or 1886 parliaments. Municipality-level controls include: log population in 1871, presence of post office, telegraph, port, railway or railway station on the territory, total number of secondary schools and libraries, terrain ruggedness, agricultural suitability, log land area, elevation, presence of a coast, quality and quantity of water available, percentage of roads with sewage system, percentage of houses with toilets, per capita number of doctors, per capita number of farmacies, per capita number of hospital beds, number of years with registered cases of cholera, per capita municipal budget measures on tax revenue and expenditures. Robust standard errors, clustered by circondario and year, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

measure captures the share of votes obtained in the municipality by the candidate winning in the district (for the 1880-1882 legislature), or by the last candidate elected (for the 1882-86 and 1886-90 legislatures). We still focus on the sample of municipalities belonging to marginal districts employed in the previous specification. Our IV specification is based on the following first and second stages:

$$Rail_{mpd} = \gamma S_{mpd} + \mu_{pt} + \nu_{dt} + \beta X_{mpd} + \varepsilon_{mpdt} \tag{4}$$

$$ln(P_{mpdt}) = \delta \widehat{Rail}_{mpd} + \sigma_{pt} + \phi_{dt} + \beta X_{mpd} + \epsilon_{mpdt} \tag{5}$$

where S_{mpd} is the instrument, and μ_{pt} and σ_{pt} are province-year and district-year fixed effects.

For our instrument to satisfy the exclusion restriction, we require that alignment with the winning candidate should not have a direct effect on population growth, other than through the construction of the railways. This assumption is likely to be met, given the prominence of the railway constructions amongst public investment programs in this period. As we already mentioned, the Italian State ended up allocating in excess of 2 billion Lire to the construction on the Baccarini Law railways between 1879–1899 (Ferrucci 1898), the bulk of which was spent in the 1880s. By comparison, State investment in the construction of roads (the second most expensive public work of the period) stood at an average 17 million Lire in 1880, 1884 and 1885 (Ministero di Agricoltura, Industria e Commercio 1884, p. 607; Ministero di Agricoltura, Industria e Commercio 1886, p. 383.) This was also a period in which transfers to the municipalities were not very large. For example, subsidies from the State and Provinces to the municipalities (to cover things such as local roads, other public works and public education) stood at a meagre 20 million Lire in 1897 (Ministero di Agricoltura, Industria e Commercio 1899, p. XII). Results for our IV specification are reported in column (7) of Table 6. The point estimate for δ rises even more (to 22.8%), confirming that our earlier estimates were downward biased.

Up to this point, the analysis has focused on the impact of the Baccarini Law railways on the average *level* of population during the period 1901-1991. We next investigate its impact on population *growth* over the same time span. To do so, we estimate the following model:

$$ln(P_{mpdt}) = \delta t Rail_{mpd} + \beta t X_{mpd} + \alpha_m + \sigma_{pt} + \phi_{dt} + \epsilon_{mpdt}. \quad (6)$$

where α_m are municipality fixed effects (which absorb the average population level), and with our railway dummy $Rail_{mpd}$ – together with the controls in X_{mpd} – interacted with a time trend t=(1901...1991). The coefficient δ then captures any differential (linear) trend in population in municipalities that obtained the railways, compared to those that did not.[35]

Such a specification also allows us to employ an IV-FE estimation strategy, based on the following model:

$$tRail_{mpd} = \gamma t S_{mpd} + \rho t X_{mpd} + \mu_m + \nu_{pt} + \upsilon_{dt} + \varepsilon_{mpdt} \quad (7)$$

[35] To interpret δ note that:

$$\frac{\hat{\delta}\,(t_1 - t_0)}{10} = E\left[ln\left(\frac{P_{mpd,t_1}}{P_{mpd,t_0}}\right)|Rail_{mpd} = 1\right] - E\left[ln\left(\frac{P_{mpd,t_1}}{P_{mpd,t_0}}\right)|Rail_{mpd} = 0\right]$$

where t_0 and t_1 are sample years, with $t_1 > t_0$ (in the above, we divide by 10, as the interval between one observation and the other in our sample is on average 10 years). δ thus represents the constant differential impact of the Baccarini Law railways on the population growth rate of a municipality crossed by a line compared to one that was not.

$$ln(P_{mpdt}) = \delta t\widehat{Rail_{mpd}} + \beta t X_{mpd} + \alpha_m + \sigma_{pt} + \phi_{dt} + \epsilon_{mpdt}. \tag{8}$$

Table 7 reports the results based on the restricted sample of municipalities in a 10 km proximity of the railways and that belonged to marginal electoral districts. We first present the results for OLS and IV regressions without municipality fixed effects (columns (1) and (2), respectively), whereas in columns (3) and (4) we include them. As we can see, including the municipality fixed effects does not substantially change the estimates. On the other hand, the IV estimates of δ are overall larger in magnitude compared to the non-instrumented ones. The coefficient estimated in column (4) implies that – between 1901 and 1991 – the population of municipalities reached by one of the Baccarini Law railways grew 52.8% faster than those that were not connected.[36] This result is in line with earlier findings in the literature – e.g. see Berger and Enflo (2017).

So far we have assumed the impact of railway access to be linear over time. We now employ the following more flexible specification to allow this effect to vary over time:

$$ln(P_{mpdt}) = \delta_t Rail_{mpd} + \beta_t X_{mpd} + \alpha_m + \sigma_{pt} + \phi_{dt} + \epsilon_{mpdt} \tag{9}$$

with t=(1861...1991) and 1861 being the base year omitted in the estimates. In this specification, δ_t captures the average difference in population growth (relative to the base year) across different years between municipalities reached by the railways and those that were not reached.

The results of our event study are reported in Figure 2. To disentangle the effect of political factors prevailing at the time of the railroad construction on long-run population growth, analogously to our previous analysis, we focus on municipalities lying within 10 km of the Baccarini Law railways and that belonged to districts where election outcomes were determined by a close margin.

Figure 2-(A) illustrates the pattern of population growth for all municipalities within 10 km of the railways, whereas Figure 2-(B) reports the results further restricting the sample to municipalities that belonged to electoral districts characterised by marginal elections at the time the construction of the railways was contracted out. The results in Figure 2-(A) suggest that the positive impact of railways can be only observed starting in the post-WWII period. Note though that the insignificant coefficients in the short and medium run could be due to the non-random placement of the railways, which we have found evidence of earlier in this section. To the extent that the railways were disproportionately allocated to municipalities which would grow less, our estimated coefficients would be downward biased. In Figure 2-(B) we alleviate this concern by restricting our attention to municipalities that belonged to

[36] I.e. $e^{0.047(1991-1901)/10} - 1 = 0.528$.

Table 7. Long-Run Impact of Baccarini Law's Railway Expansion on Municipalities' Population: Comparing Specifications

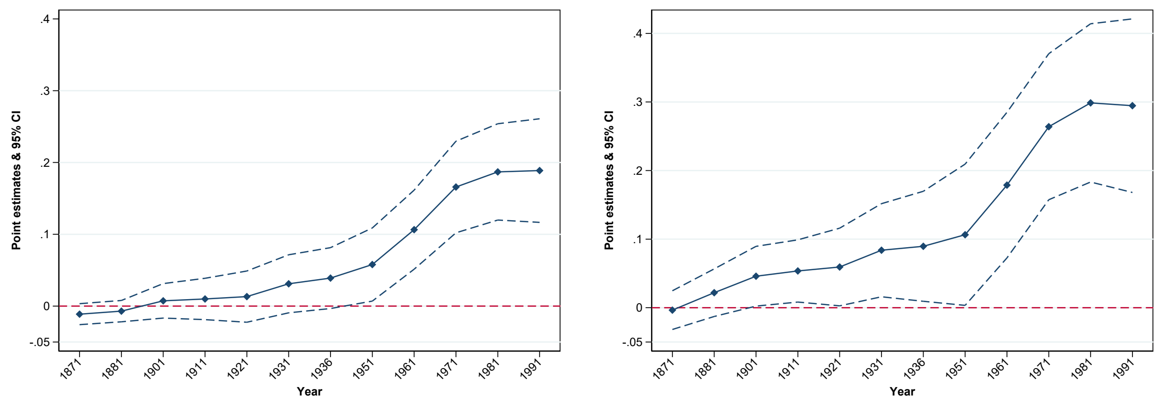
	Dependent variable: *Log population*			
	OLS	IV	FE	IV-FE
	(1)	(2)	(3)	(4)
Railway access × year	0.0293***	0.0436**	0.0243***	0.0471*
	(0.0041)	(0.0186)	(0.0055)	(0.0255)
First stage:				
Winner share in municipality		0.0778***		0.0775***
		(0.0208)		(0.0214)
diff. in population growth in 1901-1991	30.17%	48.05%	24.44%	52.79%
Observations	5,370	5,370	5,370	5,370
Adjusted R^2	0.913	0.912	0.533	0.521
KP F-stat		13.97		13.10
Municipality-level controls × year	✓	✓	✓	✓
Municipality FE			✓	✓
Province × year FE	✓	✓	✓	✓
District × year FE	✓	✓	✓	✓

Notes: Table presents panel estimates regressing population on railway access. The unit of observation is a municipality. The dependent variable is the log of municipality population for periods between 1901 and 1991 (census years are 1901, 1911, 1921, 1931, 1936, 1951, 1961, 1971, 1981 and 1991). The explanatory variable is a dummy denoting municipalities reached by one of the Baccarini Law railways, interacted with year. IV estimates in columns (2) and (4) instrument railway access with the share of votes obtained in the municipality by the candidate elected in the electoral district if construction of the railway was contracted out in the 1880 parliament, and with the share of votes obtained in the municipality by the candidate elected in the last seat of the district if construction of the railway was contracted out in the 1882 or 1886 parliaments. All specification employ municipality-level pre-determined controls, interacted with year, province-year fixed effects and district-year fixed effects. Columns (3) and (4) employ municipality fixed effects. All specification limit the sample to municipalities in a 10 km proximity of the railways and that belonged to marginal electoral districts when the railway construction was contracted out. Municipality-level controls include: log population in 1871, presence of post office, telegraph, port, railway or railway station on the territory, total number of secondary schools and libraries, terrain ruggedness, agricultural suitability, log land area, elevation, presence of a coast, quality and quantity of water available, percentage of roads with sewage system, percentage of houses with toilets, per capita number of doctors, per capita number of farmacies, per capita number of hospital beds, number of years with registered cases of cholera, per capita municipal budget measures on tax revenue and expenditures. Robust standard errors, clustered by circondario and year, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

marginal electoral districts. Our estimates in this case indicate that the positive effect of railroad access on population growth starts earlier – i.e. already in 1901 – and are larger. In fact, for municipalities with railroad access we can observe differences in population growth around 5% by 1901, increasing to almost 30% by 1991.

As previously mentioned, the substantial impact of the Baccarini Law railways on population growth is observed starting in the post-WWII period (this can be seen in both Figure 2-(A) and Figure 2-(B)). To explore this more formally, we repeat the IV regressions previously employed on restricted time periods: 1901-1936, 1936-1951, and 1951-1991. Appendix Table A.21 reports the estimates for the specification without municipality fixed effects (for comparison, see column (7) in Table 6). As can be inferred, in the first two periods we consider the effect

Figure 2. Long-Run Impact of Baccarini Law's Railway Expansion on Municipalities' Population: Event Study 1871-1991



(A) 10 km proximity of railways

(B) 10 km proximity and marginal districts

Notes: Figures plot coefficients from fixed effects estimates, with connected solid lines corresponding to point estimates and dashed lines to 95% confidence intervals based on robust standard errors clustered at the circondario level. The dependent variable is the log of municipality population. The explanatory variable is a dummy denoting municipalites reached by one of the Baccarini Law railways, interacted with year fixed effects. All specifications include municipality fixed effects, municipality-level controls interacted with year, province-year fixed effects and electoral district-year fixed effects. Figure 2-(A) employs the sample of municipalities that are in a 10 km proximity of the railways. Figure 2-(B) further restricts the sample to municipalities that belonged to electoral districts characterised by marginal elections when the railway construction was contracted out. Municipality-level controls include: log population in 1861, presence of post office, telegraph, port, railway or railway station on the territory, total number of secondary schools and libraries, terrain ruggedness, agricultural suitability, log land area, elevation, presence of a coast, quality and quantity of water available, percentage of roads with sewage system, percentage of houses with toilets, per capita number of doctors, per capita number of farmacies, per capita number of hospital beds, number of years with registered cases of cholera, per capita municipal budget measures on tax revenue and expenditures.

on population is positive, being around 10%, though it is only significant, at the 10-percent level, for the period 1901-1936, while the effect in the post-WWII period is substantially larger, growing to 34%. In Appendix Table A.22 we present IV estimates additionally employing municipality fixed effects (as in column (4) of Table 7). The observed pattern is very similar. The effect in the first two periods under consideration is not significantly different from zero, while in the post-WWII period municipalities reached by the Baccarini Law railways grew faster by around 11% per 10-year period. All the findings above indicate that railroad access had strong and persistent effects on growth and that short-run political factors influencing infrastructural development had long-lasting consequences.

6 Conclusions

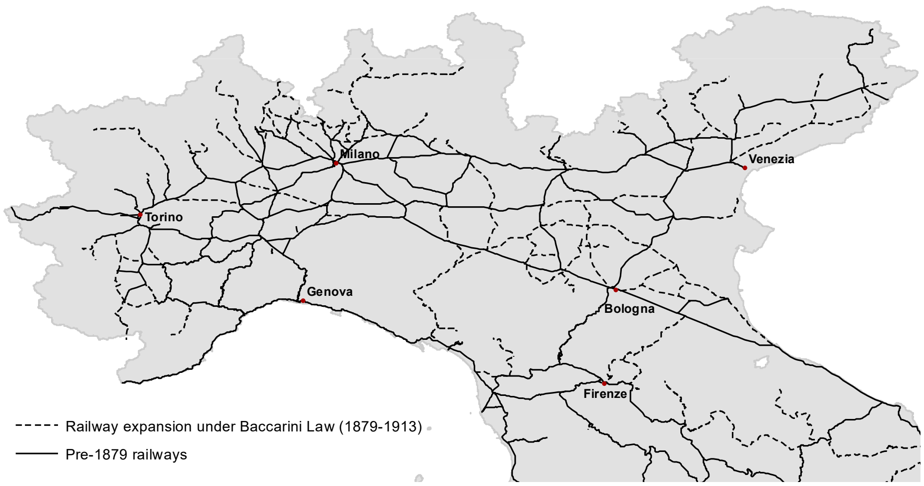
In this paper, we have shown that electoral politics has heavily influenced the allocation of secondary railways in Italy in 1876-1890. Politics mattered at two distinct levels. First, the initial choice as to which city pairs to connect was tilted in favour of connections which cut across districts that had voted for the government in the 1876 election. Second, once it was decided that a city pair would be connected, the actual route followed by the railway was tilted in favour of municipalities that had voted with the government in the election just preceding the construction of the railway. Our identification strategy allows us to attribute these effects to a causal impact of political alignment on the arrival of the railway. We interpret this as an example of pork-barrel politics.

We have also shown that, by influencing the allocation of the railways, political alignment in the 1876, 1880, 1882 and 1886 elections had important, long-run implications for growth and the spatial organisation of the Italian economy. Municipalities that voted for the government in the relevant election - and as a result obtained the railways - experienced considerably faster population growth in the century that followed (1901-1991). Growth differentials were already present at the beginning of this period, but widened after World War II, suggesting that the diverging trajectories created by political alignment were further pulled apart by the post-World War II economic miracle. This seems a remarkable example of how even minor political events occurring at a “critical juncture” (Acemoglu and Robinson 2012) may be later compounded by unrelated events, and thus have a big effect on long-run development.

Our research could be extended in two directions. First, one could study whether the impact of political alignment differed systematically across Italian regions, depending on the quality of pre-unitarian institutions. Since our events occurred shortly after unification (1861), it is possible that pre-unitarian institutions still determined the functioning of politics at the local level. Thus, one might be able to gauge the role of political institutions in limiting the prevalence of pork-barrel politics. Second, one could go beyond our reduced form estimates and attempt the estimate the parameters of a structural model of politics, infrastructure allocation and economic development. The role of such model would be to quantify the welfare cost of having infrastructure decisions mediated by electoral politics (as opposed to determined by a benign planner), also depending on the quality of political institutions. Ultimately, such a model might reveal the “true” welfare gains of a large infrastructure investment, that is those to be expected when the infrastructure is allocated neither randomly nor optimally, but rather through the functioning of real-world political institutions.

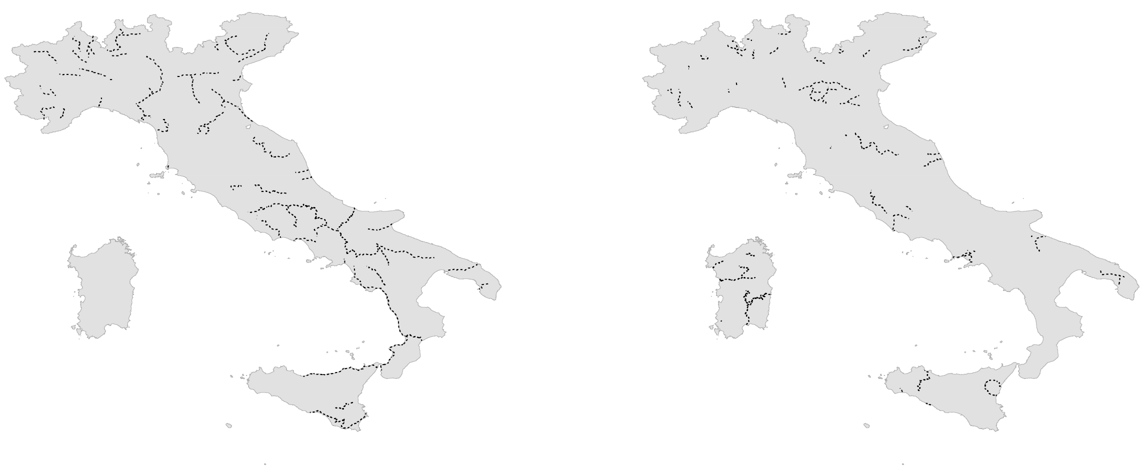
Appendix A - Tables and Figures

Figure A.1. Railways Expansion in Northern Italy, Pre- and Post-1879



Notes: Extent of the northern Italian railway network before 1879 and lines constructed between 1879 and 1913 under the Baccarini Law. Authors' elaborations based on the data by Ciccarelli and Groote (2017).

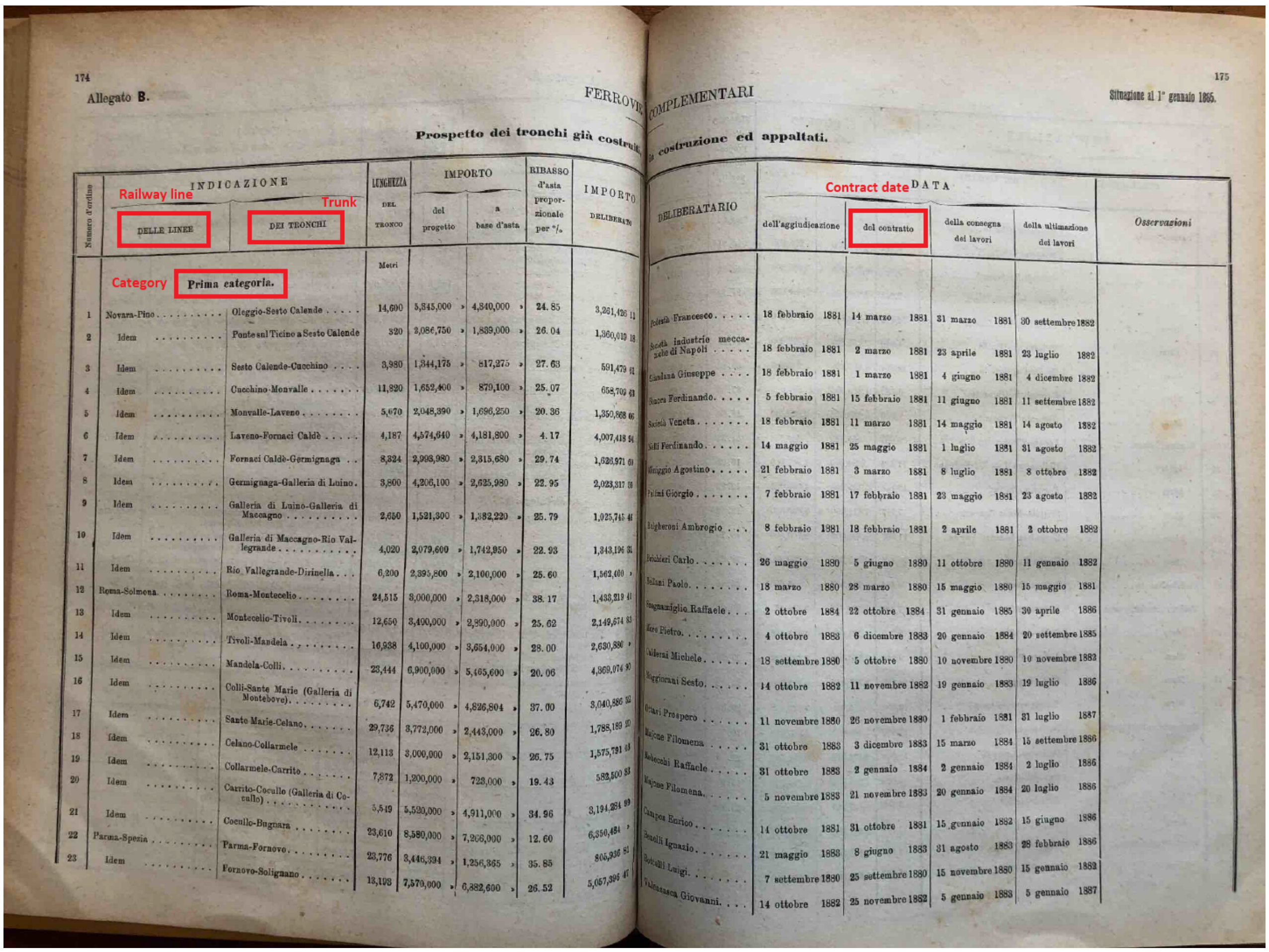
Figure A.2. Railways Expansion in Italy under Baccarini Law



(A) Railway expansion 1879-1913, category 1-3 lines (B) Railway expansion 1879-1913, category 4-5 lines

Notes: Railways constructed between 1879 and 1913 under the Baccarini Law, divided between category 1-3 lines and category 4-5 lines. Authors' elaborations based on the data by Ciccarelli and Groote (2017).

FIGURE A.3. CONSTRUCTION REPORTS



174

Allegato **B.**

FERROVIE COMPLEMENTARI

Prospetto dei tronchi già costruiti, in costruzione ed appaltati.

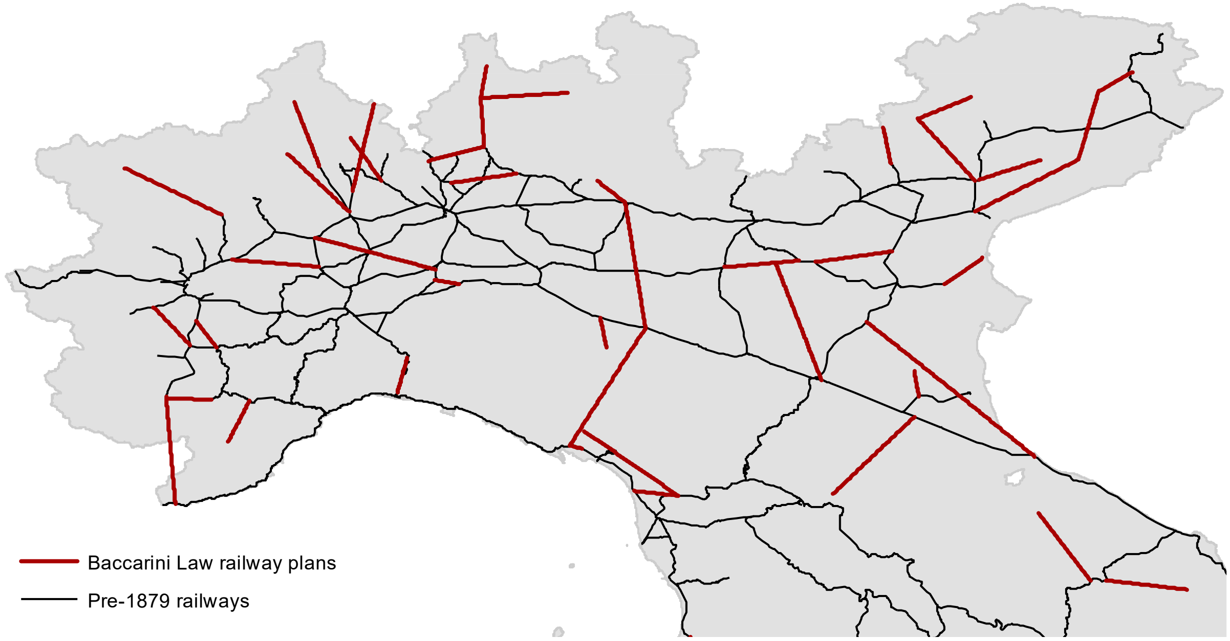
175

Situazione al 1° gennaio 1885.

Numero d'ordine	INDICAZIONE delle linee	INDICAZIONE dei tronchi	LUNGHEZZA del tronco	IMPORTO del progetto	IMPORTO a base d'asta	RIBASSO d'asta proporzionale per °/₀	IMPORTO DELIBERATO	DELIBERATARIO	DATA dell'aggiudicazione	DATA del contratto	DATA della consegna dei lavori	DATA della ultimazione dei lavori	Osservazioni
			Metri										
	Prima categoria.												
1	Novara-Pino	Oleggio-Sesto Calende	14,600	5,345,000 »	4,340,000 »	24. 85	3,261,426 11	Podestà Francesco	18 febbraio 1881	14 marzo 1881	31 marzo 1881	30 settembre 1882	
2	Idem	Ponte sul Ticino a Sesto Calende	320	2,086,750 »	1,839,000 »	26. 04	1,360,019 18	Società industrie meccaniche di Napoli	18 febbraio 1881	2 marzo 1881	23 aprile 1881	23 luglio 1882	
3	Idem	Sesto Calende-Cucchino	3,980	1,344,175 »	817,275 »	27. 63	591,479 61	Giandana Giuseppe	18 febbraio 1881	1 marzo 1881	4 giugno 1881	4 dicembre 1882	
4	Idem	Cucchino-Monvalle	11,820	1,652,400 »	879,100 »	25. 07	658,709 43	[illegible] Ferdinando	5 febbraio 1881	15 febbraio 1881	11 giugno 1881	11 settembre 1882	
5	Idem	Monvalle-Laveno	5,670	2,048,390 »	1,696,250 »	20. 36	1,350,868 06	Società Veneta	18 febbraio 1881	11 marzo 1881	14 maggio 1881	14 agosto 1882	
6	Idem	Laveno-Fornaci Caldè	4,187	4,574,640 »	4,181,800 »	4. 17	4,007,418 94	[illegible] Ferdinando	14 maggio 1881	25 maggio 1881	1 luglio 1881	31 agosto 1882	
7	Idem	Fornaci Caldè-Germignaga	8,324	2,993,980 »	2,315,680 »	29. 74	1,626,971 60	[illegible] Agostino	21 febbraio 1881	3 marzo 1881	8 luglio 1881	8 ottobre 1882	
8	Idem	Germignaga-Galleria di Luino	3,800	4,206,100 »	2,625,980 »	22. 95	2,023,317 [illegible]	Palmi Giorgio	7 febbraio 1881	17 febbraio 1881	23 maggio 1881	23 agosto 1882	
9	Idem	Galleria di Luino-Galleria di Maccagno	2,650	1,521,300 »	1,382,220 »	25. 79	1,025,745 46	Balgheroni Ambrogio	8 febbraio 1881	18 febbraio 1881	2 aprile 1881	2 ottobre 1882	
10	Idem	Galleria di Maccagno-Rio Vallegrande	4,020	2,079,600 »	1,742,950 »	22. 93	1,343,196 31	Brichieri Carlo	26 maggio 1880	5 giugno 1880	11 ottobre 1880	11 gennaio 1882	
11	Idem	Rio Vallegrande-Dirinella	6,200	2,395,800 »	2,100,000 »	25. 60	1,562,400 »	Bellani Paolo	18 marzo 1880	28 marzo 1880	15 maggio 1880	15 maggio 1881	
12	Roma-Solmona	Roma-Montecelio	24,515	3,000,000 »	2,318,000 »	38. 17	1,433,219 41	Scognamiglio Raffaele	2 ottobre 1884	22 ottobre 1884	31 gennaio 1885	30 aprile 1886	
13	Idem	Montecelio-Tivoli	12,650	3,400,000 »	2,890,000 »	25. 62	2,149,674 83	[illegible] Pietro	4 ottobre 1883	6 dicembre 1883	20 gennaio 1884	20 settembre 1885	
14	Idem	Tivoli-Mandela	16,938	4,100,000 »	3,654,000 »	28. 00	2,630,880 »	Calderai Michele	18 settembre 1880	5 ottobre 1880	10 novembre 1880	10 novembre 1882	
15	Idem	Mandela-Colli	23,444	6,900,000 »	5,465,600 »	20. 06	4,369,074 90	Maggiorani Sesto	14 ottobre 1882	11 novembre 1882	19 gennaio 1883	19 luglio 1886	
16	Idem	Colli-Sante Marie (Galleria di Montebove)	6,742	5,470,000 »	4,826,804 »	37. 00	3,040,886 32	Ottavi Prospero	11 novembre 1880	26 novembre 1880	1 febbraio 1881	31 luglio 1887	
17	Idem	Sante Marie-Celano	29,736	3,772,000 »	2,443,000 »	26. 80	1,788,189 20	Majone Filomena	31 ottobre 1883	3 dicembre 1883	15 marzo 1884	15 settembre 1886	
18	Idem	Celano-Collarmele	12,113	3,000,000 »	2,151,300 »	26. 75	1,575,791 43	Rebecchi Raffaele	31 ottobre 1883	2 gennaio 1884	2 gennaio 1884	2 luglio 1886	
19	Idem	Collarmele-Carrito	7,872	1,200,000 »	723,000 »	19. 43	583,500 83	Majone Filomena	5 novembre 1883	21 novembre 1883	20 gennaio 1884	20 luglio 1886	
20	Idem	Carrito-Cocullo (Galleria di Cocullo)	5,549	5,520,000 »	4,911,000 »	34. 96	3,194,284 99	Campos Enrico	14 ottobre 1881	31 ottobre 1881	15 gennaio 1882	15 giugno 1886	
21	Idem	Cocullo-Bugnara	23,610	8,580,000 »	7,266,000 »	12. 60	6,350,484 »	Benelli Ignazio	21 maggio 1883	8 giugno 1883	31 agosto 1883	28 febbraio 1886	
22	Parma-Spezia	Parma-Fornovo	23,776	3,446,394 »	1,256,365 »	35. 85	805,936 81	Bottelli Luigi	7 settembre 1880	25 settembre 1880	15 novembre 1880	15 gennaio 1882	
23	Idem	Fornovo-Solignano	13,193	7,570,000 »	6,882,600 »	26. 52	5,057,396 47	[illegible] Giovanni	14 ottobre 1882	25 novembre 1882	5 gennaio 1883	5 gennaio 1887	

Notes: Construction reports from Ministero dei Lavori Pubblici (1885).

Figure A.4. Baccarini Law Railway Plans



Notes: Extent of the northern Italian railway network before 1879 and straight-line railway plans as based on the focal points indicated by the Baccarini Law.

Table A.1. Baccarini Law (1879), Total Km Predicted and Timing of Construction

	1^{st} category	2^{nd} category	3^{rd} category	4^{th} category	5^{th} category	Total
Km. completed by end of legislature:						
1876 - 1880	0	0	0	0	0	0
1880 - 1882	165	36	53	16	0	270
1882 - 1886	347	250	449	305	140	1,491
1886 - 1890	773	715	996	838	621	3,944
1890 - 1892	805	910	1,417	894	738	4,764
1892 - 1895	1,009	1,086	1,544	894	1,083	5,617
1895 - 1897	1,102	1,155	1,654	1,067	1,091	6,069
1897 - 1900	1,109	1,160	1,871	1,068	1,096	6,305
1900 - 1904	1,109	1,203	1,915	1,068	1,157	6,453
1904 - 1909	1,109	1,203	1,915	1,247	1,157	6,632
1909 - 1913	1,141	1,233	1,986	1,276	1,157	6,794
Km. predicted:	1,153	1,267	2,070	2,530	-	

Notes: The 2530 km of predicted category-4 lines include the 1000 km added to this category by the Legge 27 aprile 1885 n. 3048. Authors' elaborations based on the data by Ciccarelli and Groote (2017).

Table A.2. Summary Statistics on District-Level Voting

	Mean	*Standard Deviation*	*Number of districts*
1876 elections (2-12 November 1876)			
Population	50,579	11,544	508
Franchise	0.027	0.038	508
Turnout	0.637	0.137	508
1880 elections (16-23 May 1880)			
Population	52,098	12,156	508
Franchise	0.027	0.035	508
Turnout	0.651	0.130	508
1882 elections (29 October-5 November 1882)			
Population	206,823	62,485	135
Franchise	0.076	0.031	135
Turnout	0.624	0.123	135
1886 elections (23-30 May 1886)			
Population	212,181	65,621	135
Franchise	0.088	0.034	135
Turnout	0.604	0.135	135

Notes: Authors' elaborations based on the data by Corbetta and Piretti (2009).

Table A.3. Summary Statistics on Municipality-Level Population

Year	*Mean*	*S.D.*	*Min*	*Max*
1861	3,314	9,546	56	484,026
1871	3,539	10,113	59	489,008
1881	3,754	11,270	60	535,206
1901	4,273	14,425	56	621,213
1911	4,647	17,393	58	751,211
1921	4,990	20,440	58	859,629
1931	5,202	23,022	93	960,660
1936	5,376	25,794	116	1,150,338
1951	6,027	31,779	74	1,651,393
1961	6,420	39,942	90	2,187,682
1971	6,864	46,507	51	2,781,385
1981	7,174	46,269	32	2,839,638
1991	7,196	43,324	10	2,775,250

Notes: Authors' elaborations based on the data by Sistema Statistico Nazionale (1994).

Table A.4. Baccarini Law's Railway Projects and 1876 District-Level Elections Regression Discontinuity Estimates: Controls' Coefficients

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	*log(km) dist. from projects*		*10 km proximity to projects*		*5 km proximity to projects*	
RD polynomial order:	3rd	2nd	3rd	2nd	3rd	2nd
Margin:	-	± 5%	-	± 5%	-	± 5%
Government-party win in district	-0.3081***	-0.3587**	0.1892**	0.2494**	0.1898***	0.3508***
	(0.1180)	(0.1722)	(0.0789)	(0.1154)	(0.0659)	(0.0952)
Log population density 1871	-0.1109***	0.0125	0.0587***	-0.0524**	0.0566***	-0.0155
	(0.0194)	(0.0366)	(0.0117)	(0.0218)	(0.0108)	(0.0214)
Post office in 1871	0.0154	0.0365	0.0068	0.0014	-0.0073	-0.0235
	(0.0232)	(0.0341)	(0.0145)	(0.0252)	(0.0136)	(0.0242)
Telegraph office in 1871	-0.1217***	0.0147	0.0394**	0.0146	0.0442**	0.0137
	(0.0360)	(0.0513)	(0.0200)	(0.0337)	(0.0193)	(0.0352)
Railway station in 1871	0.1246***	-0.0759	-0.0689***	0.0026	-0.0530**	0.0492
	(0.0442)	(0.0609)	(0.0258)	(0.0440)	(0.0239)	(0.0405)
Sea port in 1871	-0.2204	-0.3049	0.1136*	0.2056	0.1263*	0.1100
	(0.1549)	(0.2850)	(0.0661)	(0.1522)	(0.0710)	(0.1803)
N. secondary schools and libraries	-0.0301	-0.1430***	0.0080	0.0665***	0.0192*	0.0928***
	(0.0207)	(0.0459)	(0.0097)	(0.0189)	(0.0105)	(0.0207)
Terrain ruggedness	-0.4320***	0.1504	0.2046***	-0.0435	0.1478**	-0.2153*
	(0.1240)	(0.2115)	(0.0705)	(0.1277)	(0.0677)	(0.1214)
Log land area	-0.0506***	-0.0037	0.0769***	0.0208	0.0641***	0.0271
	(0.0167)	(0.0268)	(0.0098)	(0.0179)	(0.0093)	(0.0169)
Coast	0.1047**	0.0990	-0.0524**	-0.0413	-0.0200	-0.0454
	(0.0434)	(0.0744)	(0.0237)	(0.0429)	(0.0220)	(0.0449)
Elevation	1.2261***	0.6561***	-0.6553***	-0.3633**	-0.5453***	-0.2076
	(0.1250)	(0.2217)	(0.0728)	(0.1750)	(0.0674)	(0.1388)
Market access	-1.3829***	-1.6180**	0.7915***	0.1036	0.5275***	0.3649
	(0.2246)	(0.7343)	(0.1164)	(0.4185)	(0.1157)	(0.4308)
Adjusted R^2	0.515	0.733	0.260	0.528	0.196	0.370
Observations	6,957	1,471	6,957	1,471	6,957	1,471
Number of electoral district	458	102	458	102	458	102
Municipality-level controls	✓	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓

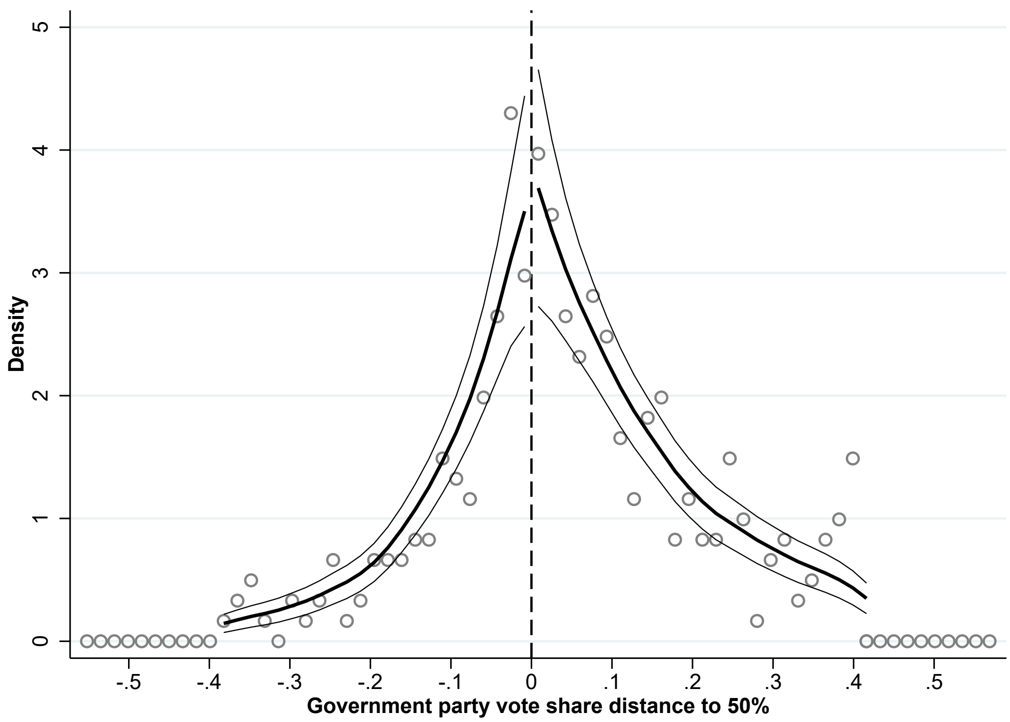
Notes: The table reports reports RD estimates. The unit of observation is a municipality. The dependent variable is the distance in log(km) from the railway projects in columns (1)-(2); a binary indicator for being in a 10 km proximity to the railway projects in columns (3)-(4); a binary indicator for being in a 5 km proximity to the railway projects in columns (5)-(6). The distance and proximity measures are based on straight-line overall projects, connecting the focal points of the line as indicated on the Baccarini Law. The explanatory variable indicates a government-party win in the electoral district, with 3rd-order polynomials on the margin of victory being employed in odd-numbered columns and 2nd order polynomials in even-numbered columns. In even-numbered columns, the sample is limited to electoral districts where the government-party candidate's vote share was within a 5% distance of the 50% threshold needed for election. Municipality-level controls include: log population density in 1871, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation and presence of a coast. All specifications employ province fixed effects. Standard errors adjusted for spatial correlation, based on a 20 km window, in parenthesis. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.5. Balance Test across Marginal Districts in 1876

	(1) *Gov. win*	(2) *Gov. loss*	(3) *Difference*
Pre-1879 and geographic characteristics			
Log population density in 1871	4.6714	4.7117	0.0403
Post office in 1871	0.4716	0.3805	-0.0911
Telegraph office in 1871	0.2371	0.1885	-0.0486
Railway station in 1871	0.1567	0.1132	-0.0435
Sea port in 1871	0.0113	0.0000	-0.0113
Number of secondary schools and libraries in 1863	0.1372	0.1835	0.0463
Market access	0.4607	0.5371	0.0765
Wheat suitability	0.3755	0.3483	-0.0273
Cereals suitability	0.3086	0.3177	0.0091
Rice suitability	0.0393	0.0653	0.0260
Cotton suitability	0.2541	0.2004	-0.0538
Barley suitability	0.3812	0.3527	-0.0285
Rye suitability	0.3957	0.3716	-0.0241
Olive suitability	0.4267	0.3136	-0.1130**
Citrus suitability	0.2752	0.0473	-0.2279***
Oat suitability	0.3810	0.3498	-0.0313
Bean suitability	0.3505	0.3214	-0.0291
Terrain ruggedness	0.2202	0.2279	0.0077
Log land area	3.3642	3.2655	-0.0987
Elevation (m)	312.4882	324.4631	11.9750
Coast	0.1091	0.0322	-0.0769
Additional characteristics (post-1879)			
Quantity of water	2.6552	2.6291	-0.0261
Quality of water	2.6456	2.8201	0.1745*
% of roads with sewage	0.1748	0.2089	0.0341
% of houses with toilets	0.4509	0.4873	0.0364
Number of farmacies (per capita)	0.0004	0.0003	-0.0001*
Number of medics (per capita)	0.0006	0.0004	-0.0001**
Number of years with colera epidemics	1.9167	1.5915	-0.3252
Number of hospital beds (per capita)	0.0010	0.0008	-0.0001
Revenue tax on terrains (log per capita)	1.5445	1.5095	-0.0350
Revenue tax on buildings (log per capita)	0.5665	0.4781	-0.0883*
Municipal surtax (log per capita)	1.4997	1.6603	0.1606
Total tax revenues (log per capita)	2.2480	2.2071	-0.0409
Education ordinary expenses (log per capita)	0.7432	0.7493	0.0061
Education extraordinary expenses (log per capita)	0.0539	0.0868	0.0329*
Education optional expenses (log per capita)	0.1409	0.1625	0.0215

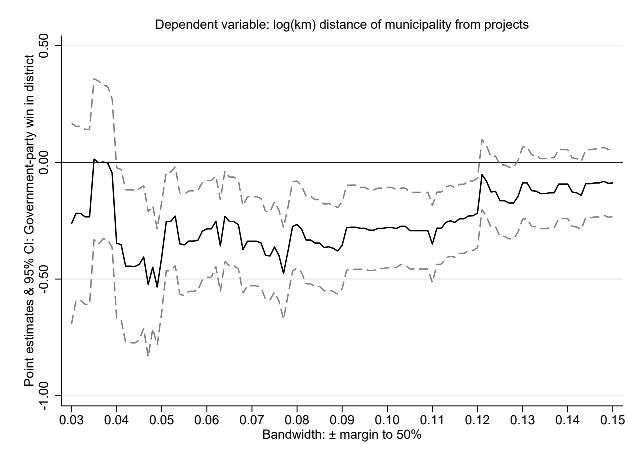
Notes: Balance test of municipality characteristics across marginal electoral districts. Column (1) reports mean values in electoral districts won by the government party with a vote share margin within 5%. Column (2) reports mean values in electoral districts lost by the government party with a vote share margin within 5%. Column (3) reports the difference in means, with significance stars based on a t-test on the equality of means. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Figure A.5. Government Party Vote Share - McCrary Test

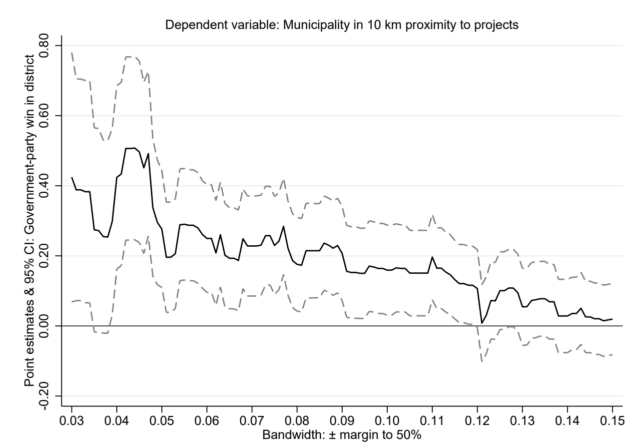


Notes: The figure implements the sorting test suggested by McCrary (2008), by plotting the density of observations in each government party vote share bin to test if there is a discountinuity at the 50% threshold.

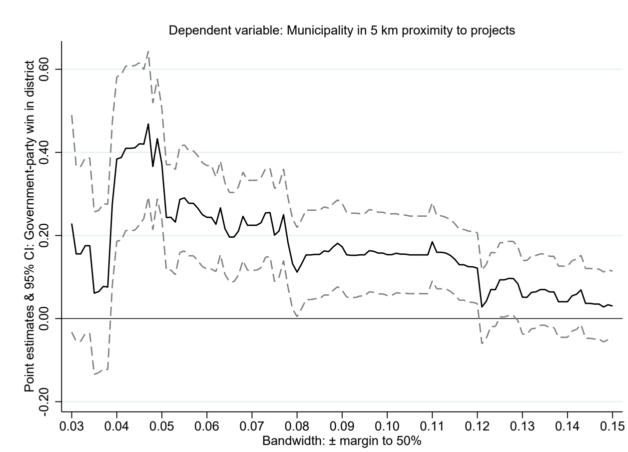
Figure A.6. Baccarini Law's Railway Projects and 1876 District-Level Elections: Sensitivity of RD Estimates to Margin



(A)



(B)



(C)

Notes: Figures plot coefficients from RD estimates, with solid lines corresponding to point estimates and dashed lines to 95% confidence intervals, based on different margin of victory bandwidths.

Table A.6. Baccarini Law's Railway Projects and 1876 District-Level Elections
Regression Discontinuity Estimates: Contemporary Controls

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	*log(km) dist. from projects*		*10 km proximity to projects*		*5 km proximity to projects*	
RD polynomial order:	3rd	2nd	3rd	2nd	3rd	2nd
Margin:	-	± 5%	-	± 5%	-	± 5%
Government-party win in district	-0.2757**	-0.4053**	0.1764**	0.2764**	0.1757***	0.3721***
	(0.1167)	(0.1721)	(0.0782)	(0.1122)	(0.0656)	(0.0932)
Adjusted R^2	0.518	0.735	0.260	0.531	0.190	0.375
Observations	6,801	1,464	6,801	1,464	6,801	1,464
Number of electoral district	458	102	458	102	458	102
Municipality-level controls	✓	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓

Notes: The table reports reports RD estimates. The unit of observation is a municipality. The dependent variable is the distance in log(km) from the railway projects in columns (1)-(2); a binary indicator for being in a 10 km proximity to the railway projects in columns (3)-(4); a binary indicator for being in a 5 km proximity to the railway projects in columns (5)-(6). The distance and proximity measures are based on straight-line overall projects, connecting the focal points of the line as indicated on the Baccarini Law. The explanatory variable indicates a government-party win in the electoral district, with 3rd-order polynomials on the margin of victory being employed in odd-numbered columns and 2nd order polynomials in even-numbered columns. In even-numbered columns, the sample is limited to electoral districts where the government-party candidate's vote share was within a 5% distance of the 50% threshold needed for election. Municipality-level controls include: log population density in 1871, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation, presence of a coast, quality and quantity of water available, percentage of roads with sewage system, percentage of houses with toilets, per capita number of doctors, per capita number of farmacies, per capita number of hospital beds, number of years with registered cases of cholera, per capita municipal budget measures on tax revenue and expenditures. All specifications employ province fixed effects. Standard errors adjusted for spatial correlation, based on a 20 km window, in parenthesis. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

TABLE A.7. BACCARINI LAW'S RAILWAY PROJECTS AND 1876 DISTRICT-LEVEL ELECTIONS
REGRESSION DISCONTINUITY ESTIMATES: DIFFERENT CLUSTERING METHODS

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	*log(km) dist. from projects*		*10 km proximity to projects*		*5 km proximity to projects*	
RD polynomial order:	3rd	2nd	3rd	2nd	3rd	2nd
Margin:	-	± 5%	-	± 5%	-	± 5%
Government-party win in district	-0.3081	-0.3587	0.1892	0.2494	0.1898	0.3508
District cluster s.e.	(0.1472)**	(0.1907)*	(0.0947)**	(0.1174)**	(0.0789)**	(0.1075)***
District-Line double-cluster s.e.	[0.1194]***	[0.1270]***	[0.0807]**	[0.0798]***	[0.0873]**	[0.1167]***
Spatially-adjusted s.e. - 50km	{0.1303}**	{0.1977}*	{0.0783}**	{0.1342}*	{0.0741}**	{0.1135}***
Adjusted R^2	0.515	0.733	0.260	0.528	0.196	0.370
Observations	6,957	1,471	6,957	1,471	6,957	1,471
Number of electoral district	458	102	458	102	458	102
Municipality-level controls	✓	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓	✓

Notes: The table reports reports RD estimates. The unit of observation is a municipality. The dependent variable is the distance in log(km) from the railway projects in columns (1)-(2); a binary indicator for being in a 10 km proximity to the railway projects in columns (3)-(4); a binary indicator for being in a 5 km proximity to the railway projects in columns (5)-(6). The distance and proximity measures are based on straight-line overall projects, connecting the focal points of the line as indicated on the Baccarini Law. The explanatory variable indicates a government-party win in the electoral district, with 3rd-order polynomials on the margin of victory being employed in odd-numbered columns and 2nd order polynomials in even-numbered columns. In even-numbered columns, the sample is limited to electoral districts where the government-party candidate's vote share was within a 5% distance of the 50% threshold needed for election. Municipality-level controls include: log population density in 1871, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation and presence of a coast. All specifications employ province fixed effects. In parenthesis are robust standard errors clustered by electoral district. In square brackets are robust standard errors double-clustered by electoral district and railway line. In curly brackets are standard errors adjusted for spatial correlation based on a 50 km window. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.8. Implementation of Baccarini Law's Railway Projects and 1880 District-Level Elections: Regression Discontinuity Estimates

	(1)	(2)	(3)	(4)
Dependent variable:	*Municipality obtains rail during legislature*			
Proximity to projects:	10 km		5 km	
RD polynomial order:	3rd	2nd	3rd	2nd
Margin:	-	± 5%	-	± 5%
Government-party win in district	-0.0391	0.2338	-0.0622	0.6238
	(0.0371)	(0.1598)	(0.0709)	(0.4354)
Adjusted R^2	0.268	0.282	0.287	0.325
Observations	2,774	650	1,727	439
Number of electoral district	298	75	243	61
Municipality-level controls	✓	✓	✓	✓
Province FE	✓	✓	✓	✓

Notes: The table reports reports RD estimates. The unit of observation is a municipality. The dependent variable indicates the passage of a railway on the municipality territory, which construction was contracted out during the 1880 legislature. The explanatory variable indicates a government-party win in the electoral district, with 3rd-order polynomials on the margin of victory being employed in odd-numbered columns and 2nd order polynomials in even-numbered columns. Columns (1)-(2) employ the sample of municipalities in a 10 km proximity of the railway plans, while columns (3)-(4) employ the sample of municipalities a 5 km proximity of the railway plans. In even-numbered columns, the sample is limited to electoral districts where the government-party candidate's vote share was within a 5% distance of the 50% threshold needed for election. Municipality-level controls include: log population density in 1871, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation and presence of a coast. All specifications employ province fixed effects. Standard errors adjusted for spatial correlation, based on a 20 km window, in parenthesis. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.9. Implementation of Baccarini Law's Railway Projects and 1880 Municipality-Level Electoral Outcomes: Controls' Coefficients

	(1)	(2)
Dependent variable:	*Municipality obtains rail during legislature*	
Proximity to projects:	10 km	5 km
Margin of victory:	$\leq$ 5%	$\leq$ 5%
Winner share in municipality	-0.0179	-0.0505
× Opposition-party win in district	(0.0241)	(0.0336)
Winner share in municipality	0.0470***	0.0486**
× Government-party win in district	(0.0157)	(0.0192)
Log population density 1871	0.0874**	0.1395***
	(0.0338)	(0.0425)
Post office in 1871	0.0120	-0.0167
	(0.0234)	(0.0341)
Telegraph office in 1871	0.0395	0.0637
	(0.0603)	(0.0764)
Railway station in 1871	-0.2147***	-0.1694*
	(0.0664)	(0.0925)
Sea port in 1871	-0.4292**	-0.4288**
	(0.1641)	(0.1645)
Total secondary schools and libraries	-0.0303	-0.0472
	(0.0469)	(0.0559)
Terrain ruggedness	-0.0224	0.0209
	(0.1354)	(0.1723)
Log land area	0.0343	0.0652**
	(0.0249)	(0.0312)
Coast	-0.0936**	-0.0941**
	(0.0358)	(0.0441)
Elevation	0.1814	0.1624
	(0.1411)	(0.1790)
Market access	-1.0577	-0.9719
	(0.6917)	(0.7303)
Adjusted R^2	0.313	0.362
Observations	619	420
Number of electoral district	75	61
Municipality-level controls	✓	✓
Province FE	✓	✓
District FE	✓	✓

Notes: The table reports reports OLS estimates. The unit of observation is a municipality. The dependent variable indicates the passage of a railway on the municipality territory, which construction was contracted out during the 1880 legislature. The main explanatory variables indicate the share of votes obtained in the municipality by the candidate elected in the electoral district, interacted with the party of the elected candidate. Column (1) employs the sample of municipalities in a 10 km proximity to the railway plans, while column (2) employs the sample of municipalities a 5 km proximity to the railway plans. The sample is restricted to municipalities belonging to districts where the candidate was elected with a margin of victory below 5%. Municipality-level controls include: log population density in 1871, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation and presence of a coast. All specifications employ province and district fixed effects. Standard errors adjusted for spatial correlation, based on a 20 km window, in parenthesis. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.10. Implementation of Baccarini Law's Railway Projects and 1882-86 Municipality-Level Electoral Outcomes: Controls' Coefficients

	(1)	(2)
Dependent variable:	*Municipality obtains rail during legislature*	
Proximity to projects:	10 km	5 km
Marginality percentile:	$\leq$ 20th	$\leq$ 20th
Last winner's share in municipality	0.0494	0.0371
× Opposition party	(0.0318)	(0.0417)
Last winner's share in municipality	0.0378***	0.0346**
× Government party	(0.0121)	(0.0163)
Log population density 1881	0.0282	0.0293
	(0.0237)	(0.0374)
Post office in 1871	-0.0087	0.0169
	(0.0212)	(0.0298)
Telegraph office in 1871	0.0943**	0.0774
	(0.0378)	(0.0529)
Railway station in 1871	-0.1296***	-0.1480*
	(0.0487)	(0.0807)
Sea port in 1871	-0.1600	-0.3051**
	(0.1272)	(0.1396)
Total secondary schools and libraries	-0.0168	-0.0193
	(0.0187)	(0.0220)
Terrain ruggedness	0.0960	-0.0542
	(0.1359)	(0.1846)
Log land area	-0.0176	-0.0256
	(0.0193)	(0.0289)
Coast	0.2151***	0.2178***
	(0.0398)	(0.0507)
Elevation	-0.3301**	-0.2799
	(0.1681)	(0.2035)
Market access	-0.7004***	-0.8272***
	(0.2073)	(0.2441)
Adjusted R^2	0.368	0.386
Observations	1,052	671
Number of electoral district	49	46
Municipality-level controls	✓	✓
Province FE	✓	✓
District FE	✓	✓

Notes: The table reports reports OLS estimates. The unit of observation is a municipality. The dependent variable indicates the passage of a railway on the municipality territory, which construction was contracted out during the 1882 and 1886 legislatures. The main explanatory variables indicate the share of total votes obtained in the municipality by the candidate elected in the last seat of the district, interacted with the party of the elected candidate. Column (1) employs the sample of municipalities in a 10 km proximity to the railway plans, while column (2) employs the sample of municipalities a 5 km proximity to the railway plans. The sample is restricted to municipalities in districts belonging to the top quintile of the marginality measure. Municipality-level controls include: log population density in 1881, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation and presence of a coast. All specifications employ province and district fixed effects. Standard errors adjusted for spatial correlation, based on a 20 km window, in parenthesis. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.11. Implementation of Baccarini Law's Railway Projects and 1880 Municipality-Level Electoral Outcomes: Exclude Large Municipalities

	(1)	(2)
Dependent variable:	*Municipality obtains rail during legislature*	
Proximity to projects:	10 km	5 km
Margin of victory:	$\leq$ 5%	$\leq$ 5%
	Panel A	
Winner share in municipality	0.0311**	0.0286
	(0.0152)	(0.0199)
Adjusted R^2	0.358	0.420
	Panel B	
Winner share in municipality	-0.0147	-0.0371
$\times$ Opposition-party win in district	(0.0239)	(0.0319)
Winner share in municipality	0.0507***	0.0516**
$\times$ Government-party win in district	(0.0176)	(0.0223)
Adjusted R^2	0.363	0.426
Observations	571	383
Number of electoral district	75	61
Municipality-level controls	✓	✓
Province FE	✓	✓
District FE	✓	✓

Notes: The table reports reports OLS estimates. The unit of observation is a municipality. The dependent variable indicates the passage of a railway on the municipality territory, which construction was contracted out during the 1880 legislature. In Panel A the explanatory variable indicates the share of votes obtained in the municipality by the candidate elected in the electoral district. In Panel B the share of votes obtained in the municipality is interacted with the party of the elected candidate. Column (1) employs the sample of municipalities in a 10 km proximity to the railway plans, while column (2) employs the sample of municipalities a 5 km proximity to the railway plans. The sample is restricted to municipalities belonging to districts where the candidate was elected with a margin of victory below 5%. Municipalities above the 95th percentile of population are excluded. Municipality-level controls include: log population density in 1871, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation and presence of a coast. All specifications employ province and district fixed effects. Standard errors adjusted for spatial correlation, based on a 20 km window, in parenthesis. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.12. Implementation of Baccarini Law's Railway Projects and 1882-1886 Municipality-Level Electoral Outcomes: Exclude Large Municipalities

	(1)	(2)
Dependent variable:	*Municipality obtains rail during legislature*	
Proximity to projects:	10 km	5 km
Marginality percentile:	$\leq$ 20th	$\leq$ 20th
	Panel A	
Last winner's share in municipality	0.0367***	0.0350**
	(0.0121)	(0.0158)
Adjusted R^2	0.317	0.329
	Panel B	
Last winner's share in municipality	0.0374	0.0167
$\times$ Opposition party	(0.0289)	(0.0394)
Last winner's share in municipality	0.0365***	0.0378**
$\times$ Government party	(0.0129)	(0.0167)
Adjusted R^2	0.317	0.329
Observations	972	606
Number of electoral district	49	46
Municipality-level controls	✓	✓
Province FE	✓	✓
District FE	✓	✓

Notes: The table reports reports OLS estimates. The unit of observation is a municipality. The dependent variable indicates the passage of a railway on the municipality territory, which construction was contracted out during the 1882 and 1886 legislatures. Panel A employs the share of total votes obtained in the municipality by the candidate elected in the last seat of the district. In Panel B the share of votes obtained in the municipality is interacted with the party of the elected candidate. Column (1) employs the sample of municipalities in a 10 km proximity to the railway plans, while column (2) employs the sample of municipalities a 5 km proximity to the railway plans. The sample is restricted to municipalities in districts belonging to the top quintile of the marginality measure. Municipalities above the 95th percentile of population are excluded. Municipality-level controls include: log population density in 1881, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation and presence of a coast. All specifications employ province and district fixed effects. Standard errors adjusted for spatial correlation, based on a 20 km window, in parenthesis. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.13. Implementation of Baccarini Law's Railway Projects and 1880 Municipality-Level Electoral Outcomes: Railway Line FE

	(1)	(2)
Dependent variable:	*Municipality obtains rail during legislature*	
Proximity to projects:	10 km	5 km
Margin of victory:	$\leq$ 5%	$\leq$ 5%
	Panel A	
Winner share in municipality	0.0335**	0.0258
	(0.0141)	(0.0184)
Adjusted R^2	0.292	0.341
	Panel B	
Winner share in municipality	-0.0140	-0.0607
$\times$ Opposition-party win in district	(0.0302)	(0.0456)
Winner share in municipality	0.0463***	0.0441**
$\times$ Government-party win in district	(0.0159)	(0.0199)
Adjusted R^2	0.295	0.347
Observations	619	420
Number of electoral district	75	61
Municipality-level controls	✓	✓
Railway line FE	✓	✓
District FE	✓	✓

Notes: The table reports reports OLS estimates. The unit of observation is a municipality. The dependent variable indicates the passage of a railway on the municipality territory, which construction was contracted out during the 1880 legislature. In Panel A the explanatory variable indicates the share of votes obtained in the municipality by the candidate elected in the electoral district. In Panel B the share of votes obtained in the municipality is interacted with the party of the elected candidate. Column (1) employs the sample of municipalities in a 10 km proximity to the railway plans, while column (2) employs the sample of municipalities a 5 km proximity to the railway plans. The sample is restricted to municipalities belonging to districts where the candidate was elected with a margin of victory below 5%. Municipality-level controls include: log population density in 1871, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation and presence of a coast. All specifications employ railway line and district fixed effects. Standard errors adjusted for spatial correlation, based on a 20 km window, in parenthesis. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.14. Implementation of Baccarini Law's Railway Projects and 1882-1886 Municipality-Level Electoral Outcomes: Railway Line FE

	(1)	(2)
Dependent variable:	*Municipality obtains rail during legislature*	
Proximity to projects:	10 km	5 km
Marginality percentile:	$\leq$ 20th	$\leq$ 20th
	Panel A	
Last winner's share in municipality	0.0282***	0.0250*
	(0.0095)	(0.0128)
Adjusted R^2	0.554	0.565
	Panel B	
Last winner's share in municipality	0.0396	0.0252
$\times$ Opposition party	(0.0244)	(0.0336)
Last winner's share in municipality	0.0263***	0.0250*
$\times$ Government party	(0.0101)	(0.0135)
Adjusted R^2	0.553	0.564
Observations	1,052	671
Number of electoral district	49	46
Municipality-level controls	✓	✓
Railway line FE	✓	✓
District FE	✓	✓

Notes: The table reports reports OLS estimates. The unit of observation is a municipality. The dependent variable indicates the passage of a railway on the municipality territory, which construction was contracted out during the 1882 and 1886 legislatures. Panel A employs the share of total votes obtained in the municipality by the candidate elected in the last seat of the district. In Panel B the share of votes obtained in the municipality is interacted with the party of the elected candidate. Column (1) employs the sample of municipalities in a 10 km proximity to the railway plans, while column (2) employs the sample of municipalities a 5 km proximity to the railway plans. The sample is restricted to municipalities in districts belonging to the top quintile of the marginality measure. Municipality-level controls include: log population density in 1881, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation and presence of a coast. All specifications employ railway line and district fixed effects. Standard errors adjusted for spatial correlation, based on a 20 km window, in parenthesis. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.15. Implementation of Baccarini Law's Railway Projects and 1880 Municipality-Level Electoral Outcomes: Exclude Incumbents

	(1)	(2)
Dependent variable:	*Municipality obtains rail during legislature*	
Proximity to projects:	10 km	5 km
Margin of victory:	$\leq$ 5%	$\leq$ 5%
	Panel A	
Winner share in municipality	0.0330**	0.0300
	(0.0156)	(0.0193)
Adjusted R^2	0.281	0.345
	Panel B	
Winner share in municipality	-0.0199	-0.0611
$\times$ Opposition-party win in district	(0.0253)	(0.0382)
Winner share in municipality	0.0543***	0.0577***
$\times$ Government-party win in district	(0.0173)	(0.0205)
Adjusted R^2	0.287	0.354
Observations	547	373
Number of electoral district	60	48
Municipality-level controls	✓	✓
Province FE	✓	✓
District FE	✓	✓

Notes: The table reports reports OLS estimates. The unit of observation is a municipality. The dependent variable indicates the passage of a railway on the municipality territory, which construction was contracted out during the 1880 legislature. In Panel A the explanatory variable indicates the share of votes obtained in the municipality by the candidate elected in the electoral district. In Panel B the share of votes obtained in the municipality is interacted with the party of the elected candidate. Column (1) employs the sample of municipalities in a 10 km proximity to the railway plans, while column (2) employs the sample of municipalities a 5 km proximity to the railway plans. The sample is restricted to municipalities belonging to districts where the candidate was elected with a margin of victory below 5%. Municipalities belonging to districts carried by an incumbent are excluded. Municipality-level controls include: log population density in 1871, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation and presence of a coast. All specifications employ province and district fixed effects. Standard errors adjusted for spatial correlation, based on a 20 km window, in parenthesis. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.16. Implementation of Baccarini Law's Railway Projects and 1882-1886 Municipality-Level Electoral Outcomes: Exclude Incumbents

	(1)	(2)
Dependent variable:	*Municipality obtains rail during legislature*	
Proximity to projects:	10 km	5 km
Marginality percentile:	$\leq$ 20th	$\leq$ 20th
	Panel A	
Last winner's share in municipality	0.0391***	0.0345**
	(0.0116)	(0.0157)
Adjusted R^2	0.371	0.386
	Panel B	
Last winner's share in municipality	0.0492	0.0345
$\times$ Opposition party	(0.0318)	(0.0417)
Last winner's share in municipality	0.0374***	0.0345**
$\times$ Government party	(0.0120)	(0.0163)
Adjusted R^2	0.370	0.385
Observations	980	625
Number of electoral district	47	44
Municipality-level controls	✓	✓
Province FE	✓	✓
District FE	✓	✓

Notes: The table reports reports OLS estimates. The unit of observation is a municipality. The dependent variable indicates the passage of a railway on the municipality territory, which construction was contracted out during the 1882 and 1886 legislatures. Panel A employs the share of total votes obtained in the municipality by the candidate elected in the last seat of the district. In Panel B the share of votes obtained in the municipality is interacted with the party of the elected candidate. Column (1) employs the sample of municipalities in a 10 km proximity to the railway plans, while column (2) employs the sample of municipalities a 5 km proximity to the railway plans. The sample is restricted to municipalities in districts belonging to the top quintile of the marginality measure. Municipalities belonging to districts where the last seat is carried by an incumbent are excluded. Municipality-level controls include: log population density in 1881, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation and presence of a coast. All specifications employ province and district fixed effects. Standard errors adjusted for spatial correlation, based on a 20 km window, in parenthesis. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.17. Implementation of Baccarini Law's Railway Projects and 1880 Municipality-Level Electoral Outcomes: Contemporary Controls

	(1)	(2)
Dependent variable:	*Municipality obtains rail during legislature*	
Proximity to projects:	10 km	5 km
Margin of victory:	$\leq$ 5%	$\leq$ 5%
	Panel A	
Winner share in municipality	0.0311**	0.0268
	(0.0141)	(0.0183)
Adjusted R^2	0.308	0.351
	Panel B	
Winner share in municipality	-0.0171	-0.0555
× Opposition-party win in district	(0.0258)	(0.0366)
Winner share in municipality	0.0503***	0.0535***
× Government-party win in district	(0.0161)	(0.0199)
Adjusted R^2	0.313	0.361
Observations	597	404
Number of electoral district	75	61
Municipality-level controls	✓	✓
Province FE	✓	✓
District FE	✓	✓

Notes: The table reports reports OLS estimates. The unit of observation is a municipality. The dependent variable indicates the passage of a railway on the municipality territory, which construction was contracted out during the 1880 legislature. In Panel A the explanatory variable indicates the share of votes obtained in the municipality by the candidate elected in the electoral district. In Panel B the share of votes obtained in the municipality is interacted with the party of the elected candidate. Column (1) employs the sample of municipalities in a 10 km proximity to the railway plans, while column (2) employs the sample of municipalities a 5 km proximity to the railway plans. The sample is restricted to municipalities belonging to districts where the candidate was elected with a margin of victory below 5%. Municipality-level controls include: log population density in 1871, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation, presence of a coast, quality and quantity of water available, percentage of roads with sewage system, percentage of houses with toilets, per capita number of doctors, per capita number of farmacies, per capita number of hospital beds, number of years with registered cases of cholera, per capita municipal budget measures on tax revenue and expenditures. All specifications employ province and district fixed effects. Standard errors adjusted for spatial correlation, based on a 20 km window, in parenthesis. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.18. Implementation of Baccarini Law's Railway Projects and 1882-1886 Municipality-Level Electoral Outcomes: Contemporary Controls

	(1)	(2)
Dependent variable:	*Municipality obtains rail during legislature*	
Proximity to projects:	10 km	5 km
Marginality percentile:	$\leq$ 20th	$\leq$ 20th
	Panel A	
Last winner's share in municipality	0.0397***	0.0263
	(0.0134)	(0.0164)
Adjusted R^2	0.264	0.330
	Panel B	
Last winner's share in municipality	0.0315	0.0017
× Opposition party	(0.0322)	(0.0385)
Last winner's share in municipality	0.0412***	0.0308*
× Government party	(0.0141)	(0.0174)
Adjusted R^2	0.263	0.329
Observations	980	611
Number of electoral district	49	46
Municipality-level controls	✓	✓
Province FE	✓	✓
District FE	✓	✓

Notes: The table reports reports OLS estimates. The unit of observation is a municipality. The dependent variable indicates the passage of a railway on the municipality territory, which construction was contracted out during the 1882 and 1886 legislatures. Panel A employs the share of total votes obtained in the municipality by the candidate elected in the last seat of the district. In Panel B the share of votes obtained in the municipality is interacted with the party of the elected candidate. Column (1) employs the sample of municipalities in a 10 km proximity to the railway plans, while column (2) employs the sample of municipalities a 5 km proximity to the railway plans. The sample is restricted to municipalities in districts belonging to the top quintile of the marginality measure. Municipality-level controls include: log population density in 1881, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation, presence of a coast, quality and quantity of water available, percentage of roads with sewage system, percentage of houses with toilets, per capita number of doctors, per capita number of farmacies, per capita number of hospital beds, number of years with registered cases of cholera, per capita municipal budget measures on tax revenue and expenditures. All specifications employ province and district fixed effects. Standard errors adjusted for spatial correlation, based on a 20 km window, in parenthesis. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.19. Implementation of Baccarini Law's Railway Projects and 1880 Municipality-Level Electoral Outcomes: Different Clustering Methods

	(1)	(2)
Dependent variable:	*Municipality obtains rail during legislature*	
Proximity to projects:	10 km	5 km
Margin of victory:	$\leq$ 5%	$\leq$ 5%
	Panel A	
Winner share in municipality	0.0292	0.0255
District cluster s.e.	(0.0173)*	(0.0215)
District-Line double-cluster s.e.	[0.0136]**	[0.0152]*
Spatially-adjusted s.e. - 50km	{0.0149}**	{0.0179}
Adjusted R^2	0.309	0.354
	Panel B	
Winner share in municipality $\times$ Opposition-party win	-0.0179	-0.0505
District cluster s.e.	(0.0271)	(0.0390)
District-Line double-cluster s.e.	[0.0259]	[0.0380]
Spatially-adjusted s.e. - 50km	{0.0249}	{0.0359}
Winner share in municipality $\times$ Government-party win	0.0470	0.0486
District cluster s.e.	(0.0194)**	(0.0236)**
District-Line double-cluster s.e.	[0.0174]***	[0.0195]**
Spatially-adjusted s.e. - 50km	{0.0171}***	{0.0197}**
Adjusted R^2	0.313	0.362
Observations	619	420
Number of electoral district	75	61
Municipality-level controls	✓	✓
Province FE	✓	✓
District FE	✓	✓

Notes: The table reports reports OLS estimates. The unit of observation is a municipality. The dependent variable indicates the passage of a railway on the municipality territory, which construction was contracted out during the 1880 legislature. In Panel A the explanatory variable indicates the share of votes obtained in the municipality by the candidate elected in the electoral district. In Panel B the share of votes obtained in the municipality is interacted with the party of the elected candidate. Column (1) employs the sample of municipalities in a 10 km proximity to the railway plans, while column (2) employs the sample of municipalities a 5 km proximity to the railway plans. The sample is restricted to municipalities belonging to districts where the candidate was elected with a margin of victory below 5%. Municipality-level controls include: log population density in 1871, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation and presence of a coast. All specifications employ province and district fixed effects. In parenthesis are robust standard errors clustered by electoral district. In square brackets are robust standard errors double-clustered by electoral district and railway line. In curly brackets are standard errors adjusted for spatial correlation based on a 50 km window. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.20. Implementation of Baccarini Law's Railway Projects and 1882-1886 Municipality-Level Electoral Outcomes: Different Clustering Methods

	(1)	(2)
Dependent variable:	*Municipality obtains rail during legislature*	
Proximity to projects:	10 km	5 km
Marginality percentile:	$\leq$ 20th	$\leq$ 20th
	Panel A	
Last winner's share in municipality	0.0395	0.0349
District cluster s.e.	(0.0107)***	(0.0164)**
District-Line double-cluster s.e.	[0.0103]***	[0.0124]***
Spatially-adjusted s.e. - 50km	{0.0115}***	{0.0152}**
Adjusted R^2	0.368	0.387
	Panel B	
Last winner's share in municipality $\times$ Opposition party	0.0494	0.0371
District cluster s.e.	(0.0385)	(0.0484)
District-Line double-cluster s.e.	[0.0433]	[0.0521]
Spatially-adjusted s.e. - 50km	{0.0372}	{0.0465}
Last winner's share in municipality $\times$ Government party	0.0378	0.0346
District cluster s.e.	(0.0100)***	(0.0166)**
District-Line double-cluster s.e.	[0.0078]***	[0.0103]***
Spatially-adjusted s.e. - 50km	{0.0114}***	{0.0153}**
Adjusted R^2	0.368	0.386
Observations	1,052	671
Number of electoral district	49	46
Municipality-level controls	✓	✓
Province FE	✓	✓
District FE	✓	✓

Notes: The table reports reports OLS estimates. The unit of observation is a municipality. The dependent variable indicates the passage of a railway on the municipality territory, which construction was contracted out during the 1882 and 1886 legislatures. Panel A employs the share of total votes obtained in the municipality by the candidate elected in the last seat of the district. In Panel B the share of votes obtained in the municipality is interacted with the party of the elected candidate. Column (1) employs the sample of municipalities in a 10 km proximity to the railway plans, while column (2) employs the sample of municipalities a 5 km proximity to the railway plans. The sample is restricted to municipalities in districts belonging to the top quintile of the marginality measure. Municipality-level controls include: log population density in 1881, presence of post office, telegraph, port, railway or railway station on the territory in 1871, total number of secondary schools and libraries in 1863, market access, terrain ruggedness, agricultural suitability, log land area, elevation and presence of a coast. All specifications employ province and district fixed effects. In parenthesis are robust standard errors clustered by electoral district. In square brackets are robust standard errors double-clustered by electoral district and railway line. In curly brackets are standard errors adjusted for spatial correlation based on a 50 km window. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.21. Long-Run Impact of Baccarini Law's Railway Expansion on Municipalities' Population: IV Estimates by Period

	Dependent variable: *Log population*			
	1901-1991	1901-1936	1936-1951	1951-1991
	(1)	(2)	(3)	(4)
Railway access	0.2282**	0.1190*	0.1022	0.3380**
	(0.1156)	(0.0707)	(0.0742)	(0.1684)
Observations	5,370	2,685	1,074	2,685
Adjusted R^2	0.911	0.962	0.951	0.891
F-stat	11.43	11.34	11.07	11.34
Municipality-level controls × year	✓	✓	✓	✓
Province × year FE	✓	✓	✓	✓
District × year FE	✓	✓	✓	✓

Notes: Table presents panel estimates regressing population on railway access. The unit of observation is a municipality. The dependent variable is the log of municipality population for periods between 1901 and 1991 (census years are 1901, 1911, 1921, 1931, 1936, 1951, 1961, 1971, 1981 and 1991). The explanatory variable is a dummy denoting municipalities reached by one of the Baccarini Law railways. Estimates instrument railway access with the share of votes obtained in the municipality by the candidate elected in the electoral district if construction of the railway was authorized in the 1880 parliament, and with the share of votes obtained in the municipality by the candidate elected in the last seat of the district if construction of the railway was authorized in the 1882 or 1886 parliaments. All specification employ municipality-level pre-determined controls, interacted with year, province-year fixed effects and district-year fixed effects. All specification limit the sample to municipalities in a 10 km proximity of the railways and that belonged to marginal electoral districts when the railway construction was authorized. Municipality-level controls include: log population in 1871, presence of post office, telegraph, port, railway or railway station on the territory, total number of secondary schools and libraries, terrain ruggedness, agricultural suitability, log land area, elevation, presence of a coast, quality and quantity of water available, percentage of roads with sewage system, percentage of houses with toilets, per capita number of doctors, per capita number of farmacies, per capita number of hospital beds, number of years with registered cases of cholera, per capita municipal budget measures on tax revenue and expenditures. Robust standard errors, clustered by circondario and year, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.22. Long-Run Impact of Baccarini Law's Railway Expansion on Municipalities' Population: IV-FE Estimates by Period

	Dependent variable: *Log population*			
	1901-1991	1901-1936	1936-1951	1951-1991
	(1)	(2)	(3)	(4)
Railway access × year	0.0471*	0.0042	-0.0279	0.1101***
	(0.0260)	(0.0181)	(0.0472)	(0.0424)
Observations	5,370	2,685	1,074	2,685
Adjusted R^2	0.521	0.561	0.160	0.362
F-stat	12.69	14.22	11.83	14.11
Municipality FE	✓	✓	✓	✓
Municipality-level controls × year	✓	✓	✓	✓
Province × year FE	✓	✓	✓	✓
District × year FE	✓	✓	✓	✓

Notes: Table presents panel estimates regressing population on railway access. The unit of observation is a municipality. The dependent variable is the log of municipality population for periods between 1901 and 1991 (census years are 1901, 1911, 1921, 1931, 1936, 1951, 1961, 1971, 1981 and 1991). The explanatory variable is a dummy denoting municipalities reached by one of the Baccarini Law railways, interacted with year. Estimates instrument railway access with the share of votes obtained in the municipality by the candidate elected in the electoral district if construction of the railway was authorized in the 1880 parliament, and with the share of votes obtained in the municipality by the candidate elected in the last seat of the district if construction of the railway was authorized in the 1882 or 1886 parliaments. All specification employ municipality-level pre-determined controls, interacted with year, municipality fixed effects, province-year fixed effects and district-year fixed effects. All specification limit the sample to municipalities in a 10 km proximity of the railways and that belonged to marginal electoral districts when the railway construction was authorized. Municipality-level controls include: log population in 1871, presence of post office, telegraph, port, railway or railway station on the territory, total number of secondary schools and libraries, terrain ruggedness, agricultural suitability, log land area, elevation, presence of a coast, quality and quantity of water available, percentage of roads with sewage system, percentage of houses with toilets, per capita number of doctors, per capita number of farmacies, per capita number of hospital beds, number of years with registered cases of cholera, per capita municipal budget measures on tax revenue and expenditures. Robust standard errors, clustered by circondario and year, are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

CHAPTER III

TWO SIDES OF THE SAME COIN: CO-EVOLUTION OF KIN TIES AND INSTITUTIONS

1 Introduction

This paper builds on a growing body of research in economics studying the relationship between the fundamental causes of growth, by investigating the joint evolution of culture and institutions across the world and over time. While both culture and institutions are broad concepts with differing definitions, here I refer to culture as, following the definition adopted by Guiso, Sapienza, and Zingales (2006), "those customary beliefs and values that ethnic, religious, and social groups transmit fairly unchanged from generation to generation", whereas I refer to institutions as, following the definition adopted by North (1990), "the humanly devised constraints that structure human interactions... made up of formal constraints (rules, laws, constitutions)... and their enforcement characteristics." In contrast to the literature that tries to gauge the impact of cultural traits while isolating the role of formal institutions, or vice-versa, an emerging group of papers has been advocating that these two forces interact through a two-way relationship (Tabellini, 2010; Bisin and Verdier, 2017; Guiso, Sapienza, and Zingales, 2016; Alesina and Giuliano, 2015). Take for example the following question: What is the scope of situations in which individuals are willing to cooperate? It is easy to see how answers to this kind of question could be attributed to either, or both, cultural and institutional factors. While there has been a proliferation of theoretical studies on the link between culture and institutions, the empirical evidence on this link is scarce. With the present study, I aim to move the literature forward by focusing on the interplay between formal institutions and a cultural feature that has received a great deal of attention, particularly so in economics in recent years, that is family and kin ties.

In this paper, I first explore the common geographical forces that have historically driven the evolution of kin ties and state institutions, based on the assumption that the scope of socio-economic relations was historically reflected in an equilibrium balance between kin ties and state institutions. I establish empirically that climatic conditions, specifically climate intertemporal volatility and spatial variability, acted as an exogenous force that shaped the prevailing economic relations and, as a consequence, both the strength of kin ties and the development of state institutions in historical times. Subsequently, I investigate the long-run persistence of historical differences in the strength of kin ties and the development of state institutions. I advance the hypothesis and establish empirically that the institutional quality of modern countries persistently reflects historical differences in the importance of kin ties and state institutions, even when the factors that lead to this divergence are no longer relevant.

More in detail, in the first part of the paper I employ data from Murdock's (1967) *Ethnographic Atlas*, providing measures of cultural, economic and institutional characteristics of over 1300 societies around the world, measured before industrialisation or colonisation, and combine it with weather data capturing climatic intertemporal volatility and spatial variability. The analysis allows me to establish that exogenous changes, deriving from climatic conditions, to the prevailing economic relations shifted the historical equilibrium between kin ties and state institutions. To capture the strength of kin ties I exploit, following Enke (2019), a group of measures from the *Ethnographic Atlas* that the anthropological literature has indicated as descriptive of kinship tightness, that is: extended vs. nuclear families, post-marriage residence with parents, unilineal lineage tracing and presence of localised clans. Combining these different measures I compute an index capturing the degree of kinship tightness. To gauge the development of state institutions, I employ an indicator of jurisdictional hierarchies beyond the local community, a measure that is employed as the standard to capture institutional sophistication and state development in the *Ethnographic Atlas* (Gennaioli and Rainer, 2006, 2007; Michalopoulos and Papaioannou, 2013; Nunn, 2008).

To proxy for the historical impact of climatic shocks, I construct, following the methodology of Durante (2009), measures of climate intertemporal volatility and spatial variability based on 20th century data from the *Climatic Research Unit (CRU)* database. The *CRU* data, derived from meteorological stations records, comes at a fine spatial resolution of 0.5°× 0.5°, with 0.5° being approximately 55 km at the equator, and at a high temporal frequency, consisting of monthly observations of temperature and precipitation over the period 1900-2017, allowing for the construction of precise and granular measures of climatic conditions. Relating the *CRU* data with paleoclimatic reconstructions of climate in the past centuries confirms that modern and historical volatilities of climate are highly correlated, alleviating potential concerns about specific changes in climatic conditions over the course of human history that would prevent from using the high-resolution *CRU* data as a proxy for historical climate shocks.

Through the analysis, I establish historical societies being characterised by weaker kin ties and more developed state institutions when subject to a climate featuring (i) higher intertemporal volatility and (ii) higher spatial variability. Therefore, consistently with the hypothesised mechanism, weather-related shocks which can be smoothed over space historically caused individuals to lower the degree of reliance on their kin and expand the geographical and social radius of their socio-economic relations, thus weakening kin ties while strengthening the need for formal state institutions. The established impact of intertemporal volatility and spatial variability is consistent both when considering aggregate measures of climate and when exploring the role of the temperature and precipitation components separately. Findings are robust to the inclusion of regional fixed effects and accounting for the potentially confounding role of a range of geographical, climatic and ethnographic characteristics. Further analysis suggests the established effect of climatic conditions to be stronger for agriculture-reliant societies. In fact, consistent with the idea of agriculturalist being particularly sensitive to the impact of weather shocks (Diamond, 1987; Harari, 2015), I determine that societies that relied more heavily on agriculture as a form of subsistence and that practised more intensive forms of agriculture show an impact of higher climate intertemporal volatility that goes beyond the average effect across the sample, resulting in even weaker kin ties and stronger state institutions. Finally, I ascertain that results hold after a variety of robustness checks, namely accounting for potential cross-equation correlation in error terms, computation of variables based on geographical areas of different size and spatial correlation in the estimated impact of climatic conditions.

I then explore the long-run persistence of historical kinship tightness and state development, based on the hypothesis that modern countries persistently reflect such historical differences even after their geographical determinants are no longer relevant. Specifically, I employ country-level indicators capturing institutional quality from the 19th century to the present and link the population of modern countries with the characteristics of their historical ancestors, to establish that countries whose populations' ancestors featured more developed state institutions and weaker kin ties are associated with modern institutions of significantly higher quality.

More in detail, I employ data on the institutional quality of modern countries from the *V-Dem* project (Coppedge, Gerring, Knutsen, Lindberg, Skaaning, Teorell, Altman, Bernhard, Fish, Cornell, et al., 2018), providing yearly multidimensional measures of institutions dating as far back as 1800. The long-term structure of the data allows me to study not only differences in institutions across countries, but also the evolution of institutions across time. Thus, to explore long-run patterns in the institutional quality of modern countries, I construct measures capturing the average institutional quality across 50-year intervals between 1800 and 2000. Next, to bring the historical data from the *Ethnographic Atlas* to the modern country level, I employ Giuliano and Nunn's (2018) *Ancestral Characteristics of Modern Populations* dataset, which applies a language-based matching to construct population-weighted averages of the prevalence of characteristics from *Ethnographic Atlas* among the ancestors of people living in each country. Via this data, I therefore compute measures of the historical characteristics of modern countries' ancestors.

Through the analysis, I show that historically weaker kin ties and greater state development are associated with higher-quality institutions across the modern period. Results provide consistent patterns across different institutional indexes, namely rule of law, property rights, level of democracy and control of corruption, and whether the analysis pools the different time periods or employs them individually. Furthermore, results are robust to the inclusion of continent fixed effects and other potentially confounding historical, geographic and contemporary factors. Further checks indicate that findings hold even after controlling for initial GDP, and formal tests on the selection of unobservables suggest that omitted variables are not driving estimation results. Finally, regressions examining the joint impact of ancestral kinship tightness and ancestral state institutions reveal a significant and meaningful effect of both dimensions on modern institutional quality even when the two historical measures are employed together.

With this research I build on and contribute to different strands of the literature. Although employing varying conceptual definitions, the values and beliefs that are fostered and transmitted within the family and the strength of kin networks have been studied in relation with features such as "amoral familism" (Banfield, 1958), differences in political and institutional systems across the world (Todd, 1983; Greif, 1994), trust and cooperation (Alesina and Giuliano, 2014; Moscona, Nunn, and Robinson, 2017; Lowes, 2018) and a plethora of economic and non-economic outcomes both at the micro and macro level (Greif and Tabellini, 2017; Gorodnichenko and Roland, 2017; Alesina and Giuliano, 2011). Particularly relevant papers are Enke (2019), relating kin structure to differences in moral systems, and Schulz (2017), linking cousin marriage rates to large variations in democracy. With my paper, I advance this literature by providing novel evidence on the geographical forces that have historically shaped the importance of kin networks and the impact of kinship tightness on modern institutional quality. I also contribute to the recent empirical literature where climatic shocks are put in relation to persistent cultural features, such as loss aversion (Galor and Savitskiy, 2018), the importance of tradition (Giuliano and Nunn, 2017) and trust (Buggle and Durante, 2017), by providing direct evidence of the historical link between climatic intertemporal volatility and spatial variability and the strength of kin ties. Another relevant body of research to which I relate is the one studying the relationship between culture and institutions (Greif, 1994; Tabellini, 2008a; Tabellini, 2008b; Bisin and Verdier, 2017; Tabellini, 2010; Alesina and Giuliano, 2015), with my paper contributing to this literature by providing first cut empirical evidence supportive of kin ties and state institutions being part of an historical equilibrium reflecting the scope of economic relations. On a broader scale, this work contributes to the literature on the geographical origins of cultural and institutional features (Diamond, 1997; Acemoglu, Johnson, and Robinson, 2001, 2002; Engerman and Sokoloff, 2000; Galor and Özak, 2016; Michalopoulos, 2012; Ashraf and Michalopoulos, 2015; Bates, 1987; Fenske, 2014). My paper contributes to this literature by providing empirical evidence supportive of the interconnectedness of three fundamental causes of growth: geography, culture and institutions.

The remainder of the paper is organised as follows. Section 2 focuses on the historical analysis, discussing its estimation framework, data and results. Section 3 turns to the long-run analysis, discussing its estimation framework, data and results. Section 4 provides a summary and concluding remarks.

2 Historical Analysis

This section explores the common forces that have historically driven the evolution of both kin ties and state institutions.

2.1 Conceptual Framework

This first part of the paper is based on the assumption that the scope of socio-economic relations was historically reflected in an equilibrium balance between kin ties and state institutions. The underlying idea is that in historical times, by which I refer to the period prior to the first industrial revolution and the process of colonisation, societies where the prevalence of economic relations was inside the family or extended kin network would be characterised by cultures featuring strong ties to said family and kin, and weak formal institutions, since most relations could be sustained through informal mechanisms. Specularly, realities where a larger share of economic activity involved socially- and possibly geographically-distant relations would be characterised by weak ties to kin and, on the other hand, strong formal institutions regulating and sustaining repeated interaction with non-related individuals.

I thus study exogenous determinants of the prevailing economic relations that consequently shaped the equilibrium between kin ties and state institutions in historical times. More in detail, I focus on climate intertemporal volatility and spatial variability, capturing the temporal and spatial dimensions of weather shocks, as an exogenous force that shaped the prevailing economic relations and, as a consequence, both the strength of kin ties and the development of state institutions in historical times. Focusing on the pre-industrial or pre-colonial period allows me to exploit the role of exogenous climatic conditions in a time when the influence of geography still loomed large. Historically, the need to protect subsistence from individual shocks was a major determinant of the organisation of societies. Under this aspect, family and kin have always been effective in acting as a buffer. This came as a natural consequence to the low cost of coordination and monitoring inside the family and amongst individuals belonging to the same kin network. For the same reason, economic activity was prominently organised around this dimension in historical times. On the other hand, shocks of larger scale and intensity such as weather-related shocks often made it so that the family was unable to provide proper protection and relief against their economic consequences, due to the reduced dimension and spatial concentration of this kind of basic organisation. Expanding the radius of socio-economic relations outside of the family and kin network, for example by establishing large-scale cooperation or trade networks, was one available strategy to smooth the risk deriving from weather shocks. In fact, individuals part of a broader community might be affected by shocks in a different way, or even not be subjected to the same kind of shocks altogether. Nonetheless, expanding the geographical and social radius of relations, while providing potential gains, also introduced more severe cooperation challenges, due to the increased costs of coordination and monitoring. Such a situation therefore raises the need for formal institutional arrangements, sets of "rules", such as those coming with a centralised state, allowing of repeated interaction and cooperation outside of the family and kin network.

Based on this framework, I formulate the hypothesis that higher degrees of climatic intertemporal volatility and spatial variability were historically associated with weaker kin ties and more developed state institutions.

2.2 Empirical Model

To formally test for the hypothesis set out above, I estimate the following specification:

$$\left.\begin{array}{l} KT_i \\ SI_i \end{array}\right\} = \gamma_0 + \gamma_1\ C_i^V + \gamma_2\ C_i^S + \delta + X_i'\ \Phi + \varepsilon_i \tag{1}$$

where on the left-hand side I employ as dependent variables measures of the kinship tightness, KT_i, and state institutions, SI_i, of historical society i. C_i^V and C_i^S indicate intertemporal volatility and spatial variability, respectively, of climate characterising the area on which society i is based. δ indicates region fixed effects, with the regions being North America, Latin America and Caribbean, Sub-Saharan Africa, Middle East and North Africa, Europe and Central Asia, South Asia, East Asia and Pacific region. X_i is a series of controls capturing the economic and geographic characteristics of society i.

2.3 Data

In the following, I describe the data employed to estimate equation (1).

2.3.1 Ethnographic Data

I employ data from Murdock's (1967) *Ethnographic Atlas*, containing categorical measures of cultural, economic and institutional characteristics of 1265 historical societies from around the world. The *Ethnographic Atlas* is arguably the best source of data that allows for a rich cross-cultural comparison of historical societies on a global scale. This data is intended to cover the characteristics of societies at an idealised moment coinciding with the time prior to the first industrial revolution in Europe and to European contact, and therefore in some cases colonisation, for the rest of the world. In practice, it is mostly based on the works of ethnographers and anthropologists at various points in time, with the average time of observation being 1898. Nevertheless, even for the observations that were sampled in the 20th century, the data is meant to describe the characteristics and living conditions of societies as they were in historical times. To ensure comparability, I remove from the sample eight societies that were sampled before 1500, namely Ancient Egypt, Aryans, Babylonia, Romans, Icelander, Uzbeg, Khmer and Hebrews. A shortcoming of the *Ethnographic Atlas* is that it significantly under-represents Europeans. To partially overcome this shortcoming, I employ a recent version of the database, updated by Giuliano and Nunn (2018) to include 44 additional observations broadening coverage in Europe. The distribution across the globe of the societies sampled in the *Ethnographic Atlas* is shown in Figure 1, with each dot being based on the coordinates of the estimated centroid of their historical location.

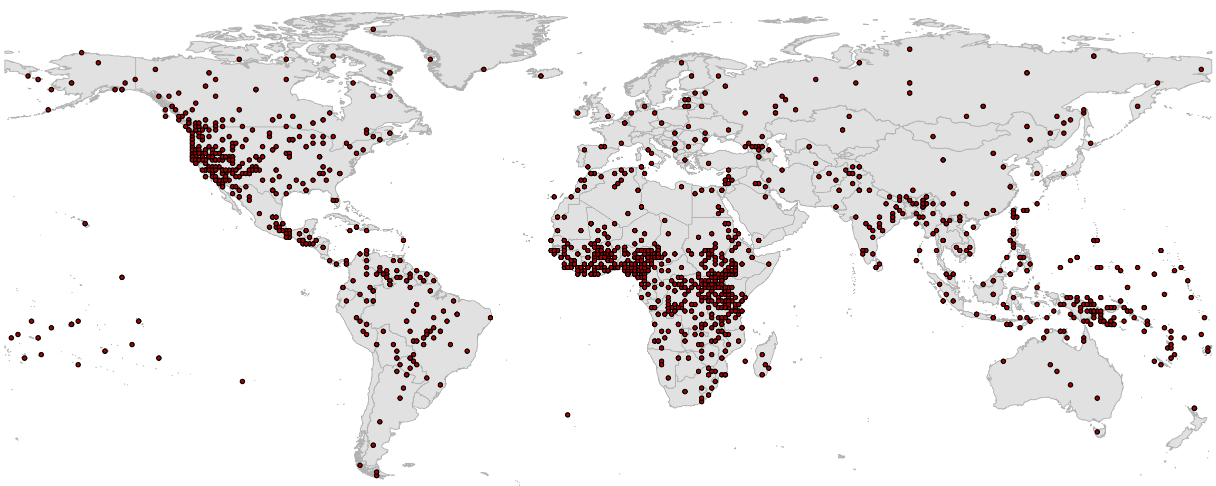
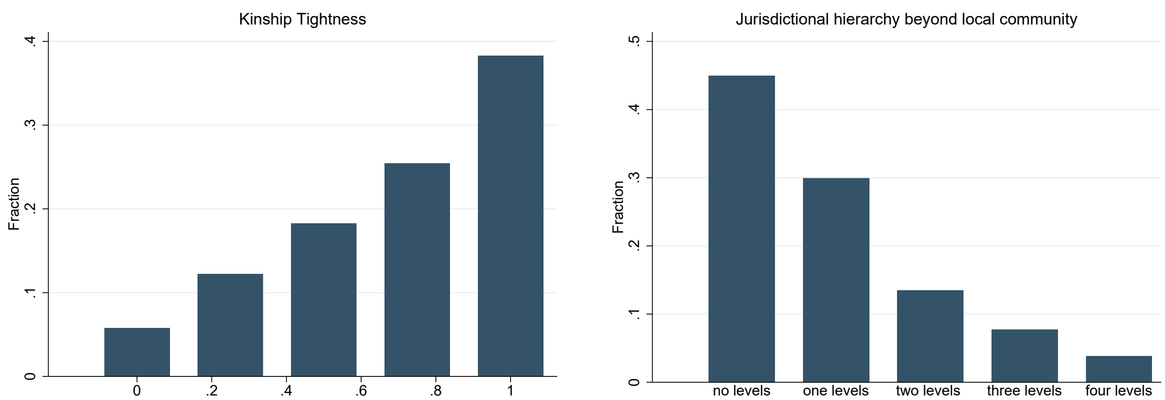


FIGURE 1. Location of historical societies in the *Ethnographic Atlas*

Based on the data from the *Ethnographic Atlas*, to capture the strength of kin ties I employ a kinship tightness index following Enke (2019). In particular, Enke's (2019) index is based on information on family structure and descent systems, which are features employed by anthropologists in order to describe kinship systems. Specifically, I employ dichotomous indicators capturing: extended vs. nuclear family system (based on variable v8 from the *EA*), post-marriage residence with parents (based on variable v11 from the *EA*), segmented communities and localised clans (based on variable v15 from the *EA*) and unilineal descent systems (based on variable v43 from the *EA*). The measure capturing kinship tightness is then computed through the average of the four illustrated indicators. The distribution of the kinship tightness index based on the *Ethnographic Atlas* is represented in Figure 2-(A), with higher values indicating stronger kin ties.

To capture state institutions, I employ the "Jurisdictional Hierarchy Beyond the Local Community" index (variable v33 from the *EA*). This ordered variable, ranging from 0 to 4, measures the number of jurisdictional hierarchies beyond the local community, and is the standard variable employed in the literature to capture institutional sophistication and state development in the *Ethnographic Atlas* (Gennaioli and Rainer, 2006, 2007; Michalopoulos and Papaioannou, 2013; Nunn, 2008). The distribution of the jurisdictional hierarchies index is represented in Figure 2-(B), with higher values indicating more developed state institutions.



(A) (B)

FIGURE 2. Kinship tightness (A) and jurisdictional hierarchies beyond the local community (B) from the *Ethnographic Atlas*.

Overall, the *Ethnographic Atlas* data provides a rich source of variation on kin ties and state institutions of historical societies across the globe. Beyond the distributions provided by Figure 2, some basic data patterns can help provide an idea of the relationship between these two characteristics of historical societies. In Appendix Figure A.1 I show through a bin scatter plot a negative relationship between kin ties and state institutions. Other telling examples can be provided by some of the observations in the *Ethnographic Atlas*. For comparison, the English, as sampled in 1600, score a value of 3 in terms of state institutions and the minimum value of 0 in terms of kinship tightness. Similarly, the Tagalog of the Philippines, sampled in 1900 score a value of 4 in terms of state institutions and a value of 0 in terms of kinship tightness. On the other end of the spectrum, the Western Apache Native Americans and the Makonde of Tanzania, sampled in 1870 and 1900, respectively, both score a value of 0 in terms of state institutions and the maximum value of 1 in terms of kinship tightness.

2.3.2 Climate Data

To capture the impact of exogenous climatic conditions, I exploit variation in temperature and precipitation. These two dimensions of climate have a considerable impact on all subsistence activities dependent on natural resources, with their role being particular relevant in historical times during which subsistence consumption was very close to being a binding constraint. Specifically, I proxy for historical climatic conditions by employing data from the CRU TS v.4.02 database constructed by the *Climatic Research Unit* of the University of East Anglia (Harris, Jones, Osborn, and Lister, 2014). This dataset is based on high-frequency observation from meteorological stations and provides monthly measures of temperature and precipitation over the period 1900-2017 at a 0.5°× 0.5° resolution.[1] Following the methodology of Durante (2009), I construct measures of intertemporal volatility and spatial variability for both temperature and precipitation, based on the raw *CRU* monthly data.

Relating to the construction of the intertemporal volatility measure, the starting point is $x_{i,m,y}$, the observation of x (temperature, precipitation) in cell i in month m (1,12) in year y (1900,2017). For each month m, I compute $\sigma_{i,m} = [\frac{1}{Y} \sum_{y=1900}^{2017} (x_{i,m,y} - \bar{x}_{i,m})^2]^{\frac{1}{2}}$, the standard deviation of $x_{i,m,y}$ over all years, capturing the month-specific variability of x in cell i. I then average the month-specific volatilities over the twelve months to obtain $\sigma_i = \frac{1}{12} \sum_{m=1}^{12} \sigma_{i,m}$, the overall year-to-year variability of x in cell i. Figure 3 shows the global distribution of temperature and precipitation intertemporal volatility, as captured by σ_i.

To compute the spatial dimension of climatic fluctuations, I consider the covariance between climate in a grid cell and climate in a set of neighbouring cells. More specifically, for each grid cell i I identify a set of J neighbours defined as those cells sharing a border or vertex with cell i, such that each cell can have up to eight neighbours. Again starting from $x_{i,m,y}$, the observation of x (temperature, precipitation) in cell i in month m (1,12) in year y (1900,2017), I first compute $\Delta_{i,m,y} = (x_{i,m,y} - \bar{x}_{i,m})$, the deviation of $x_{i,m,y}$ from its monthly mean $\bar{x}_{i,m}$. Then, for each ij pair of cells I compute $\rho_{ij} = corr(\Delta_{i,m,y}, \Delta_{j,m,y})$, the correlation between monthly deviations in i and j across all months and years. Finally, I aggregate the various ρ_{ij} into $\rho_i = \frac{1}{J} \sum_{j=1}^{J} \rho_{ij}$, providing a unique measure of spatial correlation for variable x in cell i. Figure 4 provides an illustration of the spatial correlation of temperature and precipitation across the globe, as captured by ρ_i. Based on this last measure, I compute spatial variability as the reverse of spatial correlation, i.e. $\upsilon_i = -\rho_i$.

To employ them in the analysis, the next step is to take the intertemporal volatility σ and spatial variability υ measures computed at the grid cell level and aggregate them at the

[1] Although the size of a degree varies based on latitude, 0.5° corresponds to approximately 55 kilometres at the equator.

historical society level. To do this, I employ the interior centroid of the society's location, as indicated in the *Ethnographic Atlas* and shown in Figure 1. Using these coordinates, I calculate the average values of σ and υ in a 100 km radius of the centroid, allowing me to obtain measures of intertemporal volatility and spatial variability specific to the area that was inhabited by society.[2] Finally, I compute aggregate measures of climatic intertemporal volatility and spatial variability based on averages of the individual temperature and precipitation variables.

While the high resolution data of the *Climatic Research Unit* allows me to obtain very precise measures of local variation in climate, as previously mentioned the 20th century data is employed as a proxy for historical climatic conditions. Therefore, a potential concern might arise with respect to changes in climate over the course of human history making the 20th century data a bad proxy for historical climate. To confirm the appropriateness of this proxy, I employ paleoclimatic data from Mann, Zhang, Rutherford, Bradley, Hughes, Shindell, Ammann, Faluvegi, and Ni (2009), measuring temperature anomalies from the 5th century to the present across the globe, to construct a measure a historical climatic volatility.[3] As shown through binscatter plots in Figure B.2, volatility in the 20th century is highly correlated with the corresponding volatility in the past fourteen centuries, with correlation coefficients between 0.5 and 0.9. While the Mann, Zhang, Rutherford, Bradley, Hughes, Shindell, Ammann, Faluvegi, and Ni (2009) data provides actual measures of historical climate, it has various shortcomings. First of all, this data is much more coarse, with a $5° \times 5°$ resolution, i.e. with grid cells 10 times bigger than the *CRU* data. Furthermore, the data only comes with annual frequency, with various years being imputed. Finally, the measures are only for temperature. These various aspects therefore make it difficult to employ this data to construct precise measures of local climatic differences. Thus, based on the assumption that historical and modern climatic volatilities are similar, supported by the evidence provided above, I will employ the *CRU* data as a proxy for historical climatic conditions.

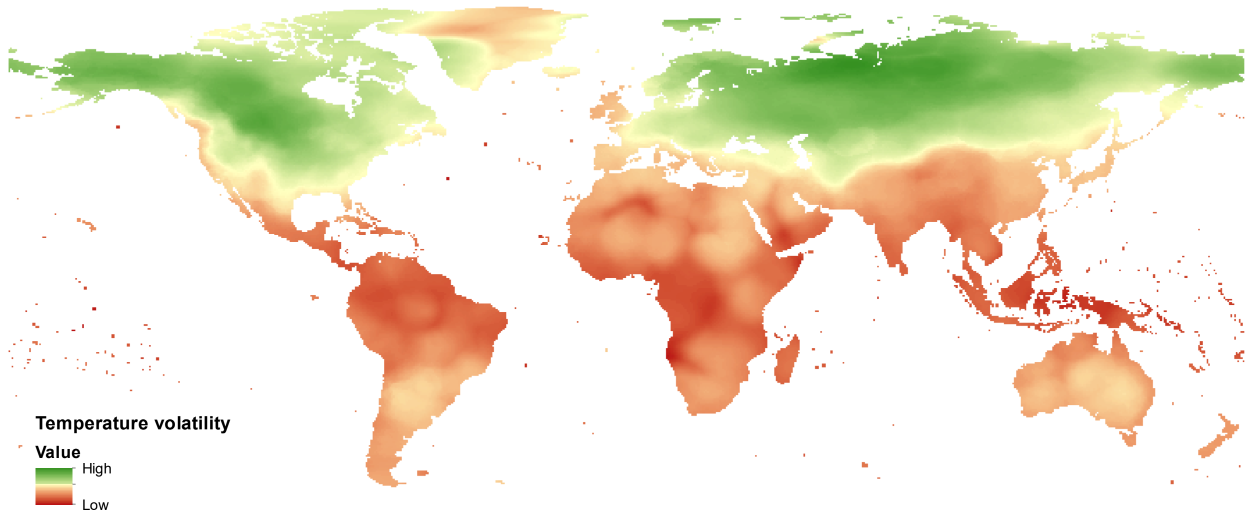
2.3.3 Controls

To address potential concerns deriving from omitted cultural, economic and geographical characteristics of a society that might bias the estimates, I employ a rich and diverse set of control variables.

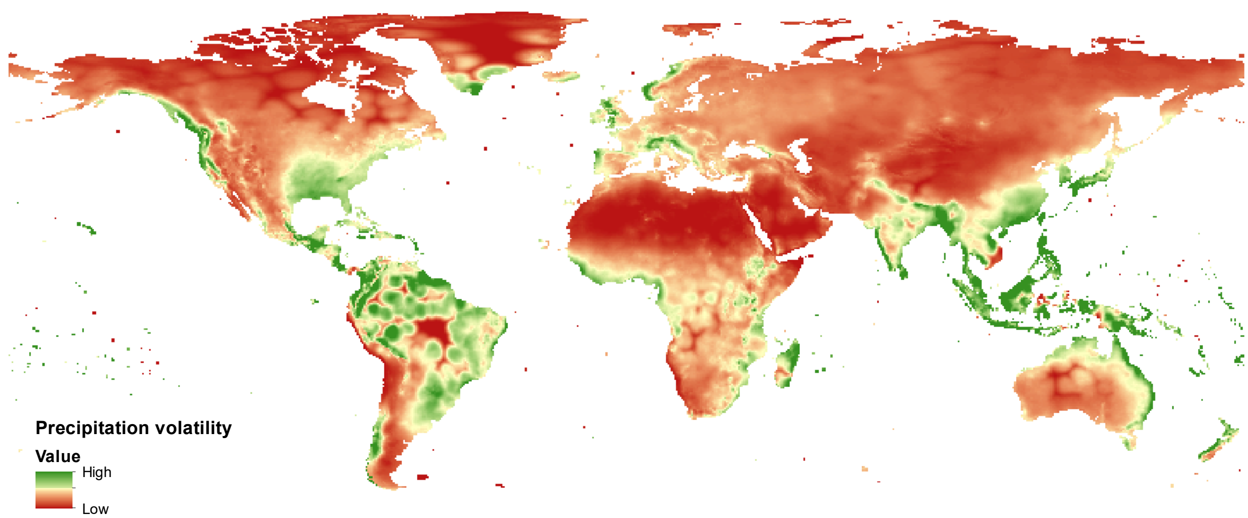
First, I account for cultural and economic characteristics of the society, as recorded in the *Ethnographic Atlas*, that might determine the degree of institutional development and the strength of kin ties. Specifically, I control for the share of subsistence derived from hunting,

[2]While the 100 km radius is employed as baseline throughout the analysis, as robustness check I also employ variables based on different radiuses.

[3]Please refer to Appendix B for more details on the data and construction of the measure.



(A)

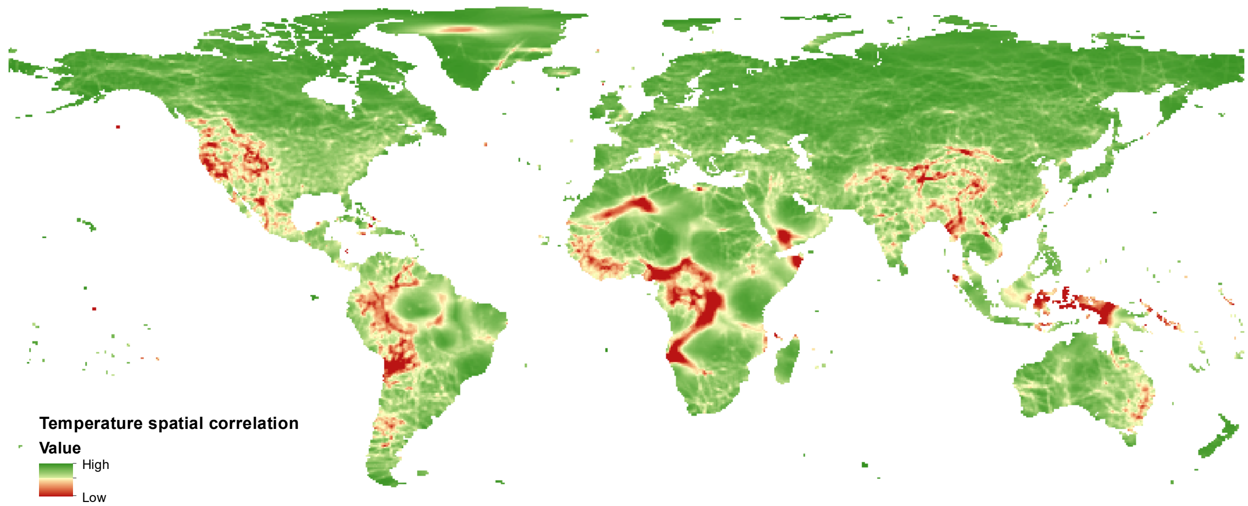


(B)

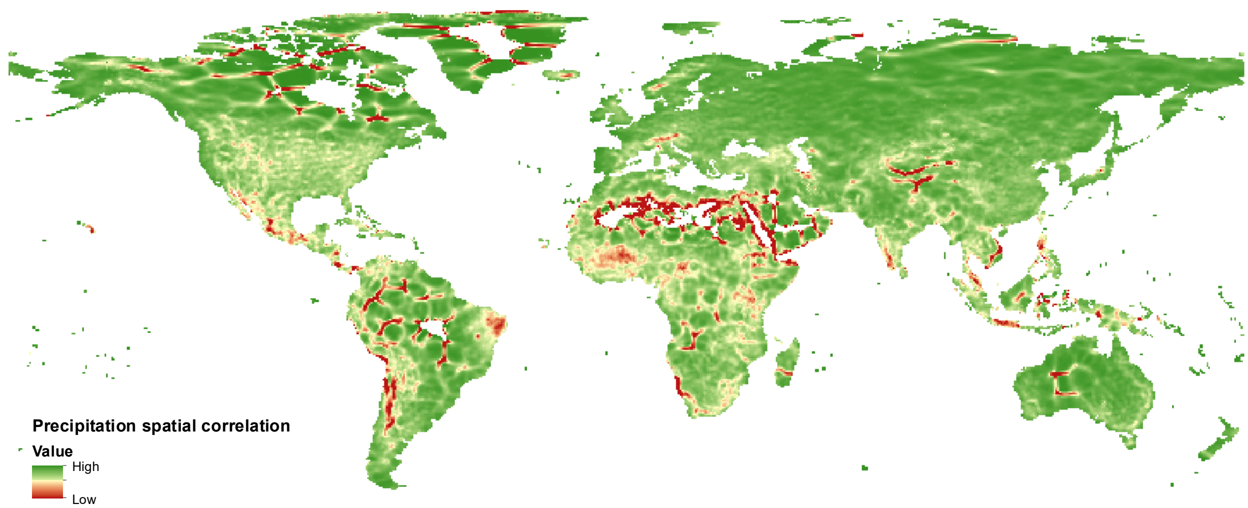
FIGURE 3. Global distribution of intertemporal temperature (A) and precipitation (B) volatility. Index constructed based on monthly measurements of temperature and precipitation over the 20th century from the *Climatic Research Unit (CRU)* database.

gathering, agriculture and animal husbandry, respectively, the intensity of agriculture and settlement structure indicators, distinguishing settlements in categories ranging from fully nomadic to complex settlements.[4] To the extent the date at which the society was sampled

[4]In detail, the settlement structure variable provides 8 ordered categories with 1 indicating fully nomadic (migratory), 2 semi-nomadic, 3 semi-sedentary, 4 compact and impermanent settlements, 5 neighbourhoods of dispersed family homes, 6 separated hamlets forming a single community, 7 compact and relatively permanent



(A)



(B)

FIGURE 4. Global distribution of temperature (A) and precipitation (B) spatial correlation. Index constructed based on monthly measurements of temperature and precipitation over the 20th century from the *Climatic Research Unit (CRU)* database.

is a proxy for measurement error in capturing historical characteristics, I employ the year of ethnographic record as an additional control in the analysis.

Second, to isolate the effect of intertemporal volatility and spatial variability of climate, and to ensure they are not merely proxying for other geographical features, I account for a series of potentially confounding geographic characteristics. These include mean temperature

settlements, and 8 complex settlements.

and precipitation, absolute latitude, elevation, ruggedness, mean and standard deviation of land suitability for agriculture. For these variables, I assign their value to a society following the same approach employed for the main climatic measures, where I compute the average value within an area drawn around the interior centroid of the society's location.

2.4 Results

I begin the analysis by exploring broad patterns in the historical impact of climatic shocks on kin ties and state institutions. Specifically, regressions employ the aggregate measures of climatic intertemporal volatility, capturing the temporal variance of climate in a location, and spatial variability, capturing the spatial variance of weather shocks across neighbouring locations, obtained by combining the individual components of temperature and precipitation.

Table 1 presents estimates of equation (1), with the dependent variable being state institutions in columns (1)-(3) and kinship tightness in columns (4)-(6). For both sets of regressions, the first column initially employs region fixed effects, the second column introduces ethnographic controls and the third column employs the full set of controls including geographic characteristics.

Estimation results indicate that a higher level of intertemporal volatility is associated with more developed state institutions and weaker kin ties. In particular, once the various controls are included, the effect of aggregate climate volatility on state institutions and kin ties is similar both in terms of statistical and economic significance. In fact, as shown by the reported standardised coefficients, a standard deviation increase in climate volatility is associated with an increase in state institutions equal to 10% of a standard deviation in column (3), and a similar decrease in kinship tightness equal to 11.7% of a standard deviation in column (6). With both variables normalised to vary on a scale between 0 and 1 for ease of comparability, and resulting standard deviations of 0.28 and 0.31, respectively, the coefficients imply that a standard deviation increase in climate volatility is associated with a 2.8% increase in state institutions and a 3.6% decrease in kinship tightness.

Overall, results are consistent with the hypothesised mechanism, according to which a more erratic weather shifts the equilibrium between state institutions and kin ties by inducing changes in the predominant economic relations. Specifically, individuals increasing the geographical and social scope of their relations in response to a greater amount of volatility implies a diminishing importance of the family and kin network, whereby increasing the need for formal institutions.

Turning to the spatial component of climate, estimates indicate that a greater level of spatial variability is associated with significantly weaker kin ties, with a standard deviation increase in spatial variability being associated with a decrease in kinship tightness equal to

8.2% of a standard deviation in column (6). The standardised coefficient therefore implies a 2.5% decrease in kinship tightness following a standard deviation increase in climate spatial variability. As hypothesised, the incentive to interact and cooperate with a wider community and other regions in response to climate shocks will be crucially affected by whether these shocks are uniform across different regions, or in other words by how much climate varies across space. Therefore, the gains of large-scale cooperation and trade as a geographical differentiation strategy are going to be a function of the covariance of shocks. Estimation results in fact indicate that the less climate is covarying, the greater is the incentive to increase the geographical and social scope of interaction as opposed to relying on family and kin, therefore decreasing the strength of kin ties.

On the other hand, no significant effect of spatial variability is found for state institutions. The latter result could imply that, contrary to the intertemporal volatility dimension, spatial variability had a significant impact only at a more local level. On the other hand, it is also possible that employing the aggregate version of this measure, combining the individual components of climate, fails to properly capture the role of spatial variability. Next, I therefore move away from the aggregate climate measures and turn my attention to studying the role played by the individual components of temperature and precipitation.

2.4.1 Temperature and Precipitation

In Table 2, I examine the separate roles played by temperature and precipitation, by employing variables capturing the intertemporal volatility and spatial variability of these two components of climate. Estimation results provide patterns in line with those obtained with employing aggregate measures in Table 1, with greater amounts of climate intertemporal volatility and spatial variability associated with more developed state institutions and weaker kin ties.

Specifically, greater intertemporal volatilities of temperature and precipitation have a significantly positive effect on state institutions, and a significantly negative effect on kinship tightness, although only for precipitation in this case. The spatial components of temperature and precipitation, as captured by the spatial variability measures, indicate a positive effect on state institutions and a negative effect on kin ties. In particular, an increase in the spatial variability of temperature has a significantly positive effect on the development of state institutions, while the effect in the case of precipitation is not statistically significant. On the other hand, an increase in the spatial variability of precipitation leads to significantly lower kinship tightness, while the effect is not statistically significant when it comes to the spatial variability of temperature. An explanation for the heterogeneous effect of these two components of climate on state institutions and kin ties can be derived by drawing insights from climatology, according to which temperature is more spatially homogeneous

Table 1. Historical Impact of Climate on State Institutions and Kin Ties

Dependent variable:	*State Institutions*			*Kinship Tightness*		
	(1)	(2)	(3)	(4)	(5)	(6)
Climate volatility	0.121***	0.129***	0.100**	-0.014	-0.105**	-0.117**
	(0.044)	(0.043)	(0.046)	(0.043)	(0.043)	(0.048)
Climate spatial variability	-0.003	-0.042	-0.007	-0.059**	-0.078***	-0.082***
	(0.030)	(0.030)	(0.031)	(0.026)	(0.025)	(0.025)
Mean of dep. var.	0.24	0.24	0.24	0.69	0.69	0.69
St. dev. of dep. var.	0.28	0.28	0.28	0.31	0.31	0.31
Adjusted R^2	0.25	0.40	0.43	0.34	0.39	0.41
Observations	1,089	1,078	1,053	1,145	1,096	1,071
Region FE	✓	✓	✓	✓	✓	✓
Ethnographic controls		✓	✓		✓	✓
Geographic controls			✓			✓

Notes: The table reports OLS estimates. The unit of observation is a society from the *Ethnographic Atlas*. Columns (1)-(3) focus on state institutions (variable 33 in the *Ethnographic Atlas*), whereas columns (4)-(6) focus on kinship tightness (based on variables 8, 11, 15 and 23 in the *Ethnographic Atlas*). Volatility and spatial variability measures are based on climate data from the Climatic Research Unit (CRU). Region fixed effects include dummy variables for North America, Latin America and Caribbean, Sub-Saharan Africa, Middle East and North Africa, Europe and Central Asia, South Asia, East Asia and Pacific region. Ethnographic controls include settlement structure, intensity of agriculture, share of subsistence coming from hunting, gathering, agriculture and animal husbandry and time of ethnographic record. Geographic controls include the level of temperature and precipitation, absolute latitude, elevation, ruggedness, mean and standard deviation of land suitability. Reported coefficients have been standardised, therefore indicating the number of standard deviation changes in the dependent variable for a one–standard deviation change in the explanatory variable. Heteroskedasticity-robust standard errors in parentheses. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

than precipitation, or, in other words, varies across greater distances. Thus, the extent at which temperature and precipitation vary spatially is arguably reflected in their respective influence on the local kin ties and supra-local state institutions. This would imply that precipitation shocks, contrary to temperature shocks, primarily had an impact at a more local level, while temperature shocks had an impact at a more supra-local level.

Estimates are overall stable in terms of coefficient size and statistical significance, providing similar results across the two dependent variables. The only notable exception is the coefficient on temperature volatility when the dependent variable is state institutions, with the coefficient size undergoing consistent changes as the various controls are employed, although statistical significance levels remain similar. This possibly indicates the importance of accounting for the various economic and geographic features of the society in order to precisely estimate the effect of climatic volatility. Once the various controls are accounted for, the estimated effect of a standard deviation increase in the intertemporal volatility measures is associated with a

Table 2. Historical Impact of Climate on State Institutions and Kin Ties: Temperature and Precipitation

Dependent variable:	*State Institutions*			*Kinship Tightness*		
	(1)	(2)	(3)	(4)	(5)	(6)
Temperature volatility	0.272***	0.364***	0.171**	-0.048	-0.058	0.085
	(0.071)	(0.063)	(0.074)	(0.060)	(0.063)	(0.073)
Precipitation volatility	0.089**	0.091**	0.089**	-0.001	-0.110***	-0.136***
	(0.042)	(0.040)	(0.046)	(0.042)	(0.042)	(0.048)
Temperature spatial variability	0.024	0.064***	0.061**	0.015	0.010	0.022
	(0.029)	(0.025)	(0.026)	(0.021)	(0.022)	(0.023)
Precipitation spatial variability	-0.001	-0.046	-0.022	-0.066**	-0.083***	-0.094***
	(0.030)	(0.029)	(0.029)	(0.027)	(0.025)	(0.025)
Mean of dep. var.	0.24	0.24	0.24	0.69	0.69	0.69
St. dev. of dep. var.	0.28	0.28	0.28	0.31	0.31	0.31
Adjusted R^2	0.25	0.41	0.43	0.34	0.39	0.41
Observations	1,085	1,074	1,053	1,141	1,092	1,071
Region FE	✓	✓	✓	✓	✓	✓
Ethnographic controls		✓	✓		✓	✓
geographic controls			✓			✓

Notes: The table reports OLS estimates. The unit of observation is a society from the *Ethnographic Atlas*. Columns (1)-(3) focus on state institutions (variable 33 in the *Ethnographic Atlas*), whereas columns (4)-(6) focus on kinship tightness (based on variables 8, 11, 15 and 23 in the *Ethnographic Atlas*). Volatility and spatial variability measures are based on climate data from the Climatic Research Unit (CRU). Region fixed effects include dummy variables for North America, Latin America and Caribbean, Sub-Saharan Africa, Middle East and North Africa, Europe and Central Asia, South Asia, East Asia and Pacific region. Ethnographic controls include settlement structure, intensity of agriculture, share of subsistence coming from hunting, gathering, agriculture and animal husbandry and time of ethnographic record. Geographic controls include the level of temperature and precipitation, absolute latitude, elevation, ruggedness, mean and standard deviation of land suitability. Reported coefficients have been standardised, therefore indicating the number of standard deviation changes in the dependent variable for a one–standard deviation change in the explanatory variable. Heteroskedasticity-robust standard errors in parentheses. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

change in the dependent variables ranging between 13% and 17% of a standard deviation, while a standard deviation increase in the spatial variability measures is associated with change in the dependent variables ranging between 6% and 9% of a standard deviation.

2.4.2 The Role of Agriculture

So far the discussion has focused on the average effect of climatic conditions across historical societies in the *Ethnographic Atlas* sample. Nonetheless, one can expect that the estimates presented above conceal an underlying heterogeneity in the degree to which societies were historically affected by climatic shocks. Here I focus on subsistence practices as a potential source of heterogeneity, in particular focusing on the role of agriculture.

Out of all the different sources of subsistence, agriculture is possibly the most subject to climate-related shocks. Among the different challenges faced by agriculturalists, two are particularly relevant when it comes to weather risk: reliance on less variety and a general lack of geographic mobility (Diamond, 1987; Harari, 2015). Agriculturalists, in particular sedentary ones, depended on less variety of food for subsistence, compared for example to hunter-gatherers which relied on a combination of different food sources. Therefore, the failure of a crop could have very dire consequences, especially when due to unpredictable factors completely outside of individuals' control, like climate-related ones. Relating to the issue of geographic mobility, the potential output of agriculturalists was tied to single settled area, which made movement across or employment of a wider geographical area an unfeasible strategy in response to weather risk.

A possible shortcoming of this set of estimates is that agricultural practices can be argued to be endogenous, as inherently linked to the type of kin ties and state institutions that a society develops, therefore making results only suggestive. Nonetheless, they provide an insight as to how different subsistence practices can act as source of heterogeneity in the historical impact of climatic conditions.

In order to explore the degree to which agriculturalists were affected by climatic shocks, I consider two variables from the *Ethnographic Atlas* detailing agricultural practices. The first denotes the extent of dependence on agriculture as a source of subsistence. This variable is recorded on a 0-9 scale in which each unit increase corresponds to a roughly 10% increase in the share of the society's subsistence coming from agriculture, relative to hunting, gathering, fishing, and animal husbandry.

The second variable I consider indicates the type of agriculture being practised, specifically its intensity. This variable is recorded on a 0-5 scale, with the increasing categories indicating no agriculture, casual agriculture, extensive or shifting agriculture, horticulture, intensive agriculture and intensive irrigated agriculture. To empirically explore agricultural practices as source of heterogeneity, I employ the specification illustrated by equation (1), with the addition of interaction terms between the measures of climate intertemporal volatility and agricultural practices as reported in the *Ethnographic Atlas*.

The data on agricultural practices provides enough variation to allow me to analyse it as a source of heterogeneity. For example, based on the information in the *Ethnographic Atlas*, around 50% of the societies in the sample had agriculture as its primary source of subsistence. To explore how agriculture-reliant societies were differentially affected by climatic shocks, I focus on the dimension of intertemporal volatility, given that year-to-year variability in climatic conditions most closely captures the type of erratic weather leading to variation in crop yields and potentially crop failure.

Table 3 presents estimation results employing the interaction between the degree of reliance on agriculture and climate intertemporal volatility, first for the aggregate measure of climate in Panel A and then separately for temperature and precipitation in Panel B. While the coefficients on the intertemporal volatility and spatial variability measures are very similar in terms of coefficient size and statistical significance to those reported in Tables 1 and 2, the newly introduced interaction terms reveal an even stronger effect of climate intertemporal volatility when it came to societies relying more heavily on agriculture. With reported coefficients being standardised "beta" coefficients, interaction terms can be interpreted as the differential effect of intertemporal volatility for agriculture-dependent societies as compared to its mean effect. In particular, in column (1) a higher intertemporal volatility of temperature is associated with even stronger state institutions for societies that derived more of their subsistence from agriculture. The results for kinship tightness in column (2) reveal an equivalent picture, with higher intertemporal volatilities of temperature and precipitation leading to even weaker kin ties for agriculture-dependent societies.

It is worth noting that while temperature volatility does not have a statistically significant effect on kinship tightness when considered alone, its interaction with the reliance on agriculture measure has a significantly negative impact. Estimation results indicate that a standard deviation increase in the interaction term between intertemporal volatility and degree of reliance on agriculture is associated with an additional impact, positive for state institutions and negative for kinship tightness, between 8% and 15% of a standard deviation, once all controls are included. In other words, the standardised coefficients imply a further impact of intertemporal volatility for agriculture-dependent societies consisting of increased state institutions and decreased kinship tightness varying between 2.2% and 4.1%.

Similar results are obtained when employing the interaction between intensity of agriculture and intertemporal volatility, as presented in Appendix Table A.2. These estimate provide a picture consistent with the one provided by Table 3, which is not surprising considering that the degree of reliance on agriculture and intensity of agriculture are positively correlated. In particular, the interaction terms between intensity of agriculture and intertemporal volatility, aggregate in Panel A while separated in temperature and precipitation in Panel B, reveal an additional positive impact on state institutions and an additional negative impact on tightness tightness. Although the coefficients on the interaction terms with the intensity of agriculture are smaller in terms of size and weaker in statistical significance when compared to the ones with the reliance on agriculture, both sets of results support a common story: societies heavily dependent on more intensive and sedentary forms of agriculture have weaker kin ties and more developed state institutions when exposed to a greater degree of climatic intertemporal volatility.

Table 3. Historical Impact of Climate on State Institutions and Kin Ties: Reliance on Agriculture

Dependent variable:	*State Institutions*	*Kinship Tightness*
	(1)	(2)
	Panel A	
Climate volatility	0.098**	-0.112**
	(0.046)	(0.049)
Climate spatial variability	-0.009	-0.080***
	(0.031)	(0.025)
Climate volatility × Reliance on agriculture	0.079**	-0.113***
	(0.033)	(0.035)
Adjusted R^2	0.43	0.42
	Panel B	
Temperature volatility	0.218***	0.065
	(0.071)	(0.076)
Precipitation volatility	0.084**	-0.130***
	(0.046)	(0.049)
Temperature spatial variability	0.070***	0.016
	(0.025)	(0.023)
Precipitation spatial variability	-0.019	-0.091***
	(0.028)	(0.025)
Temperature volatility × Reliance on agriculture	0.145***	-0.116***
	(0.035)	(0.034)
Precipitation volatility × Reliance on agriculture	0.022	-0.084**
	(0.032)	(0.044)
Adjusted R^2	0.44	0.42
Mean of dep. var.	0.24	0.69
St. dev. of dep. var.	0.28	0.31
Observations	1,053	1,071
Region FE	✓	✓
Ethnographic controls	✓	✓
Geographic controls	✓	✓

Notes: The table reports OLS estimates. The unit of observation is a society from the *Ethnographic Atlas*. Column (1) focuses on state institutions (variable 33 in the *Ethnographic Atlas*), whereas columns (2) focuses on kinship tightness (based on variables 8, 11, 15 and 23 in the *Ethnographic Atlas*). Volatility and spatial variability measures are based on climate data from the Climatic Research Unit (CRU). Panel A employs aggregate climate measures, while Panel B employs separate temperature and precipitation measures. The reliance on agriculture (variable 5 in the *EA*) is recorded on a 0-9 scale in which each unit increase corresponds to a 10% increase in the share of the society's subsistence coming from agriculture relative to hunting, gathering, fishing, and animal husbandry. Region fixed effects include dummy variables for North America, Latin America and Caribbean, Sub-Saharan Africa, Middle East and North Africa, Europe and Central Asia, South Asia, East Asia and Pacific region. Ethnographic controls include settlement structure, intensity of agriculture, share of subsistence coming from hunting, gathering, agriculture and animal husbandry and time of ethnographic record. Geographic controls include the level of temperature and precipitation, absolute latitude, elevation, ruggedness, mean and standard deviation of land suitability. Reported coefficients have been standardised, therefore indicating the number of standard deviation changes in the dependent variable for a one–standard deviation change in the explanatory variable. Heteroskedasticity-robust standard errors in parentheses. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

2.4.3 Seemingly Unrelated Regressions

The analysis up to this point has employed a system of two separate equations, one for each dependent variable, based on the specification illustrated by equation (1). A potential issue with this approach would manifest in case the estimates of one of the two equations part of the system were to be driven by unobservable factors inherently related to the other equation. For example, the effect of climatic conditions on kinship tightness could be driven by unobserved factors which are in turn related to the effect of said climatic conditions on state institutions, or vice-versa. Another, perhaps even more direct, form of this problem is the case where one of the two outcomes is not being affected by climatic conditions, but is instead indirectly impacted through the other outcome. Specifically, climatic conditions could, for example, only impact state institutions in a society, whose change subsequently affects the degree of kinship tightness. The opposite could also be true.

Regarding the first form of this concern, replicating the baseline estimates presented in Tables 1 and 2 with sets of Seemingly Unrelated Regressions shows through a Breusch-Pagan test for independent equations that error terms are uncorrelated (p=0.1803 for the version with separate temperature and precipitation components; p=0.2163 for the version employing aggregate climate measures). In other words, the test results indicate that there is no significant cross-equation correlation of residuals which would bias the estimates of the equations if run separately.

To direct the second form of this potential issue, that is whether results are being driven only by the impact of climatic conditions on either state institutions or kin ties, I repeat the regressions presented in Tables 1 and 2 through Seemingly Unrelated Regressions while controlling for the other outcome variable. In other words, I estimate the impact of intertemporal volatility and spatial variability of climate on kinship tightness while controlling for the level of state institutions, and vice-versa. While this approach could potentially introduce endogeneity concerns, which is why such specification is not employed as baseline, it can allow to establish if results are being driven by the aforementioned potential issue. Results presented in Appendix Table A.3 show that the estimated impact of intertemporal volatility and spatial variability of climate, both with the aggregate measure of climate in Panel A and with separate temperature and precipitation components in Panel B, is unchanged, with estimation results being remarkably similar to the baseline ones presented in Tables 1 and 2.

2.4.4 Robustness

Throughout the analysis, I assign the intertemporal volatility and spatial variability measures by computing their average value in a 100 km radius of the interior centroid of societies' location, as indicated in the *Ethnographic Atlas*. In Appendix Table A.4 I re-estimate the main regressions using variables constructed on the basis of a smaller radius of 50 km and a bigger radius of 200 km. Results are consistent when considering different radiuses.

Next, I consider the potential correlation of observed and unobserved features across space, which might bias the estimates. Specifically, to account for the possibility that error terms are correlated across regions, I employ spatially-adjusted standard errors using the methodology of Conley (1999). Appendix Table A.5 reports the main estimates with standard errors adjusted for two-way spatial correlation based on windows of 100, 200 and 500 km, which all provide similar levels of statistical significance.

3 Long-Run Analysis

Over the past two centuries, industrialisation and structural change has presumably made climatic conditions a less important factor for many economies. Thus, it would reasonable to expect the historical link between climate and cultural and institutional features to have disappeared. On the other hand, institutions and culture are generally deemed to persist over time (see Nunn (2014) for a review), so features shaped by past climatic conditions could still potentially characterise modern societies. In this section I explore the long-run impact of historical kinship tightness and state development, based on the hypothesis that modern countries persistently reflect historical differences in the role and importance of kin ties and state institutions, even when the factors that led to the formation of these differences are no longer salient. To establish the persistent impact of these historical characteristics, I explore their relation with the evolution of institutions over time and until today. In particular, I formulate the hypothesis that modern countries' institutional quality negatively reflects their historical kinship tightness and positively reflects their historical state development.

3.1 Empirical Model

To formally test for the hypothesis set out above, the analysis in this section is based on estimates of the following specification:

$$IQ_{ct} = \beta_0 + \beta_1 \; KT_c + \beta_2 \; SI_c + \alpha + Z_c' \, \Theta + \epsilon_{ct} \tag{2}$$

where IQ_{ct} captures the institutional quality of country c at time t. The variables KT_c and SI_c reflect the ones employed in the previous section, and denote, respectively, the kinship tightness and state development of the ancestors of country c. α denotes continent fixed effects. Z_c indicates a series of country-level controls capturing the historical, contemporary and geographic features of country c.

3.2 Data

In the following, I introduce the data employed to estimate equation (2).

3.2.1 Institutional Quality Data

To measure the institutional quality of modern countries, I employ data from the *V-Dem* project (Coppedge, Gerring, Knutsen, Lindberg, Skaaning, Teorell, Altman, Bernhard, Fish, Cornell, et al., 2018), which provides yearly multidimensional indicators of the countries' institutions. Specifically, I employ the 8th *V-Dem* dataset, which integrates the *Historical V-Dem*, providing institutional scores dating as far back as 1800 for a number of countries. From the *V-Dem* database, I employ measures of four institutional dimensions that are well-established in the literature, namely Rule of Law, Property Rights, Electoral Democracy and Control of Corruption. Institutional quality is measured through indexes varying between 0 and 1, with 1 indicating better scores.[5] The sample of countries is not balanced for all years, especially for the first part of the 19th century. Country coverage improves in the latter half of the 19th century and almost consistently covers the entire world starting from the first half of the 20th century.

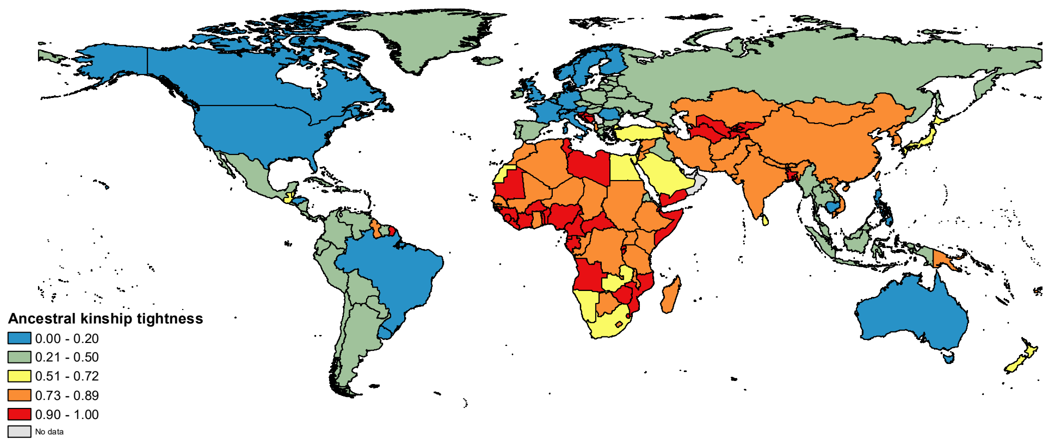
To explore long-run patterns I employ measures of average institutional quality across periods of time, instead of year-by-year changes. Specifically, I follow the approach of Giuliano and Nunn (2013), computing the average institutional quality across 50-year intervals between 1800 and 2000, therefore resulting in four t time periods, with t={1800-1850, 1850-1900, 1900-1950, 1950-2000}.

3.2.2 Ancestral Characteristics Data

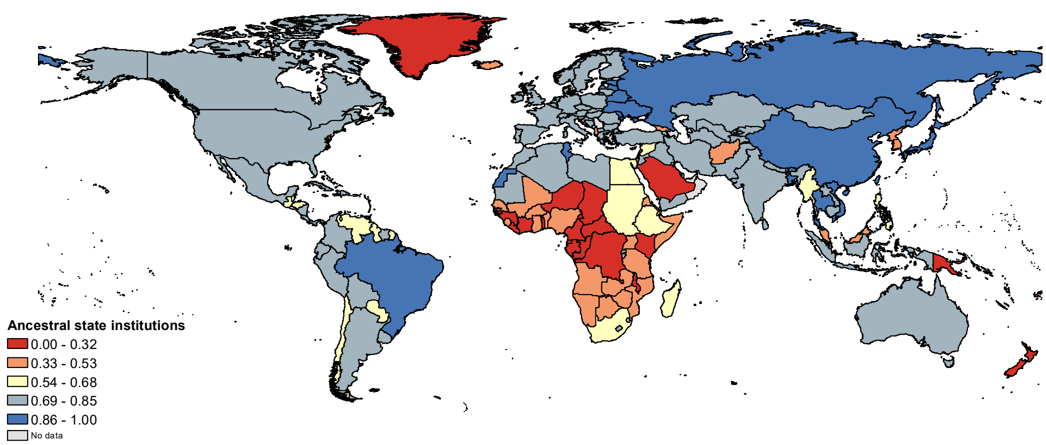
I link the historical data from the *Ethnographic Atlas* to modern countries by employing Giuliano and Nunn's (2018) *Ancestral Characteristics of Modern Populations* dataset. This data applies a language-based matching to construct population-weighted averages of the prevalence of the characteristics from *Ethnographic Atlas* among the ancestors of the people composing the population of each country. In Figure 5, I show the average value of kinship tightness and state institutions among modern countries' ancestors. It is important to stress that through this data I obtain country-level measures of the ancestral characteristics of those who inhabit the country in modern times, instead of those that inhabited the area associated with the country in the past. An illustrative example can be provided by considering the case of the US. While in the data provided by the *Ethnographic Atlas*, as shown in Figure 1, the observations associated with what is today's US territory are Native Americans, in the country-level data constructed through *Ancestral Characteristics of Modern Populations* only a small part of the country's population has ancestry tracing to Native Americans, while

[5]The original corruption index from *V-Dem*, the Political Corruption index, runs from less corrupt to more corrupt. The Control of Corruption index is obtained by inverting the values of the original index.

the major part of the population will have ancestral characteristics tracing to European populations.



(A)



(B)

FIGURE 5. Average kinship tightness (A) and state institutions (B) among modern countries' ancestors.

3.2.3 Controls

To attenuate the risk of the estimates of the long-run impact of historical kinship tightness and state development from being biased by omitted features, I employ a rich a set of country-level controls. First, I include a series of country-level ethnographic controls, based on the *Ancestral Characteristics of Modern Populations* dataset, capturing the ethnographic characteristic of the country populations' ancestors. These characteristics include dependence on hunting, gathering, animal husbandry and agriculture, average complexity of settlements and average time of ethnographic record. Additionally, I include the years passed since the country experienced the Neolithic Revolution. Second, I control for geographic and climatic characteristics of the country, including absolute latitude, average land suitability for agriculture, average elevation and ruggedness, percentage of the country's area in tropical and temperate climates, average distance of the country's surface to a coast or river, average level of precipitation and temperature, land area, an indicator for islands and an indicator for landlocked countries. Finally, I include a series of country-level contemporary controls, which include population density, share of the country's population being Catholics, Protestants, Muslims or belonging to other religions.

3.3 Results

3.3.1 Ancestral Kinship Tightness

I begin the analysis by focusing on the long-run impact of ancestral kinship tightness. Table 4 reports estimates of equation (2), with the dependent variable being the average institutional quality across 50-year intervals between 1800 and 2000. Specifically, the institutional quality measures employed as dependent variable are rule of law in column (1), property rights in column (2), level of democracy in column (3) and control of corruption in column (4). Regressions employ continent fixed effects and period fixed effects, in addition to the country-level controls previously described. Given that I estimate the equation as a pooled regression, specifically a repeated cross-section of four 50-year periods, I employ heteroskedasticity-robust standard errors clustered by country and period.

Across the different institutional scores, results consistently point to a significantly negative effect of ancestral kinship tightness, indicating that countries whose populations' ancestors were characterised by stronger kin ties are associated with significantly lower institutional quality across time. Reported coefficient are standardised "beta" coefficient, therefore indicating the number of standard deviation changes in the dependent variable for a one-standard deviation change in the explanatory variable. Thus, the coefficients in Table 4 indicate that a standard deviation increase in ancestral kinship tightness is associated with an average decrease in institutional quality ranging between 17% and 31% of a standard deviation. Considering for example the estimates from column (1), employing rule of law as the dependent variable, a one-standard-deviation increase in ancestral kinship tightness is associated with a decrease in institutional quality of 31% of a standard deviation. With the institutional measures varying between 0 and 1, and with the standard deviation of the rule of law variable being 0.30, this implies an increase in institutional quality of approximately 9.3%,

Table 4. Long-Run Impact of Ancestral Kinship Tightness on Institutional Quality

Dependent variable:	*Rule of Law*	*Property Rights*	*Democracy*	*Control of Corr.*
	(1)	(2)	(3)	(4)
Ancestral kinship tightness	-0.309***	-0.166***	-0.220***	-0.236***
	(0.077)	(0.063)	(0.037)	(0.082)
Mean of dep. var.	0.50	0.50	0.33	0.52
St. dev. of dep. var.	0.30	0.28	0.27	0.28
Adjusted R^2	0.51	0.56	0.71	0.48
Observations	479	482	475	475
Continent FE	✓	✓	✓	✓
Period FE	✓	✓	✓	✓
Country-level controls	✓	✓	✓	✓

Notes: The table reports OLS estimates. The unit of observation is a country. The dependent variable indicates the average institutional quality across 50-year intervals between 1800 and 2000. The institutional quality measures are rule of law in column (1), property rights in column (2), level of democracy in column (3) and control of corruption in column (4). The explanatory variable captures the average kinship tightness of the ancestors of the country's population. Country-level controls include ethnographic, geographic and contemporary controls. Ethnographic controls include ancestral dependence on hunting, gathering, animal husbandry and agriculture, ancestral complexity of settlements, average time of ethnographic record and years since the Neolithic Revolution. Geographic controls include absolute latitude, average elevation, average ruggedness, percentage of country's land in tropical and temperate climates, average distance to a coast or river, mean level of temperature and precipitation, average land quality, land area, an indicator for islands and an indicator for landlocked countries. Contemporary controls include population density, share of the population being Catholics, Protestants, Muslims or belonging to other religions. All specifications include continent fixed effects and period fixed effects. Reported coefficients have been standardised, therefore indicating the number of standard deviation changes in the dependent variable for a one–standard deviation change in the independent variable. Heteroskedasticity-robust standard errors, clustered by country and period, are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

showing how the effect is not only statistically significant, but also economically meaningful.

Going beyond the overall effect of ancestral kinship tightness on institutional quality over 1800-2000, columns (1)-(4) of Appendix Table A.6 employ as dependent variable the average institutional quality of the individual 50-year periods. The number of observations differs across columns due to unbalanced country samples over time. While there is variation in the size of the coefficients and their statistical significance, period estimates are overall consistent with the results from Table 4, with a significantly negative impact of ancestral kinship tightness on institutional quality across time. It is worth noting that for all institutional quality indexes results are particularly weak for the period 1800-1850, with only one coefficient being significant at the 10% level. One potential explanation may be that the small sample of countries available for this period causes the estimates to lack precision. On the other hand, it may also indicate that acestral kinship tightness did not have an impact on institutional quality during the period corresponding to the beginning of the first industrial revolution, instead starting to have an impact only at a later point in time.

3.3.2 Ancestral State Institutions

Next, I turn to the long-run impact of ancestral state institutions, with estimates of equation (2) presented in Table 5. Similar to the previous set of results, the level of ancestral state institutions is shown to have a consistent effect across the different institutional quality indexes. Specifically, estimation results indicate that countries with populations' ancestors characterised by more developed state institutions in the historical period are associated with a significantly higher quality of institutions across 1800-2000. In terms of effect size, the estimated impact of ancestral state institutions is similar to the one of ancestral kinship tightness, although comparatively lower on average, with a standard deviation increase in ancestral state institutions associated with an increase in institutional quality ranging between 10% and 26% of a standard deviation. The measure of property rights is the only institutional dimension for which the impact of ancestral state institutions is estimated to be bigger than the one of ancestral kinship tightness, whereas for all the other three institutional quality indexes the effect of ancestral kinship tightness is estimated to be greater.

Similar to the previous set of results, I also study the impact of ancestral state institutions on institutional quality for each of the individual 50-year periods in 1800-2000, with results reported in Appendix Table A.7. While results are overall consistent with the pooled ones reported in Table 5, these period estimates are in general less robust in terms of statistical significance, in particular when it comes to the estimated impact of ancestral state institutions on the level of democracy and control of corruption.

3.3.3 Joint Impact of Ancestral Kinship Tightness and State Institutions

Finally, I explore the joint long-run impact of ancestral kinship tightness and ancestral state institutions by estimating equation (2) while employing both historical characteristics of the country's population as explanatory variables, with results reported in Table 6. Estimates reveal that the previously uncovered impacts of both ancestral kinship tightness and ancestral state institutions on institutional quality hold even when the two are employed together. In fact, not only the coefficients on both explanatory variables retain their high significance, but the effect size is remarkably similar to the ones estimated individually in Tables 4 and 5. Therefore, this last set of results supports the idea of kinship tightness and state institutions of the country's ancestors jointly affecting the quality of institutions across the last two centuries. This result speaks to the literature on the close relationship between culture and institutions and their persistent effect (Tabellini, 2008b), with dimensions of past culture (kin ties) and past formal institutions (state institutions) jointly and persistently affecting modern institutional quality.

Table 5. Long-Run Impact of Ancestral State Institutions on Institutional Quality

Dependent variable:	*Rule of Law*	*Property Rights*	*Democracy*	*Control of Corr.*
	(1)	(2)	(3)	(4)
Ancestral state institutions	0.261***	0.231***	0.099**	0.148**
	(0.067)	(0.071)	(0.048)	(0.067)
Mean of dep. var.	0.50	0.50	0.33	0.52
St. dev. of dep. var.	0.30	0.28	0.27	0.28
Adjusted R^2	0.50	0.57	0.70	0.47
Observations	479	482	475	475
Continent FE	✓	✓	✓	✓
Period FE	✓	✓	✓	✓
Country-level controls	✓	✓	✓	✓

Notes: The table reports OLS estimates. The unit of observation is a country. The dependent variable indicates the average institutional quality across 50-year intervals between 1800 and 2000. The institutional quality measures are rule of law in column (1), property rights in column (2), level of democracy in column (3) and control of corruption in column (4). The explanatory variable captures the average state institutions of the ancestors of the country's population. Country-level controls include ethnographic, geographic and contemporary controls. Ethnographic controls include ancestral dependence on hunting, gathering, animal husbandry and agriculture, ancestral complexity of settlements, average time of ethnographic record and years since the Neolithic Revolution. Geographic controls include absolute latitude, average elevation, average ruggedness, percentage of country's land in tropical and temperate climates, average distance to a coast or river, mean level of temperature and precipitation, average land quality, land area, an indicator for islands and an indicator for landlocked countries. Contemporary controls include population density, share of the population being Catholics, Protestants, Muslims or belonging to other religions. All specifications include continent fixed effects and period fixed effects. Reported coefficients have been standardised, therefore indicating the number of standard deviation changes in the dependent variable for a one–standard deviation change in the independent variable. Heteroskedasticity-robust standard errors, clustered by country and period, are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Furthermore, results by individual 50-year periods are reported in Appendix Table A.8 and show patterns consistent with previous results employing ancestral kinship tightness and ancestral state institutions both individually and jointly.

3.3.4 Robustness

Results are robust to continent fixed effects and other potentially confounding ethnographic, geographic and contemporary factors. While the estimates discussed up to this point employ the full set of controls, in Appendix Table A.9 I show how estimates change when gradually introducing the individual elements of the control set. For brevity, I only report results from this exercise employing the rule of law institutional quality index, for the period 1950-2000.

Starting with the ancestral kinship tightness explanatory variable in Panel A, introducing the various control sets one at a time does not significantly affect the results. Column (6) of Appendix Table A.9 further takes into account GDP per capita. Given that this may be a potentially bad control, to limit issues of reverse causality I employ GDP per capita at the beginning of the period of interest, in this case 1950. Results show that the even after

Table 6. Long-Run Impact of Ancestral Kinship Tightness and Ancestral State Institutions on Institutional Quality

Dependent variable:	*Rule of Law*	*Property Rights*	*Democracy*	*Control of Corr.*
	(1)	(2)	(3)	(4)
Ancestral kinship tightness	-0.332***	-0.186***	-0.229***	-0.250***
	(0.077)	(0.063)	(0.038)	(0.080)
Ancestral state institutions	0.283***	0.243***	0.113**	0.165***
	(0.061)	(0.071)	(0.045)	(0.065)
Mean of dep. var.	0.50	0.50	0.33	0.52
St. dev. of dep. var.	0.30	0.28	0.27	0.28
Adjusted R^2	0.53	0.58	0.72	0.49
Observations	479	482	475	475
Continent FE	✓	✓	✓	✓
Period FE	✓	✓	✓	✓
Country-level controls	✓	✓	✓	✓

Notes: The table reports OLS estimates. The unit of observation is a country. The dependent variable indicates the average institutional quality across 50-year intervals between 1800 and 2000. The institutional quality measures are rule of law in column (1), property rights in column (2), level of democracy in column (3) and control of corruption in column (4). The explanatory variables capture the average kinship tightness and average state institutions of the ancestors of the country's population. Country-level controls include ethnographic, geographic and contemporary controls. Ethnographic controls include ancestral dependence on hunting, gathering, animal husbandry and agriculture, ancestral complexity of settlements, average time of ethnographic record and years since the Neolithic Revolution. Geographic controls include absolute latitude, average elevation, average ruggedness, percentage of country's land in tropical and temperate climates, average distance to a coast or river, mean level of temperature and precipitation, average land quality, land area, an indicator for islands and an indicator for landlocked countries. Contemporary controls include population density, share of the population being Catholics, Protestants, Muslims or belonging to other religions. All specifications include continent fixed effects and period fixed effects. Reported coefficients have been standardised, therefore indicating the number of standard deviation changes in the dependent variable for a one–standard deviation change in the independent variable. Heteroskedasticity-robust standard errors, clustered by country and period, are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

controlling for initial GDP per capita the significantly negative effect of ancestral kinship tightness holds. I then repeat the exercise with ancestral state institutions in Panel B.

Next, I examine the potential role of omitted factors in altering the findings. As Appendix Table A.10 establishes, it appears to be unlikely that unobservables are driving the results of Tables 4 and 5. In particular, as indicated in columns (2), (4), (6) and (8) (using columns (1), (3), (5) and (7) as baseline specifications), the AET ratios (Altonji, Elder, and Taber, 2005; Bellows and Miguel, 2009) and Oster's δ (Oster, 2016), measuring how strongly correlated unobservables would have to be in order to account for the full size of the coefficient, are different from the critical value of 1. Additionally, Oster's β statistic, indicating the estimated effects of ancestral kinship tightness and ancestral state institutions if the proportion of selection of observables and unobservables was to be equal, is comparable to the estimated coefficients, with zero not belonging to the interval created by the estimated value and Oster's β.

4 Concluding Remarks

The impact of cultural and institutional features on economic development has been the subject of extensive research, with the relationship between these two fundamental determinants of growth being the focus of significant theoretical and empirical advancements over the last years. My paper contributes to this literatures by examining the joint evolution of kin ties and state institutions. Based on the assumption that the scope of socio-economic relations was historically reflected in the role and importance of kin and states, I first explore the common geographical forces that have historically driven kinship tightness and the development of state institutions. Subsequently, I study the long-run persistence of historical differences in kinship tightness and state institutions across the populations of modern countries.

In the first part of the paper, I explore the role of climatic conditions as an exogenous force that affected the strength of kin ties and the development of formal state institutions in the historical era. Employing data on the characteristics of historical societies from different regions of the world, and combining it with granular and high-frequency data on temperature and precipitation, I find consistent evidence of weaker kin ties and stronger state institutions in regions where climate is characterised by higher degrees of intertemporal volatility and spatial variability. In the second part of the paper, I employ country-level institutional quality indicators ranging from the 19th century to the present and link modern countries' populations with the historical characteristics of their ancestors. The analysis allows me to establish that countries whose ancestors featured greater state development and weaker ties are consistently associated with institutions of significantly higher quality over time. Combining the findings from the two parts of the paper, the evidence implies that while kinship tightness and state institutions were historically shaped by geographical forces, such as the climatic shocks on which I focus, their long-run influence is not geographically bound, with a persistent impact that is carried through the ancestral features of modern populations even after large processes of migration.

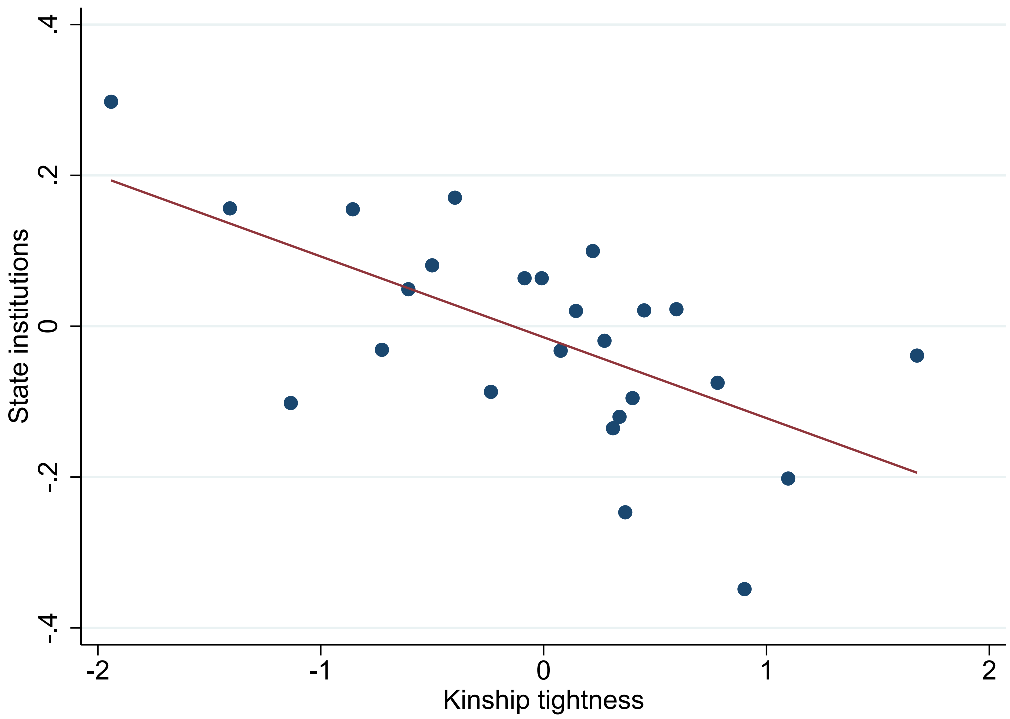
This research builds on and contributes to the literatures on the geographic forces that shaped both culture and institutions historically, and on the close relationship and long-run persistence of cultural and institutional features. Taken together, findings overall support the interconnectedness of three fundamental causes of growth, geography, culture and institutions, both historically and in the present, a rich avenue of research that calls for further study.

Appendix A - Tables and Figures

Table A.1. Summary Statistics

	Average	*St.Dev.*	*Min*	*Max*
Historical Analysis				
Kinship tightness	0.69	0.31	0	1
State institutions	0.24	0.28	0	1
Dependence gathering	1.02	1.59	0	8
Dependence hunting	1.43	1.56	0	9
Dependence animal husbandry	1.56	1.79	0	9
Dependence agriculture	4.46	2.75	0	9
Intensity of agriculture	3.32	1.59	1	6
Settlement structure	5.11	2.23	1	8
Ethnographic record	1904	53	1500	1984
Average temperature (°C)	19.32	9.61	-15.73	29.88
Average precipitation (m)	1.10	0.79	0.01	3.97
Temperature volatility (°C)	0.94	0.57	0	2.82
Precipitation volatility (m)	0.37	0.24	0	1.90
Temperature spatial correlation	0.98	0.01	0.84	0.99
Precipitation spatial correlation	0.95	0.03	0.72	0.99
Absolute latitude	21.66	17.52	0	71
Terrain ruggedness	0.13	0.15	0	1
Elevation (m)	622.17	642.33	0	4,748.82
Average land quality	0.44	0.26	0	0.98
St.dev. land quality	0.11	0.08	0	0.43
Long-Run Analysis				
Rule of law	0.55	0.31	0.03	1
Property rights	0.68	0.22	0.01	0.94
Democracy	0.55	0.25	0.02	0.91
Control of corruption	0.47	0.30	0.03	0.99
Ancestral kinship tightness	0.61	0.32	0	1
Ancestral state institutions	0.63	0.23	0	1
Ancestral dependence gathering	1.16	0.34	1	4
Ancestral dependence hunting	1.33	0.44	1	3
Ancestral dependence animal husbandry	3.85	1.58	1	9.88
Ancestral dependence agriculture	6.86	1.48	1.13	9.97
Ancestral settlement structure	6.37	1.44	1.09	8
Absolute latitude	26.07	17.43	1	65
Terrain ruggedness	0.20	0.19	0	1
Elevation (m)	540	490	20	2,670
Average land quality	0.43	0.26	0	0.97
% Land in tropical climate	37.41	44.20	0	100
% Land in temperate climate	30.74	41.28	0	100
Average temperature (°C)	18.23	8.42	-7.93	28.64
Average precipitation (m)	0.94	0.64	0.03	2.84
Distance to waterways (1000 km)	0.34	0.45	0.01	2.39
Landlocked	0.22	0.41	0	1
Island	0.17	0.38	0	1
Neolithic transition timing (10,000 years)	0.53	0.21	0.14	1.04
% Muslims in population	22.86	34.89	0	99.90
% Protestants in population	12.30	20.72	0	97.80
% Catholics in population	30.61	35.44	0	96.90

Figure A.1. Kin Ties and State Institutions in the Ethnographic Atlas



Notes: Binscatter plot between kinship tightness and state institutions across historical societies in the *Ethnographic Atlas*. Both variables are standardised into z-scores. Conditional on source of subsistence, settlement structure and region fixed effects.

Table A.2. Historical Impact of Climate on State Institutions and Kin Ties: Intensity of Agriculture

Dependent variable:	*State Institutions*	*Kinship Tightness*
	(1)	(2)
	Panel A	
Climate volatility	0.094**	-0.110**
	(0.046)	(0.048)
Climate spatial variability	-0.012	-0.075***
	(0.030)	(0.025)
Climate volatility × Intensity of agriculture	0.066*	-0.084**
	(0.036)	(0.033)
Adjusted R^2	0.43	0.41
	Panel B	
Temperature volatility	0.181**	0.078
	(0.074)	(0.076)
Precipitation volatility	0.084**	-0.130***
	(0.045)	(0.048)
Temperature spatial variability	0.063**	0.020
	(0.026)	(0.024)
Precipitation spatial variability	-0.025	-0.088***
	(0.027)	(0.025)
Temperature volatility × Intensity of agriculture	0.084*	-0.086**
	(0.039)	(0.035)
Precipitation volatility × Intensity of agriculture	0.056	-0.068*
	(0.039)	(0.039)
Adjusted R^2	0.43	0.42
Mean of dep. var.	0.24	0.69
St. dev. of dep. var.	0.28	0.31
Observations	1,053	1,071
Region FE	✓	✓
Ethnographic controls	✓	✓
Geographic controls	✓	✓

Notes: The table reports OLS estimates. The unit of observation is society from the *Ethnographic Atlas*. Column (1) focuses on state institutions (variable 33 in the *EA*), whereas column (2) focuses on kinship tightness (based on variables 8, 11, 15 and 43 in the *EA*). Volatility and spatial variability measures are based on climate data from the Climatic Research Unit (CRU). Panel A employs aggregate climate measures, while Panel B employs separate temperature and precipitation measures. The intensity of agriculture (variable 28 in the *EA*) is recorded on a 0-5 scale, with the increasing categories indicating: no agriculture, casual agriculture, extensive or shifting agriculture, horticulture, intensive agriculture and intensive irrigated agriculture. Region fixed effects include dummy variables for North America, Latin America and Caribbean, Sub-Saharan Africa, Middle East and North Africa, Europe and Central Asia, South Asia, East Asia and Pacific region. Ethnographic controls include settlement structure, intensity of agriculture, share of subsistence coming from hunting, gathering, agriculture and animal husbandry and time of ethnographic record. Geographic controls include the level of temperature and precipitation, absolute latitude, elevation, ruggedness, mean and standard deviation of land suitability. Reported coefficients have been standardised, therefore indicating the number of standard deviation changes in the dependent variable for a one–standard deviation change in the explanatory variable. Heteroskedasticity-robust standard errors in parentheses. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.3. Historical Impact of Climate on State Institutions and Kin Ties: Seemingly Unrelated Regressions

Dependent variable:	*State Institutions*	*Kinship Tightness*
	(1)	(2)
	Panel A	
Climate volatility	0.091**	-0.108***
	(0.042)	(0.042)
Climate spatial variability	-0.013	-0.081***
	(0.027)	(0.027)
Kinship tightness	-0.076**	
	(0.031)	
State institutions		-0.077**
		(0.031)
Adjusted R^2	0.44	0.44
	Panel B	
Temperature volatility	0.171**	0.104
	(0.074)	(0.074)
Precipitation volatility	0.079**	-0.130***
	(0.042)	(0.041)
Temperature spatial variability	0.060**	0.025
	(0.028)	(0.028)
Precipitation spatial variability	-0.028	-0.094***
	(0.027)	(0.027)
Kinship tightness	-0.083***	
	(0.031)	
State institutions		-0.083***
		(0.031)
Adjusted R^2	0.44	0.44
Mean of dep. var.	0.24	0.69
St. dev. of dep. var.	0.28	0.31
Observations	1,041	1,041
Region FE	✓	✓
Ethnographic controls	✓	✓
Geographic controls	✓	✓

Notes: The table reports SUR estimates. The unit of observation is a society from the *Ethnographic Atlas*. Column (1) focuses on state institutions (variable 33 in the *Ethnographic Atlas*), whereas columns (2) focuses on kinship tightness (based on variables 8, 11, 15 and 23 in the *Ethnographic Atlas*). Volatility and spatial variability measures are based on climate data from the Climatic Research Unit (CRU). Panel A employs aggregate climate measures, while Panel B employs separate temperature and precipitation measures. Region fixed effects include dummy variables for North America, Latin America and Caribbean, Sub-Saharan Africa, Middle East and North Africa, Europe and Central Asia, South Asia, East Asia and Pacific region. Ethnographic controls include settlement structure, intensity of agriculture, share of subsistence coming from hunting, gathering, agriculture and animal husbandry and time of ethnographic record. Geographic controls include the level of temperature and precipitation, absolute latitude, elevation, ruggedness, mean and standard deviation of land suitability. Reported coefficients have been standardised, therefore indicating the number of standard deviation changes in the dependent variable for a one–standard deviation change in the explanatory variable. Heteroskedasticity-robust standard errors in parentheses. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.4. Historical Impact of Climate on State Institutions and Kin Ties: Robustness to Radius Size

Dependent variable:	*State Institutions*			*Kinship Tightness*		
	50 km	100 km	200 km	50 km	100 km	200 km
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A					
Climate volatility	0.091**	0.100**	0.094*	-0.101**	-0.117**	-0.134***
	(0.044)	(0.046)	(0.051)	(0.046)	(0.048)	(0.053)
Climate spatial variability	-0.004	-0.007	-0.013	-0.067***	-0.082***	-0.098***
	(0.032)	(0.031)	(0.034)	(0.024)	(0.025)	(0.028)
Adjusted R^2	0.43	0.43	0.43	0.41	0.41	0.41
	Panel B					
Temperature volatility	0.152**	0.171**	0.181**	0.080	0.085	0.089
	(0.072)	(0.074)	(0.078)	(0.072)	(0.073)	(0.075)
Precipitation volatility	0.080**	0.089**	0.083*	-0.119***	-0.136***	-0.161***
	(0.042)	(0.046)	(0.053)	(0.045)	(0.048)	(0.054)
Temperature spatial variability	0.053**	0.061**	0.061**	0.016	0.022	0.023
	(0.025)	(0.026)	(0.030)	(0.022)	(0.023)	(0.027)
Precipitation spatial variability	-0.017	-0.022	-0.030	-0.077***	-0.094***	-0.111***
	(0.029)	(0.029)	(0.030)	(0.024)	(0.025)	(0.028)
Adjusted R^2	0.43	0.43	0.43	0.41	0.41	0.42
Mean of dep. var.	0.24	0.24	0.24	0.69	0.69	0.69
St. dev. of dep. var.	0.28	0.28	0.28	0.31	0.31	0.31
Observations	1,053	1,053	1,053	1,071	1,071	1,071
Region FE	✓	✓	✓	✓	✓	✓
Ethnographic controls	✓	✓	✓	✓	✓	✓
Geographic controls	✓	✓	✓	✓	✓	✓

Notes: The table reports OLS estimates. The unit of observation is a society from the *Ethnographic Atlas*. Columns (1)-(3) focus on state institutions (variable 33 in the *Ethnographic Atlas*), whereas columns (4)-(6) focus on kinship tightness (based on variables 8, 11, 15 and 23 in the *Ethnographic Atlas*). Volatility and spatial variability measures are based on climate data from the Climatic Research Unit (CRU). Panel A employs aggregate climate measures, while Panel B employs separate temperature and precipitation measures. Variables in columns (1) and (3) are constructed based on 50 km radius around the interior centroid of the society's location, as indicated on the *Ethnographic Atlas*. Columns (2) and (4) employ a 100 km radius, while columns (3) and (6) employ a 200 km radius. Region fixed effects include dummy variables for North America, Latin America and Caribbean, Sub-Saharan Africa, Middle East and North Africa, Europe and Central Asia, South Asia, East Asia and Pacific region. Ethnographic controls include settlement structure, intensity of agriculture, share of subsistence coming from hunting, gathering, agriculture and animal husbandry and time of ethnographic record. Geographic controls include the level of temperature and precipitation, absolute latitude, elevation, ruggedness, mean and standard deviation of land suitability. Reported coefficients have been standardised, therefore indicating the number of standard deviation changes in the dependent variable for a one–standard deviation change in the explanatory variable. Heteroskedasticity-robust standard errors in parentheses. ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.5. Historical Impact of Climate on State Institutions and Kin Ties: Spatially-Adjusted Standard Errors

Dependent variable:	*State Institutions*	*Kinship Tightness*
	(1)	(2)
	Panel A	
Climate volatility	0.100	-0.117
Spatially-adjusted s.e. - 100km	(0.052)**	(0.052)**
Spatially-adjusted s.e. - 200km	[0.054]*	[0.054]**
Spatially-adjusted s.e. - 500km	{0.057}*	{0.059}**
Climate spatial variability	-0.007	-0.082
Spatially-adjusted s.e. - 100km	(0.031)	(0.025)***
Spatially-adjusted s.e. - 200km	[0.032]	[0.026]***
Spatially-adjusted s.e. - 500km	{0.031}	{0.029}***
Adjusted R^2	0.43	0.42
	Panel B	
Temperature volatility	0.171	0.085
Spatially-adjusted s.e. - 100km	(0.077)**	(0.075)
Spatially-adjusted s.e. - 200km	[0.081]**	[0.078]
Spatially-adjusted s.e. - 500km	{0.091}*	{0.082}
Precipitation volatility	0.089	-0.136
Spatially-adjusted s.e. - 100km	(0.051)*	(0.052)***
Spatially-adjusted s.e. - 200km	[0.053]*	[0.054]***
Spatially-adjusted s.e. - 500km	{0.054}*	{0.060}**
Temperature spatial variability	0.061	0.022
Spatially-adjusted s.e. - 100km	(0.027)**	(0.024)
Spatially-adjusted s.e. - 200km	[0.028]**	[0.025]
Spatially-adjusted s.e. - 500km	{0.028}**	{0.027}
Precipitation spatial variability	-0.022	-0.094
Spatially-adjusted s.e. - 100km	(0.029)	(0.025)***
Spatially-adjusted s.e. - 200km	[0.029]	[0.026]***
Spatially-adjusted s.e. - 500km	{0.029}	{0.029}***
Adjusted R^2	0.44	0.42
Mean of dep. var.	0.24	0.69
St. dev. of dep. var.	0.28	0.31
Observations	1,053	1,071
Region FE	✓	✓
Ethnographic controls	✓	✓
Geographic controls	✓	✓

Notes: The table reports OLS estimates. The unit of observation is a society from the *Ethnographic Atlas*. Column (1) focuses on state institutions (variable 33 in the *Ethnographic Atlas*), whereas columns (2) focuses on kinship tightness (based on variables 8, 11, 15 and 23 in the *Ethnographic Atlas*). Volatility and spatial variability measures are based on climate data from the Climatic Research Unit (CRU). Panel A employs aggregate climate measures, while Panel B employs separate temperature and precipitation measures. Region fixed effects include dummy variables for North America, Latin America and Caribbean, Sub-Saharan Africa, Middle East and North Africa, Europe and Central Asia, South Asia, East Asia and Pacific region. Ethnographic controls include settlement structure, intensity of agriculture, share of subsistence coming from hunting, gathering, agriculture and animal husbandry and time of ethnographic record. Geographic controls include the level of temperature and precipitation, absolute latitude, elevation, ruggedness, mean and standard deviation of land suitability. Reported coefficients have been standardised, therefore indicating the number of standard deviation changes in the dependent variable for a one–standard deviation change in the explanatory variable. Standard errors adjusted for two-dimensional spatial correlation are employed, considering windows of 100 km (in parentheses), 200 km (in square brackets) and 500 km (in curly brackets). ***, **, and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Table A.6. Long-Run Impact of Ancestral Kinship Tightness on Institutional Quality: Period Estimates

Dependent variable:	*Institutional Quality*				
Period:	*1800-1850*	*1850-1900*	*1900-1950*	*1950-2000*	*1800-2000*
	(1)	(2)	(3)	(4)	(5)
	Panel A: Rule of Law				
Ancestral kinship tightness	-0.359	-0.414**	-0.304**	-0.228**	-0.309***
	(0.292)	(0.178)	(0.129)	(0.102)	(0.077)
Mean (s.d.) of dep. var.	0.46 (0.32)	0.50 (0.28)	0.50 (0.28)	0.55 (0.31)	0.50 (0.30)
Adjusted R^2	0.60	0.32	0.44	0.49	0.51
Observations	55	132	146	146	479
	Panel B: Property Rights				
Ancestral kinship tightness	-0.441	-0.271**	-0.227*	-0.091	-0.166***
	(0.407)	(0.126)	(0.127)	(0.082)	(0.063)
Mean (s.d.) of dep. var.	0.39 (0.22)	0.38 (0.26)	0.53 (0.27)	0.68 (0.22)	0.50 (0.28)
Adjusted R^2	0.11	0.48	0.34	0.39	0.56
Observations	58	132	146	146	482
	Panel C: Democracy				
Ancestral kinship tightness	-0.253	-0.346***	-0.308***	-0.250**	-0.220***
	(0.192)	(0.085)	(0.106)	(0.097)	(0.037)
Mean (s.d.) of dep. var.	0.19 (0.15)	0.18 (0.19)	0.37 (0.25)	0.55 (0.25)	0.33 (0.27)
Adjusted R^2	0.55	0.70	0.60	0.56	0.71
Observations	53	130	146	146	475
	Panel D: Control of Corruption				
Ancestral kinship tightness	-0.553*	-0.162	-0.244*	-0.235**	-0.236***
	(0.346)	(0.224)	(0.121)	(0.109)	(0.082)
Mean (s.d.) of dep. var.	0.49 (0.30)	0.59 (0.26)	0.52 (0.26)	0.47 (0.30)	0.52 (0.28)
Adjusted R^2	0.58	0.25	0.43	0.54	0.48
Observations	51	132	146	146	475
Continent FE	✓	✓	✓	✓	✓
Period FE					✓
Country-level controls	✓	✓	✓	✓	✓

Notes: The table reports OLS estimates. The unit of observation is a country. The dependent variable in columns (1)-(4) indicates the average institutional quality across 50-year intervals between 1800 and 2000, while column (5) employs a pooled regression including period fixed effects. The institutional quality measures are rule of law in Panel A, property rights in Panel B, level of democracy in Panel C and control of corruption in Panel D. The explanatory variable captures the average kinship tightness of the ancestors of the country's population. Country-level controls include ethnographic, geographic and contemporary controls. Ethnographic controls include ancestral dependence on hunting, gathering, animal husbandry and agriculture, ancestral complexity of settlements, average time of ethnographic record and years since the Neolithic Revolution. Geographic controls include absolute latitude, average elevation, average ruggedness, percentage of country's land in tropical and temperate climates, average distance to a coast or river, mean level of temperature and precipitation, average land quality, land area, an indicator for islands and an indicator for landlocked countries. Contemporary controls include population density, share of the population being Catholics, Protestants, Muslims or belonging to other religions. All specifications include continent fixed effects. Reported coefficients have been standardised, therefore indicating the number of standard deviation changes in the dependent variable for a one–standard deviation change in the independent variable. Heteroskedasticity-robust standard errors, clustered by country and period in column (5), are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.7. Long-Run Impact of Ancestral State Institutions on Institutional Quality: Period Estimates

Dependent variable:	*Institutional Quality*				
Period:	*1800-1850*	*1850-1900*	*1900-1950*	*1950-2000*	*1800-2000*
	(1)	(2)	(3)	(4)	(5)
	Panel A: Rule of Law				
Ancestral state institutions	0.172	0.254*	0.334***	0.258**	0.261***
	(0.338)	(0.122)	(0.097)	(0.128)	(0.067)
Mean (s.d.) of dep. var.	0.46 (0.32)	0.50 (0.28)	0.50 (0.28)	0.55 (0.31)	0.50 (0.30)
Adjusted R^2	0.58	0.30	0.45	0.50	0.50
Observations	55	132	146	146	479
	Panel B: Property Rights				
Ancestral state institutions	0.100	0.247*	0.325**	0.226	0.231***
	(0.275)	(0.118)	(0.123)	(0.107)	(0.071)
Mean (s.d.) of dep. var.	0.39 (0.22)	0.38 (0.26)	0.53 (0.27)	0.68 (0.22)	0.50 (0.28)
Adjusted R^2	0.08	0.48	0.36	0.41	0.57
Observations	58	132	146	146	482
	Panel C: Democracy				
Ancestral state institutions	0.001	0.126	0.151	0.035	0.099**
	(0.153)	(0.065)	(0.084)	(0.104)	(0.048)
Mean (s.d.) of dep. var.	0.19 (0.15)	0.18 (0.19)	0.37 (0.25)	0.55 (0.25)	0.33 (0.27)
Adjusted R^2	0.53	0.67	0.57	0.54	0.70
Observations	53	130	146	146	475
	Panel D: Control of Corruption				
Ancestral state institutions	0.185	0.077	0.169	0.232*	0.148**
	(0.409)	(0.138)	(0.104)	(0.124)	(0.067)
Mean (s.d.) of dep. var.	0.49 (0.30)	0.59 (0.26)	0.52 (0.26)	0.47 (0.30)	0.52 (0.28)
Adjusted R^2	0.52	0.25	0.42	0.54	0.47
Observations	51	132	146	146	475
Continent FE	✓	✓	✓	✓	✓
Period FE					✓
Country-level controls	✓	✓	✓	✓	✓

Notes: The table reports OLS estimates. The unit of observation is a country. The dependent variable in columns (1)-(4) indicates the average institutional quality across 50-year intervals between 1800 and 2000, while column (5) employs a pooled regression including period fixed effects. The institutional quality measures are rule of law in Panel A, property rights in Panel B, level of democracy in Panel C and control of corruption in Panel D. The explanatory variable indicates the average state institutions of the ancestors of the country's population. Country-level controls include ethnographic, geographic and contemporary controls. Ethnographic controls include ancestral dependence on hunting, gathering, animal husbandry and agriculture, ancestral complexity of settlements, average time of ethnographic record and years since the Neolithic Revolution. Geographic controls include absolute latitude, average elevation, average ruggedness, percentage of country's land in tropical and temperate climates, average distance to a coast or river, mean level of temperature and precipitation, average land quality, land area, an indicator for islands and an indicator for landlocked countries. Contemporary controls include population density, share of the population being Catholics, Protestants, Muslims or belonging to other religions. All specifications include continent fixed effects. Reported coefficients have been standardised, therefore indicating the number of standard deviation changes in the dependent variable for a one–standard deviation change in the independent variable. Heteroskedasticity-robust standard errors, clustered by country and period, are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.8. Long-Run Impact of Ancestral Kinship Tightness and Ancestral State Institutions on Institutional Quality: Period Estimates

Dependent variable:	*Institutional Quality*				
Period:	*1800-1850*	*1850-1900*	*1900-1950*	*1950-2000*	*1800-2000*
	(1)	(2)	(3)	(4)	(5)
	Panel A: Rule of Law				
Ancestral kinship tightness	-0.448	-0.440**	-0.325**	-0.246**	-0.332***
	(0.314)	(0.182)	(0.128)	(0.104)	(0.077)
Ancestral state institutions	0.264	0.287**	0.352***	0.272**	0.283***
	(0.336)	(0.115)	(0.093)	(0.129)	(0.061)
Mean (s.d.) of dep. var.	0.46 (0.32)	0.50 (0.28)	0.50 (0.28)	0.55 (0.31)	0.50 (0.30)
Adjusted R^2	0.61	0.35	0.48	0.52	0.53
Observations	55	132	146	146	479
	Panel B: Property Rights				
Ancestral kinship tightness	-0.526	-0.296**	-0.248**	-0.106	-0.186***
	(0.415)	(0.127)	(0.122)	(0.079)	(0.063)
Ancestral state institutions	0.203	0.269**	0.339**	0.233	0.243***
	(0.303)	(0.117)	(0.124)	(0.108)	(0.071)
Mean (s.d.) of dep. var.	0.39 (0.22)	0.38 (0.26)	0.53 (0.27)	0.68 (0.22)	0.50 (0.28)
Adjusted R^2	0.09	0.50	0.38	0.41	0.58
Observations	58	132	146	146	482
	Panel C: Democracy				
Ancestral kinship tightness	-0.263	-0.360***	-0.319***	-0.254**	-0.229***
	(0.191)	(0.087)	(0.106)	(0.098)	(0.038)
Ancestral state institutions	0.045	0.152*	0.169*	0.050	0.113**
	(0.138)	(0.059)	(0.082)	(0.104)	(0.045)
Mean (s.d.) of dep. var.	0.19 (0.15)	0.18 (0.19)	0.37 (0.25)	0.55 (0.25)	0.33 (0.27)
Adjusted R^2	0.52	0.71	0.60	0.56	0.72
Observations	53	130	146	146	475
	Panel D: Control of Corruption				
Ancestral kinship tightness	-0.677*	-0.170	-0.255**	-0.251**	-0.250***
	(0.456)	(0.223)	(0.121)	(0.110)	(0.080)
Ancestral state institutions	0.329	0.090	0.182	0.247**	0.165***
	(0.381)	(0.133)	(0.103)	(0.128)	(0.065)
Mean (s.d.) of dep. var.	0.49 (0.30)	0.59 (0.26)	0.52 (0.26)	0.47 (0.30)	0.52 (0.28)
Adjusted R^2	0.59	0.25	0.44	0.56	0.49
Observations	51	132	146	146	475
Continent FE	✓	✓	✓	✓	✓
Period FE					✓
Country-level controls	✓	✓	✓	✓	✓

Notes: The table reports OLS estimates. The unit of observation is a country. The dependent variable in columns (1)-(4) indicates the average institutional quality across 50-year intervals between 1800 and 2000, while column (5) employs a pooled regression including period fixed effects. The institutional quality measures are rule of law in Panel A, property rights in Panel B, level of democracy in Panel C and control of corruption in Panel D. The explanatory variables capture the average kinship tightness and average state institutions of the ancestors of the country's population. Country-level controls include ethnographic, geographic and contemporary controls. Ethnographic controls include ancestral dependence on hunting, gathering, animal husbandry and agriculture, ancestral complexity of settlements, average time of ethnographic record and years since the Neolithic Revolution. Geographic controls include absolute latitude, average elevation, average ruggedness, percentage of country's land in tropical and temperate climates, average distance to a coast or river, mean level of temperature and precipitation, average land quality, land area, an indicator for islands and an indicator for landlocked countries. Contemporary controls include population density, share of the population being Catholics, Protestants, Muslims or belonging to other religions. All specifications include continent fixed effects. Reported coefficients have been standardised, therefore indicating the number of standard deviation changes in the dependent variable for a one–standard deviation change in the independent variable. Heteroskedasticity-robust standard errors, clustered by country and period in column (5), are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.9. Long-Run Impact of Ancestral Kinship Tightness and Ancestral State Institutions on Rule of Law: Adding Controls

	Dependent variable: *Rule of Law 1950-2000*					
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A					
Ancestral kinship tightness	-0.499***	-0.289***	-0.333***	-0.323***	-0.228**	-0.227**
	(0.066)	(0.093)	(0.087)	(0.088)	(0.093)	(0.093)
Adjusted R^2	0.24	0.33	0.46	0.47	0.49	0.52
	Panel B					
Ancestral state institutions	0.273***	0.079	0.052	0.184	0.257**	0.205*
	(0.025)	(0.031)	(0.034)	(0.046)	(0.044)	(0.043)
Adjusted R^2	0.07	0.30	0.41	0.44	0.50	0.52
Mean of dep. var.	0.55	0.55	0.55	0.55	0.55	0.55
St. dev. of dep. var.	0.31	0.31	0.31	0.31	0.31	0.31
Observations	147	147	147	146	146	142
Continent FE		✓	✓	✓	✓	✓
Geographic controls			✓	✓	✓	✓
Ethnographic controls				✓	✓	✓
Contemporary controls					✓	✓
Ln(GDP per capita 1950)						✓

Notes: The table reports OLS estimates. The unit of observation is a country. The dependent variable is the average rule of law in the period 1950-2000. The explanatory variable in Panel A indicates the average kinship tightness of the ancestors of the country's population. The explanatory variable in Panel B indicates the average state institutions of the ancestors of the country's population. Country-level controls include ethnographic, geographic and contemporary controls. Ethnographic controls include ancestral dependence on hunting, gathering, animal husbandry and agriculture, ancestral complexity of settlements, average time of ethnographic record and years since the Neolithic Revolution. Geographic controls include absolute latitude, average elevation, average ruggedness, percentage of country's land in tropical and temperate climates, average distance to a coast or river, mean level of temperature and precipitation, average land quality, land area, an indicator for islands and an indicator for landlocked countries. Contemporary controls include population density, share of the population being Catholics, Protestants, Muslims or belonging to other religions. Reported coefficients have been standardised, therefore indicating the number of standard deviation changes in the dependent variable for a one–standard deviation change in the independent variable. Heteroskedasticity-robust standard errors in parentheses.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table A.10. Long-Run Impact of Ancestral Kinship Tightness and Ancestral State Institutions on Institutional Quality: Selection on Unobservables

Dependent variable:	*Rule of Law*		*Property Rights*		*Democracy*		*Control of Corruption*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A							
Anc. kinship tightness	-0.32***	-0.31***	-0.13	-0.17***	-0.20***	-0.22***	-0.28***	-0.24***
	(0.10)	(0.06)	(0.08)	(0.06)	(0.07)	(0.05)	(0.11)	(0.07)
R^2	0.30	0.54	0.22	0.59	0.29	0.73	0.25	0.52
Oster's δ		1.96		4.85		3.90		1.71
Oster's β		-0.29		-0.20		-0.25		-0.17
AET		46.73		-5.23		-9.76		5.68
	Panel B							
Anc. state institutions	0.09	0.26***	0.15**	0.23***	0.03	0.10**	0.03	0.15**
	(0.07)	(0.06)	(0.08)	(0.05)	(0.06)	(0.05)	(0.07)	(0.06)
R^2	0.27	0.54	0.22	0.60	0.27	0.72	0.22	0.51
Oster's δ		-4.78		7.42		-4.55		-2.76
Oster's β		0.57		0.36		0.22		0.36
AET		-1.57		-2.92		-1.39		-1.23
Observations	479	479	482	482	475	475	475	475
Continent FE	✓	✓	✓	✓	✓	✓	✓	✓
Country-level controls		✓		✓		✓		✓

Notes: The table reports OLS estimates. The unit of observation is a country. The dependent variable indicates the average institutional quality across 50-year intervals between 1800 and 2000, pooled together. The institutional quality measures are rule of law in columns (1)-(2), property rights in columns (3)-(4), level of democracy in columns (5)-(6) and control of corruption in columns (7)-(8). The explanatory variable in Panel A indicates the average kinship tightness of the ancestors of the country's population. The explanatory variable in Panel B indicates the average state institutions of the ancestors of the country's population. Country-level controls include ethnographic, geographic and contemporary controls. Ethnographic controls include ancestral dependence on hunting, gathering, animal husbandry and agriculture, ancestral complexity of settlements, average time of ethnographic record and years since the Neolithic Revolution. Geographic controls include absolute latitude, average elevation, average ruggedness, percentage of country's land in tropical and temperate climates, average distance to a coast or river, mean level of temperature and precipitation, average land quality, land area, an indicator for islands and an indicator for landlocked countries. Contemporary controls include population density, share of the population being Catholics, Protestants, Muslims or belonging to other religions. Reported coefficients have been standardised, therefore indicating the number of standard deviation changes in the dependent variable for a one–standard deviation change in the independent variable. Reported statistics show the robustness of results to selection by unobservables. AET ratio (Altonji, Elder, and Taber, 2005; Bellows and Miguel, 2009) and Oster's δ (Oster, 2016) measure how strongly correlated unobservables would have to be in order to account for the full size of the coefficient. Oster's β provides the estimated effect if the proportion of selection of observables and unobservables was to be equal, and the maximal R^2 equal to 1.3 times the observed R^2 (Oster, 2016). All three statistics in even-numbered columns employ odd-numbered columns as baseline specification. Heteroskedasticity-robust standard errors, clustered by country and period, are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Appendix B - Historical Volatility

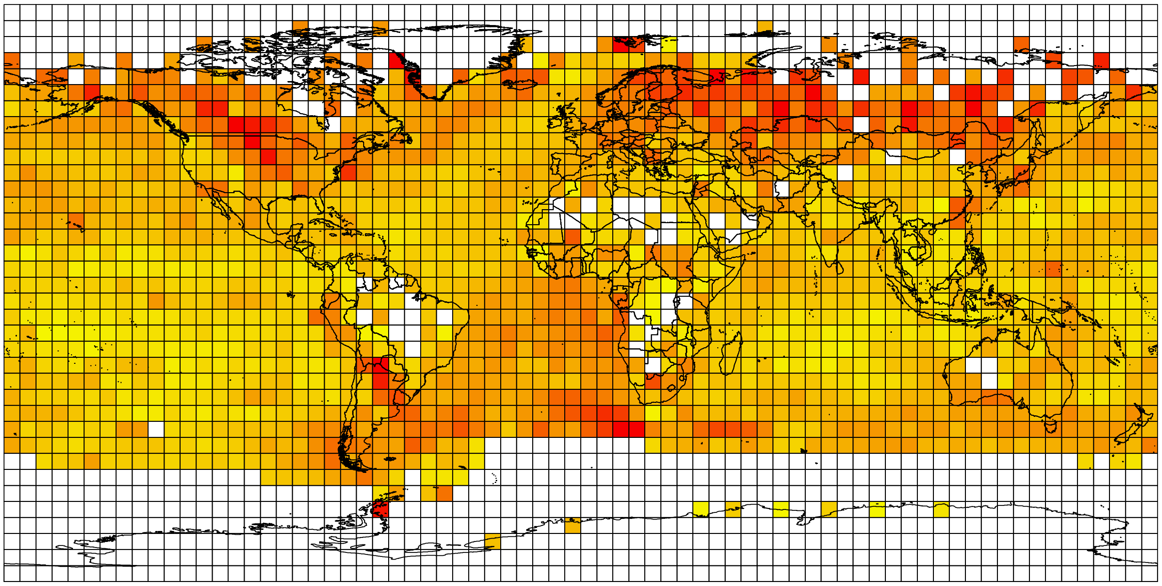
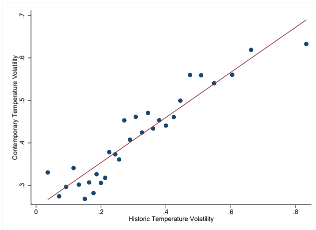


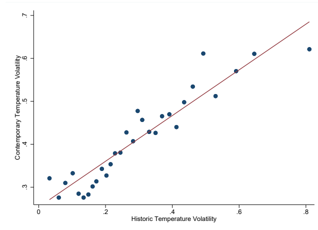
FIGURE B.1. Grid-cell-level measure of temperature variability across generations for the period 500-1900. Warmer colours indicate higher variability. White grid-cells indicate no data. Source: Mann, Zhang, Rutherford, Bradley, Hughes, Shindell, Ammann, Faluvegi, and Ni (2009)

To construct a measure of historical climate volatility, I employ paleoclimatic data from Mann, Zhang, Rutherford, Bradley, Hughes, Shindell, Ammann, Faluvegi, and Ni (2009), measuring temperature anomalies from the 5th century to the present. This data uses a climate field reconstruction approach to reconstruct global patterns of surface temperature, and has global coverage at a 5°× 5° resolution. With the dataset consisting of annual temperature anomalies, I construct a measure of historical temperature volatility across the previous centuries. First, I divide the sample of annual temperature anomalies into fourteen centuries going from 500 to 1900. Let $x_{i,y,c}$ denote the temperature measure in year y in century c for grid-cell i. Now let $\bar{x}_{i,c}$ denote the average temperature during century c in grid-cell i, The measure of climate variability for each previous century I construct is given by $\sigma_{i,c} = [\frac{1}{100}\sum_{y=1}^{100}(x_{i,y,c} - \bar{x}_{i,c})^2]^{\frac{1}{2}}$, which denotes the standard deviation of temperature across century c in grid-cell i. I also compute the overall temperature volatility for the past fourteen centuries using all the years in the period 500-1900, which I show in Figure B.1.

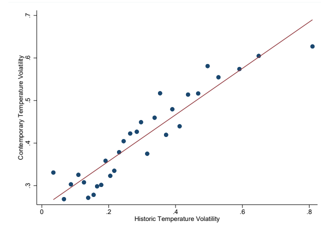
Figure B.2. Contemporary vs Historical Temperature Volatility



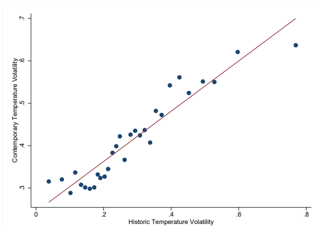
(a) 500-600 vs 1900-2017



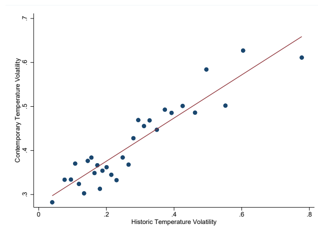
(b) 600-700 vs 1900-2017



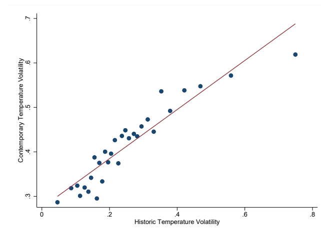
(c) 700-800 vs 1900-2017



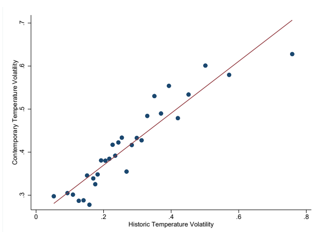
(d) 800-900 vs 1900-2017



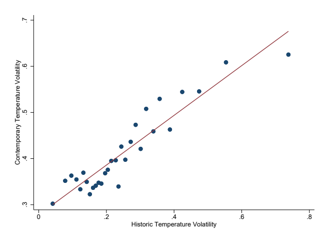
(e) 900-1000 vs 1900-2017



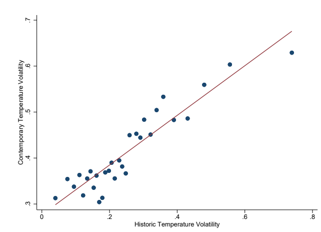
(f) 1000-1100 vs 1900-2017



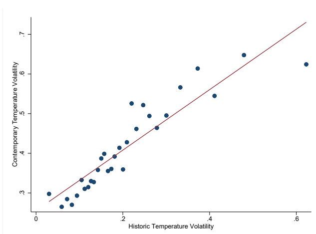
(g) 1100-1200 vs 1900-2017



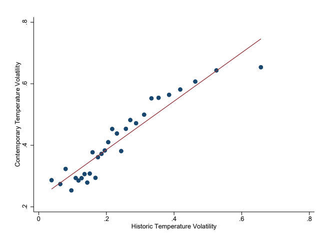
(h) 1200-1300 vs 1900-2017



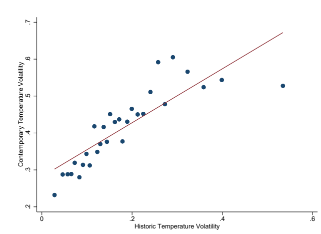
(i) 1300-1400 vs 1900-2017



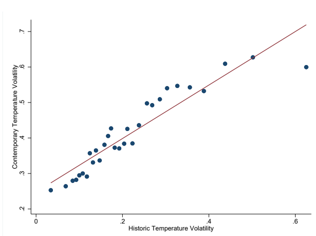
(j) 1400-1500 vs 1900-2017



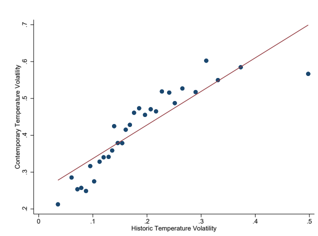
(k) 1500-1600 vs 1900-2017



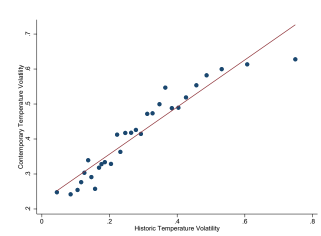
(l) 1600-1700 vs 1900-2017



(m) 1700-1800 vs 1900-2017



(n) 1800-1900 vs 1900-2017



(o) 500-1900 vs 1900-2017

Notes: Historical intertemporal temperature volatility based on paleoclimatic reconstructions of climate from Mann, Zhang, Rutherford, Bradley, Hughes, Shindell, Ammann, Faluvegi, and Ni (2009). Contemporary intertemporal temperature volatility based on the *Climatic Research Unit (CRU)* database.

BIBLIOGRAPHY

Acemoglu, D., S. Johnson, and J. A. Robinson (2001): “The Colonial Origins of Comparative Development: An Empirical Investigation,” *American Economic Review*, 91(5), 1369–1401.

——— (2002): “Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution,” *The Quarterly journal of economics*, 117(4), 1231–1294.

Acemoglu, D., and J. A. Robinson (2012): *Why nations fail: The origins of power, prosperity, and poverty.* Crown Publishing Group.

Aghion, P., Y. Algan, P. Cahuc, and A. Shleifer (2010): “Regulation and distrust,” *The Quarterly Journal of Economics*, 125(3), 1015–1049.

Alder, S., and I. Kondo (2019): “Political distortions and infrastructure networsk in China: A Quantitative Spatial Equilibrium Analysis,” mimeo, UNC Chapel Hill.

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

Alesina, A., and P. Giuliano (2011): “Family ties and political participation,” *Journal of the European Economic Association*, 9(5), 817–839.

——— (2014): “Family ties,” in *Handbook of economic growth*, vol. 2, pp. 177–215. Elsevier.

——— (2015): “Culture and institutions,” *Journal of Economic Literature*, 53(4), 898–944.

Alesina, A., P. Giuliano, and N. Nunn (2013): “On the origins of gender roles: Women and the plough,” *The Quarterly Journal of Economics*, 128(2), 469–530.

Alesina, A., and E. La Ferrara (2002): “Who trusts others?,” *Journal of public economics*, 85(2), 207–234.

Algan, Y., and P. Cahuc (2010): “Inherited trust and growth,” *The American Economic Review*, 100(5), 2060–2092.

Algan, Y., P. Cahuc, and A. Shleifer (2013): “Teaching practices and social capital,” *American Economic Journal: Applied Economics*, 5(3), 189–210.

Almond, G., and S. Verba (1963): “The civic culture: political attitudes and democracy in five countries,” *Princeton University Press.*

Alsan, M. (2015): “The effect of the tsetse fly on African development,” *American Economic Review*, 105(1), 382–410.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005): “Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools,” *Journal of political economy*, 113(1), 151–184.

Ashraf, Q., and S. Michalopoulos (2015): “Climatic fluctuations and the diffusion of agriculture," *Review of Economics and Statistics*, 97(3), 589–609.

Banerjee, A., E. Duflo, and N. Qian (2020): "On the road: Access to transportation infrastructure and economic growth in China," *Journal of Development Economics*, 145.

Banfield, E. C. (1958): *The moral basis of a backward society.* New York: Simon and Schuster, Free Press.

Bates, R. H. (1987): *Essays on the political economy of rural Africa*, vol. 38. University of California Press.

Becker, S. O., K. Boeckh, C. Hainz, and L. Woessmann (2016): "The empire is dead, long live the empire! Long-run persistence of trust and corruption in the bureaucracy," *The Economic Journal*, 126(590), 40–74.

Bellows, J., and E. Miguel (2009): "War and local collective action in Sierra Leone," *Journal of public Economics*, 93(11-12), 1144–1157.

BenYishay, A., R. Rotberg, J. Wells, Z. Lv, S. Goodman, L. Kovacevic, and D. Runfola (2017): "Geocoding afrobarometer rounds 1-6: Methodology & data quality," *AidData.*

Berger, T., and K. Enflo (2017): "Locomotives of local growth: The short-and long-term impact of railroads in Sweden," *Journal of Urban Economics*, 98, 124–138.

Besley, T. (2006): *Principled agents? The political economy of good government.* Oxford University Press.

Bisin, A., and T. Verdier (2001): "The economics of cultural transmission and the dynamics of preferences," *Journal of Economic Theory*, 97(2), 298–319.

——— (2017): "On the joint evolution of culture and institutions," *NBER Working Paper.*

Bockstette, V., A. Chanda, and L. Putterman (2002): "States and markets: The advantage of an early start," *Journal of Economic Growth*, 7(4), 347–369.

Boyd, R., and P. J. Richerson (1985): *Culture and the evolutionary process.* Chicago: University of Chicago Press.

——— (1995): "Why does culture increase human adaptability?," *Ethology and sociobiology*, 16(2), 125–143.

——— (2005): *The origin and evolution of cultures.* New York: Oxford University Press.

Brunialti, A. (1882): *Legge Elettorale Politica - commentata dal Dr Attilio Brunialti.* Unione Tipografico-Editrice, Torino-Roma-Napoli.

Büchel, K., and S. Kyburz (2020): "Fast track to growth? Railway access, population growth and local displacement in 19th century Switzerland," *Journal of Economic Geography*, 20(1), 155–195.

Buggle, J., and R. Durante (2017): "Climate risk, cooperation, and the co-evolution of culture and institutions," *CEPR Discussion Paper No. DP12380*.

Burgess, R., R. Jedwab, E. Miguel, A. Morjaria, and G. P. i Miquel (2015): "The value of democracy: evidence from road building in Kenya," *The American Economic Review*, 105, 1817–1851.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust nonparametric confidence intervals for regression-discontinuity designs," *Econometrica*, 82(6), 2295–2326.

Chandler, T. (1987): *Four thousand years of urban growth: an historical census.* Lewiston New York St. Davids University Press.

Ciccarelli, C., and P. Groote (2017): "Railway Endowment in Italy's Provinces, 1839-1913," *Rivista di storia economica*, 33(1), 45–88.

Conley, T. G. (1999): "GMM estimation with cross sectional dependence," *Journal of econometrics*, 92(1), 1–45.

Coppedge, M., J. Gerring, C. H. Knutsen, S. I. Lindberg, S.-E. Skaaning, J. Teorell, D. Altman, M. Bernhard, M. S. Fish, A. Cornell, et al. (2018): "V-Dem Dataset v8," .

Corbetta, P., and M. S. Piretti (2009): *Atlante storico-elettorale d'Italia: 1861-2008.* Zanichelli.

Curto–Grau, M., A. Herranz–Loncan, and A. Sole–Olle (2012): "Pork–Barrell Politics in Semi-Democracies: The Spanish "Parliamentary Roads"," *Journal of Economic History*, 72, 771–796.

Deconinck, K., and M. Verpoorten (2013): "Narrow and scientific replication of 'The slave trade and the origins of mistrust in Africa'," *Journal of Applied Econometrics*, 28(1), 166–169.

Dell, M. (2010): "The persistent effects of Peru's mining mita," *Econometrica*, 78(6), 1863–1903.

Depetris-Chauvin, E. (2015): "State history and contemporary conflict: Evidence from sub-saharan africa," *Available at SSRN 2679594*.

Diamond, J. (1987): "The Worst Mistake in the History of the Human Race," *Discover*.

——— (1997): *Germs, guns, and steel: the fates of human societies.* New York: Norton.

Direzione Generale della Statistica (1886): *Risultati dell'Inchiesta sulle Condizioni Igieniche e Sanitarie nei Comuni del Regno - Parte Seconda.* Tip. nell'Ospizio di San Michele, Roma.

Durante, R. (2009): "Risk, cooperation and the economic origins of social trust: an empirical investigation," *Available at SSRN 1576774*.

Easterly, W., and R. Levine (1997): "Africa's growth tragedy: policies and ethnic divisions," *The Quarterly Journal of Economics*, 112(4), 1203–1250.

Easton, D. (1965): *A systems analysis of political life*. New York: Wiley.

Engerman, S. L., and K. L. Sokoloff (2000): "History lessons: Institutions, factors endowments, and paths of development in the new world," *The Journal of Economic Perspectives*, 14(3), 217–232.

Englebert, P. (2000): "Pre-colonial institutions, post-colonial states, and economic development in tropical Africa," *Political research quarterly*, 53(1), 7–36.

Enke, B. (2019): "Kinship, cooperation, and the evolution of moral systems," *The Quarterly Journal of Economics*, 134(2), 953–1019.

Fenoaltea, S. (1983): "Italy," in *Railways and the Economic Development of Western Europe, 1830-1914*, ed. by P. O'Brien. The MacMillan Press Ltd, London.

Fenske, J. (2013): "Does land abundance explain African institutions?," *The Economic Journal*, 123(573), 1363–1390.

——— (2014): "Ecology, trade, and states in pre-colonial Africa," *Journal of the European Economic Association*, 12(3), 612–640.

Fernandez, R. (2007): "Women, work, and culture," *Journal of the European Economic Association*, 5(2-3), 305–332.

Fernandez, R., and A. Fogli (2009): "Culture: An empirical investigation of beliefs, work, and fertility," *American Economic Journal: Macroeconomics*, 1(1), 146–177.

Ferrucci, A. (1898): *Delle Spese Straordinarie dello Stato per le Ferrovie e dei Relativi Utili*. Tipografia dei Fratelli Bencini, Roma-Firenze.

Fukuyama, F. (1995): *Trust: The social virtues and the creation of prosperity*. Free Press.

Galor, O., and M. Klemp (2017): "Roots of autocracy," *NBER Working Paper*.

Galor, O., and Ö. Özak (2016): "The agricultural origins of time preference," *American Economic Review*, 106(10), 3064–3103.

Galor, O., and V. Savitskiy (2018): "Climatic Roots of Loss Aversion," .

Gennaioli, N., and I. Rainer (2006): "Precolonial Centralization and Institutional Quality in Africa," *Institutions and norms in economic development*, pp. 21–46.

——— (2007): "The modern impact of precolonial centralization in Africa," *Journal of Economic Growth*, 12(3), 185–234.

Gershman, B. (2017): "Long-run development and the new cultural economics," *Demographic Change and Long-Run Development*, p. 221.

Giavazzi, F., I. Petkov, and F. Schiantarelli (2014): "Culture: Persistence and evolution," *NBER Working Paper*.

Giuliano, P. (2007): "Living arrangements in western europe: Does cultural origin matter?," *Journal of the European Economic Association*, 5(5), 927–952.

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

——— (2017): "Understanding cultural persistence and Change," *NBER Working Paper*.

——— (2018): "Ancestral characteristics of modern populations," *Economic History of Developing Regions*, 33(1), 1–17.

Golden, M., and B. Min (2013): "Distributive Politics around the World," *Annual Review of Political Science*, 16, 73–99.

Gorodnichenko, Y., and G. Roland (2017): "Culture, institutions, and the wealth of nations," *Review of Economics and Statistics*, 99(3), 402–416.

Greene, W. H. (2003): *Econometric analysis*. Pearson.

Greif, A. (1994): "Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies," *Journal of political economy*, 102(5), 912–950.

Greif, A., and G. Tabellini (2017): "The clan and the corporation: Sustaining cooperation in China and Europe," *Journal of Comparative Economics*, 45(1), 1–35.

Guiso, L., P. Sapienza, and L. Zingales (2006): "Does Culture Affect Economic Outcomes?," *The Journal of Economic Perspectives*, 20(2), 23–48.

——— (2016): "Long-term persistence," *Journal of the European Economic Association*, 14(6), 1401–1436.

Harari, N. Y. (2015): *Sapiens: A brief history of humankind*. New York: HarperCollins.

Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister (2014): "Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset," *International journal of climatology*, 34(3), 623–642.

Heldring, L. (2018): "The Origins of Violence in Rwanda," .

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, R. McElreath, M. Alvard, A. Barr, J. Ensminger, et al. (2005): ""Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies," *Behavioral and brain sciences*, 28(6), 795–815.

Henrich, J. P. (2004): *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford University Press.

Herbst, J. (2000): *States and power in Africa: Comparative lessons in authority and control*. Princeton University Press.

Hornung, E. (2015): "Railroads and Growth in Prussia," *Journal of the European Economic Association*, 13(4), 699–736.

Houghton, A. B. (1889): "Italian Finances from 1860 to 1884," *Quarterly Journal of Economics*, 3, 373–402.

Istat (1957): *Indagine statistica sullo sviluppo del reddito nazionale dell'Italia dal 1861 al 1956*. Istituto Centrale di Statistica.

Jedwab, R., and A. Moradi (2016): "The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa," *Review of Economics and Statistics*, 89, 268–284.

Kiszewski, A., A. Mellinger, A. Spielman, P. Malaney, S. E. Sachs, and J. Sachs (2004): "A global index representing the stability of malaria transmission," *The American journal of tropical medicine and hygiene*, 70(5), 486–498.

Knight, B. (2005): "Estimating the Value of Proposal Power," *The American Economic Review*, 95, 1639–1652.

Larcinese, V., J. M. Snyder, and C. Testa (2013): "Testing Models of Distributive Politics using Exit Polls to Measure Voters Preferences and Partisanship," *British Journal of Political Science*, 43, 845–875.

Lewis, W. A. (1954): "Economic development with unlimited supplies of labour," *The manchester school*, 22(2), 139–191.

Lipset, S. M. (1959): "Some social requisites of democracy: Economic development and political legitimacy," *American political science review*, 53(1), 69–105.

Litina, A. (2014): "The geographical origins of early state formation," .

Lowes, S. (2018): "Matrilineal kinship and spousal cooperation: evidence from the Matrilineal belt," *Working Paper*.

Lowes, S., N. Nunn, J. A. Robinson, and J. L. Weigel (2017): "The evolution of culture and institutions: Evidence from the Kuba kingdom," *Econometrica*, 85(4), 1065–1091.

Maggi, S. (2003): *Le ferrovie*. Il Mulino, Bologna.

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

Marshall, J. (2016): "Coarsening bias: How coarse treatment measurement upwardly biases instrumental variable estimates," *Political Analysis*, 24(2), 157–171.

Mayshar, J., O. Moavz, Z. Neeman, and L. Pascali (2016): "Cereals, appropriability and hierarchy," .

McCrary, J. (2008): "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of econometrics*, 142(2), 698–714.

Mercurio, F. (1994): "Le ferrovie e il Mezzogiorno: i vincoli morali e le gerarchie territoriali," *Meridiana*, 19, 155–193.

Michalopoulos, S. (2012): "The origins of ethnolinguistic diversity," *The American economic review*, 102(4), 1508.

Michalopoulos, S., and E. Papaioannou (2013): "Pre-Colonial Ethnic Institutions and Contemporary African Development," *Econometrica*, 81(1), 113–152.

——— (2014): "National Institutions and Subnational Development in Africa," *The Quarterly journal of economics*, 129(1), 151–213.

——— (2016): "The long-run effects of the scramble for Africa," *American Economic Review*, 106(7), 1802–48.

Michalopoulos, S., L. Putterman, and D. N. Weil (2019): "The influence of ancestral lifeways on individual economic outcomes in Sub-Saharan Africa," *Journal of the European Economic Association*, 17(4), 1186–1231.

Ministero dei Lavori Pubblici (1885): *Relazione Statistica sulle Costruzioni e sull'Esercizio delle Strade Ferrate Italiane per l'Anno 1884*. Tipografia Eredi Botta, Roma.

——— (1889): *Relazione sulle Costruzioni e sull'Esercizio delle Strade Ferrate Italiane per gli Anni 1885-1886-1887 - Volume I*. Tipografia Nazionale G. Bertero, Roma.

——— (1891): *Relazione sulle Costruzioni e sull'Esercizio delle Strade Ferrate Italiane per gli Anni 1888-1889-1890 - Volume II*. Tipografia Nazionale G. Bertero, Roma.

Ministero dell'Educazione Nazionale (1865): *Statistica del Regno d'Italia - Biblioteche, Anno 1863*. Tipografia del Successori Le Monnier, Firenze.

——— (1866): *Statistica del Regno d'Italia - Istruzione Pubblica e Privata, Parte II e III - Ginnasi, Licei e Scuole Superiori, Anno Scolastico 1862-63; Istituti Superiori, Anno Scolastico 1863-64*. Tipografia Tofani, Firenze.

Ministero dell'Interno (1874): *Dizionario dei comuni del Regno D'Italia, 1871*. Eredi Botta, Roma.

Ministero di Agricoltura, Industria e Commercio (1884): *Annuario Statistico Italiano*. Tipografia Eredi Botta, Roma.

——— (1886): *Annuario Statistico Italiano*. Tipografia Eredi Botta, Roma.

——— (1887): *Bilanci Comunali per l'Anno 1884 - Parte II*. Stabilimento Tipografico dell'Opinione, Roma.

——— (1899): *Bilanci Comunali per l'Anno 1897*. Stabilimento Tipografico di G. Berterio, Roma.

Moscona, J., N. Nunn, and J. A. Robinson (2017): "Keeping it in the family: Lineage organization and the scope of trust in sub-Saharan Africa," *American Economic Review: Papers & Proceedings*, 107(5), 565–71.

Murdock, G. P. (1959): *Africa: its peoples and their culture history*. New York: McGraw-Hill.

——— (1967): *Ethnographic atlas*. Pittsburgh: University of Pittsburgh Press.

North, D. C. (1990): *Institutions, institutional change and economic performance*. Cambridge university press.

Nunn, N. (2008): "The long-term effects of Africa's slave trades," *The Quarterly Journal of Economics*, 123(1), 139–176.

——— (2012): "Culture and the historical process," *Economic History of Developing Regions*, 27(sup1), S108–S126.

——— (2014): "Historical development," in *Handbook of economic growth*, vol. 2, pp. 347–402. Elsevier.

Nunn, N., and D. Puga (2012): "Ruggedness: The blessing of bad geography in Africa," *Review of Economics and Statistics*, 94(1), 20–36.

Nunn, N., and L. Wantchekon (2011): "The Slave Trade and the Origins of Mistrust in Africa," *American Economic Review*, 101, 3221–3252.

Nuvoloni, L. (1898): *Le Elezioni Politiche al Parlamento subalpino e al Parlamento Italiano*. Tipografia della Camera dei Deputati, Roma.

Oster, E. (2016): "Unobservable selection and coefficient stability: Theory and evidence," *Journal of Business & Economic Statistics*, 37(2), 187–204.

Putnam, R. D. (1993): *Making democracy work: Civic traditions in modern Italy*. Princeton University Press.

——— (2000): *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.

Redding, S. J., and M. A. Turner (2015): "Transportation costs and the spatial organization of economic activity," in *Handbook of regional and urban economics*, vol. 5, pp. 1339–1398. Elsevier.

Schram, A. (1997): *Railways and the formation of the Italian state in the nineteenth century*. Cambridge University Press, Cambridge, United Kingdom.

Schulz, J. (2017): "The Churches' Bans on Consanguineous Marriages, Kin-Networks and Democracy," *Working Paper*.

Selod, H., and S. Soumahoro (2019): "Highway politics in a divided government," Working Paper 8710, World Bank Group.

Sistema Statistico Nazionale (1994): *Popolazione Residente dei Comuni: Censimenti dal 18961 al 1991*. ISTAT, Rome, Italy.

Tabellini, G. (2008a): "Institutions and culture," *Journal of the European Economic Association*, 6(2-3), 255–294.

——— (2008b): "The Scope of Cooperation: Values and Incentives," *The Quarterly journal of economics*, 123(3), 905–950.

——— (2010): "Culture and institutions: economic development in the regions of Europe," *Journal of the European Economic Association*, 8(4), 677–716.

Tajani, F. (1944): *Storia delle ferrovie italiane: a 100 anni dall'apertura della prima linea*. Garzanti, Milano.

Todd, E. (1983): *The explanation of ideology: Family structures and social systems*. New York: Blackwell.

Vogt, M., N.-C. Bormann, S. Rüegger, L.-E. Cederman, P. Hunziker, and L. Girardin (2015): "Integrating data on ethnicity, geography, and conflict: The ethnic power relations data set family," *Journal of Conflict Resolution*, 59(7), 1327–1342.

Voigtländer, N., and H.-J. Voth (2012): "Persecution perpetuated: the medieval origins of anti-Semitic violence in Nazi Germany," *The Quarterly Journal of Economics*, 127(3), 1339–1392.