Labor Scarcity, Land Tenure, and Historical Legacy: Evidence from Mexico

Emily A. Sellars* Jennifer Alix-Garcia†

Acknowledgements: We are grateful to Barbara Forrest for pointing out the existence of *cocolitzli*. For detailed comments on this manuscript, we thank Francisco Garfias, Scott Gehlbach, Horacio Larreguy, Maggie Peters, Laura Schechter, and Felicity Vabulas. We also thank seminar participants at the American Political Science Association Annual Meeting, the Bureau for Research and Economic Analysis of Development Conference, the Midwest International Economic Development Conference, the Triangle Research and Environmental Economics (TREE) seminar, Oregon State University, the Harris Political Economy Lunch, and the University of Wisconsin Development Workshop and Comparative Politics Colloquium.

*Department of Political Science, Yale University, 115 Prospect St, New Haven, CT, 06511, email: emily.sellars@yale.edu, phone: 203-432-5230

†Department of Applied Economics, Oregon State University, 213 Ballard Extension Hall, Corvallis, OR, 97331; email: jennifer.alix-garcia@oregonstate.edu, phone: 541-737-2691

Labor Scarcity, Land Tenure, and Historical Legacy: Evidence from Mexico

Abstract

How does labor scarcity shape institutional development? We examine the long-term impact of Mexico's 16th-century demographic collapse on landholding patterns through the present day. We show that the collapse, which reduced the indigenous population by between 70 and 90 percent and differed in severity across space, facilitated land concentration and the rise of a landowner class that dominated Mexican political economy for centuries. To identify the impact of the population collapse on land arrangements we use an instrumental-variables identification strategy exploiting climatic shocks associated with an epidemic in the 1570s that roughly halved the indigenous population. Though landholding patterns were transformed following the Revolution of 1910, we also show that disease-impacted areas now have more land in common-property institutions with limited property rights. The population collapse thus had a persistent impact on Mexican political economy, but the nature of this impact has changed radically over time.

JEL codes: O12, Q15, N26

Keywords: economic history, institutional development, population dynamics, land tenure

1 Introduction

Historical events and institutions have long-reaching political and economic effects. The institutions developed under colonial rule in particular were crucial in establishing contemporary development trajectories (e.g., La Porta et al. 1997; Sokoloff and Engerman 2000; Nunn 2009; Mahoney 2010; Acemoglu and Robinson 2012). However, considerable disagreement exists over why advantageous or pernicious institutions were established in different places and how the impacts of these institutions have persisted or changed over time.

One important debate in this literature centers on the role of demographic conditions in driving institutional development. Differences in population density have been central to theories explaining long-term underdevelopment in Africa, the emergence and decline of slavery, the rise of the European nation-state, the survival of democracy in the United States, the structure of property-rights institutions, and many other outcomes (e.g., Domar 1970; Herbst 2000; Epstein 2000; Boone 2003; Levine 2006; Nunn 2008; Fenske 2013; Voigtländer and Voth 2013*b*). However, these theories differ in whether relative population density (e.g., Herbst 2000; Fenske 2013) or scarcity (e.g., Levine 2006; Voigtländer and Voth 2013*b*) is associated with better long-term development outcomes. Moreover, disagreement continues about the extent to which sudden demographic shocks—such as occur following large-scale migration flows, wars, or epidemics—shape institutional development long after the population recovers.

In this paper, we examine these questions by tracing the evolving impact of a massive population collapse in 16th-century Mexico on landholding patterns through the present day. It has often been argued that the relative population density of colonial Mexico led to the creation of pernicious institutions with adverse consequences for long-term development (Engerman and Sokoloff 1997; Sokoloff and Engerman 2000; Acemoglu, Johnson and Robinson 2002; Mahoney 2010; Acemoglu and Robinson 2012). However, this generalization masks profound spatial and temporal variation in population dynamics, institutions, and development trajectories. Though central Mexico was densely populated at the time of the

Spanish Conquest, the population fell by between 70 and 90 percent in the first century of colonial rule, mostly through a series of severe epidemics (Cook and Borah 1971; Hassig 1985).[1] As a comparison, the Black Plague—the linchpin of many explanations of European political development—is thought to have reduced Europe's population by between 30 and 60 percent (Levine 2006; Voigtländer and Voth 2013*b*). The cataclysmic demographic collapse dramatically altered Mexico's economic and political trajectory. The *encomienda*, the extractive institution highlighted in many accounts of Mexican underdevelopment (e.g., Acemoglu and Robinson 2012), had essentially disappeared in central Mexico less than 50 years after the Conquest (Gibson 1964; Hassig 1985; Knight 2002). In its place, the Spanish adopted a series of institutions aimed at controlling land and labor in a context of severe population scarcity. According to numerous historians of Mexico, the labor scarcity caused by the demographic collapse—not merely labor abundance at the time of Conquest—led to the development of perverse political and economic institutions (e.g., Borah 1951; Gibson 1964; Knight 2002).

We demonstrate that Mexico's population collapse had a significant and persistent, but changing, impact on the country's landholding patterns. In areas depopulated by disease in the early colonial period, settlers were able to amass large quantities of land from indigenous villages, and this early land concentration built on itself over time. We show that just prior to the Revolution of 1910, a higher percentage of the rural population lived in communities on large estates (haciendas) in areas that experienced a more precipitous decline in population between 1570 and 1650. To address empirical concerns related to omitted-variables bias, measurement error, and reverse causality, we use an instrumental-variables empirical strategy based on the characteristics of a massive epidemic in the mid-1570s. This epidemic, which roughly halved the indigenous population, is believed to have been caused by a rodent-transmitted pathogen that emerged after several years of drought were followed by a period of above-average rainfall (Acuña Soto, Calderon Romero and Maguire 2000;

[1]For a summary of colonial population estimates, see Appendix Table A1.

Acuña Soto et al. 2002). We construct proxies for these climate conditions using precipitation data dating back to the early 1500s extracted from tree-ring chronologies (Cook and Krusic 2004). We use these proxies to predict the fall in population between the first comprehensive population count in 1570 and the nadir of the indigenous population in 1650. These climate data capture deviations from normal conditions for each point in space and are arguably orthogonal to geographic, political, and economic confounds. Our results are robust under numerous empirical specifications. They are also supported by an extensive qualitative historical literature highlighting the relationship between the population collapse and the formation of the hacienda (e.g., Chevalier 1963; Florescano 1976; Knight 2002). We are the first, to our knowledge, to examine this assertion quantitatively.

Though Mexico's land arrangements and rural political economy were transformed during the 20th century, we show that disease-impacted areas continue to follow a unique development trajectory. After the Revolution, Mexico undertook one of the most significant land reform programs in history, through which much of the country's surface area was redistributed from large estates to peasant villages in the form of *ejidos*, common-property institutions with limited property rights protections and legal constraints on landholding and land use. Because this reform disproportionately targeted regions of previously high land concentration, areas that lost more population in the early colonial period saw more land redistributed into ejidos than areas where the collapse had been less severe. As a consequence, disease-impacted areas now have more land in these unique common-property institutions than in areas where the collapse was less severe.

By rigorously linking the demographic collapse to past and present land institutions in Mexico, this paper highlights an important channel through which population shocks in the early colonial period continue to shape development. In Mexico, as in much of Latin America, land ownership has been a critical determinant of past and present political and economic power (Engerman and Sokoloff 1997; Sokoloff and Engerman 2000; Baland and Robinson 2008). Furthermore, an extensive literature has demonstrated how land rights and

landholding patterns determine patterns of investment and production, and thus economic development (e.g., Banerjee and Iyer 2005; Goldstein and Udry 2008; Fenske 2013). The specific land institutions we examine, the hacienda and the ejido, have been independently blamed for persistent underdevelopment in Mexico. However, the mechanisms through which each institution is thought to impede development are different. Whereas the hacienda system is blamed for fostering the rise of a dominant landowner class that perpetuated economic and political inequality (e.g., Engerman and Sokoloff 1997; Sokoloff and Engerman 2000), the ejido, introduced in part to address this inequality, is thought to impede development through institutionalized land fragmentation and the incomplete property rights granted to smallholders (e.g., Yates 1981; Larreguy 2013; Albertus et al. 2015). In tying both institutions to the much earlier population collapse, this paper illustrates how the legacy of the collapse continues to affect development, though the nature of its impact has changed dramatically.

The link between the population collapse, the hacienda, and the ejido also has implications for the political economy of population scarcity or abundance. Theoretical models examining the interplay between labor scarcity, outside labor options, and coercive institutions have illustrated that a fall in population does not necessarily improve the lot of labor (e.g., Domar 1970; Conning 2007; Acemoglu and Wolitzky 2011). Building on this work, we highlight two mechanisms through which population scarcity can enable and encourage land concentration by elites: by reducing the cost of acquiring land from smallholders and by increasing the benefits of controlling local labor markets once initial land concentration has been established. Whereas in Europe depopulation due to the Plague increased the bargaining power of labor and set the continent on a trajectory of prosperity (e.g., Levine 2006; Voigtländer and Voth 2013*b*), elites in Mexico used their pre-established political and economic power to obtain large quantities of land while the indigenous population was weakened by disease, forcing surviving workers into employment on large estates. Evidence suggests that this early land concentration increased over time as large landowners sought to ac-

quire additional land around their estates to deprive peasant workers of outside earnings opportunities and, later, to take advantage of export opportunities abroad (e.g., Coatsworth 1974; Conning 2007). In areas where the collapse was less severe, by contrast, indigenous villages were more able to maintain control of community lands through formal and informal protections that hindered encroachment by colonists (e.g., Florescano 1976; Knight 2002; Franco Vivanco 2018). This underscores the importance of prior institutional context in determining the political and economic consequences of sudden population scarcity and contrasts with the argument that a fall in population caused by wars or epidemics will necessarily lead to subsequent gains in productivity or income through direct labor-market effects on survivors (e.g., Young 2005).

Our work further highlights that even short-term labor scarcity may produce long-term institutional consequences. Though Mexico's population collapse was extreme, populations had largely rebounded by the end of the colonial period (Knight 2002). Furthermore, despite substantial variation in the severity of the collapse, the regional distribution of population largely returned to pre-collapse levels as well. The correlation coefficient between population density in 1570 (our baseline year) and that of 1800 (the end of the colonial period) in our dataset is 0.83. However, the drop in population in the critical 1570 to 1650 period has strong predictive power over land institutions centuries later, even when controlling for both baseline and 20th-century population density. This demonstrates that the indirect, institutional effects of a population shock may reverberate long after demographic patterns have returned to normal.

It is perhaps unsurprising that a massive demographic shock during a formative period of institutional development would have long-reaching impacts when even small shocks at critical moments can have persistent consequences (Acharya and Lee 2016). What is more interesting is that we can continue to trace the impacts of the collapse after the long-run political control of landowners had been subverted. While the Mexican Revolution arguably represents an example of rapid and "discontinuous" change in political institutions (Roland

2004), the subsequent land reform targeted long-running agrarian grievances and was tied directly to the legacy of the hacienda. While rural political economy has changed dramatically since 1910, institutional structures remain distinct in disease-impacted areas. Moreover, while it is often argued that the impact of historical events deteriorates over time (e.g., Banerjee and Iyer 2005; Nunn 2008), this study illustrates that this is not the only possible trajectory. Early patterns of land concentration during the 16th-century collapse were amplified in during the 19th-century economic boom before being radically transformed through the agrarian reform. Many papers on historical legacy associate a long-ago historical event or institution with a contemporary outcome, but this article highlights why one should be cautious in assuming the continuity of political and institutional processes between two snapshots of history (Austin 2008). This is important to remember when studying colonial institutions and development (e.g., La Porta et al. 1997; Acemoglu, Johnson and Robinson 2002; Mahoney 2010) or institutional persistence and change more generally (e.g., North 1990; Pierson 2000; Roland 2004; Thelen 2004; Mahoney and Thelen 2010).

2 Population scarcity and institutions

Demography plays a central role in many theories of institutional development. An extensive literature, dating back to the writings of Malthus in the seventeenth century, has argued that "overpopulation" encourages resource competition and conflict, potentially undermining long-run stability and growth of dense areas (e.g., Dasgupta 1995; Homer-Dixon 1999). Other work has examined the ways in which the concentration or dispersion of population shaped the short- and long-term development of political and economic institutions. One influential argument holds that population patterns at the time of colonization determined whether regions received pernicious institutions aimed at short-term wealth extraction (as occurred in dense areas) or beneficial institutions aimed at long-run investment and settlement (as occurred in sparse areas) (e.g., Engerman and Sokoloff 1997; Acemoglu, Johnson

and Robinson 2002). Beyond its effects on colonial institutions, high population density in early development has been liked with the emergence of a landed elite class, who could take advantage of labor abundance and then adopt institutions to perpetuate their political and economic dominance over time. This is one of the reasons cited for differences in political development in Latin America, and in particular why democracy was more likely to emerge and succeed in population-scarce areas.[2] A similar argument is made with respect to the United States, where the existence of a sparsely populated western "frontier" is argued to have reduced the economic and political dominance of early landed elites, promoting the development of democracy (e.g., Turner 1920). The converse of this argument also features in "land abundance" theories of African development, which hold that population scarcity on the continent "limited the surplus the powerful could extract" and "prevented the emerge of literate elites and formal institutions" with ambiguous long-term development consequences (e.g., Herbst 2000; Iliffe 2007, p. 2; Fenske 2013; Boone 2014).

How population "density" and "scarcity" are defined in these arguments is not always clear. As many have noted, this determination is often more related to preexisting political and economic conditions than to demographic fundamentals themselves (e.g., Durham 1979; Knight 1986*a*). Moreover, the demographic features of a society may change over time. Densely settled areas may become depopulated following war, disease, famine, or mass emigration. The population shocks themselves may influence long-term development trajectories, as the case of Mexico illustrates.

When the Spanish arrived in Mexico in 1519, they encountered a large and urbanized society of around 20 million people (Gibson 1964; Cook and Borah 1971; Hassig 1985). After the fall of the Triple Alliance (Aztec Empire), the Spanish usurped the Alliance's tribute system, and conquistadors quickly accumulated wealth and power by harnessing labor and tribute from the large indigenous society (Hassig 1985; Knight 2002). This was also the model adopted in other population-dense colonies (e.g., Engerman and Sokoloff 1997; Acemoglu,

[2]See Mahoney (2010) for a critical discussion of this literature.

Johnson and Robinson 2002; Mahoney 2010). However, by the middle of the sixteenth century, a combination of disease, drought, and famine had decimated Mexico's population, undermining these early institutions.

The size and significance of Mexico's demographic collapse has few parallels in history. The closest comparison is arguably the depopulation of Europe during the Black Plague. In Western Europe, demographic losses due to the Plague raised per capita incomes, labor-market bargaining power, and consumption among those who survived, which had important implications for long-run development. It is argued that the demographic shock caused or contributed to the decline in feudalism, the rise of urbanization, and the creation of modern nation-states in this region (e.g., North and Thomas 1973; Epstein 2000; Levine 2006; Voigtländer and Voth 2013*b*,*a*). However, in other contexts affected by the Plague, the rapid decline in population had very different institutional consequences. While labor scarcity can have beneficial consequences for the surviving population, it can also encourage coercion as landlords seek to extract effort from the scarce remaining workforce (Domar 1970; Acemoglu and Wolitzky 2011). This second mechanism seems to have dominated in Eastern Europe, for example, where labor scarcity led to the re-adoption of serfdom, undermining long-term political and economic development (e.g., Domar 1970; Klein and Ogilvie 2017; Markevich and Zhuravskaya 2017). In China, evidence suggests that wages increased in the short term following the Plague, but also that this did not spark a massive long-term shift in economic development as it had in Western Europe (Voigtländer and Voth 2013*b*). These differences highlight the importance of political and institutional context in mediating the effects of sudden population scarcity.

In Mexico, the 16th-century demographic collapse took place in a context of colonization and conquest. The sharp decline in population placed heavy burdens both on colonial elites, who could no longer live off of the surplus production of indigenous villages, and indigenous laborers, who were increasingly unable to economically support both themselves and these elites (e.g., Borah 1951; Zavala 1973; Hassig 1985, p. 185). Crucially, however, the elites led

the subsequent process of institutional restructuring. As acquiring wealth through tribute extraction became untenable, elites transitioned to harnessing land and labor for agricultural production and commercialization (Borah 1951; Gibson 1964; Cook and Borah 1971; Knight 2002). The Mexican hacienda—the large estate that became a symbol of economic and political power in the countryside—grew out of this transition. Many members of the new landowner class were elites who had lived off of indigenous tribute under the encomienda and used their power to take advantage of the changing economic circumstances following the collapse (Lockhart 1969; Knight 2002). They developed institutions to control labor, restrict internal migration, and force the population onto large estates, preventing survivors from leveraging their bargaining power in the labor market (Gibson 1964; Florescano 1987; Knight 2002). The preexisting political structure was thus instrumental in determining the political and economic effects of sudden population scarcity. This process built on itself over time. By the turn of the 20th century, substantial proportions of the rural population in disease-affected areas lived on large estates, which often exceeded 10,000 hectares in size (Southworth 1910).

In the remainder of this paper, we systematically evaluate the relationship between Mexico's population collapse and past and present landholding patterns. In Section 5, we present our baseline results showing that a larger proportion of the rural population lived on haciendas in 1900 in areas that experienced a more precipitous decline in population between 1570 and 1650. Building on historical evidence, we trace this outcome to two mechanisms. The sharp decline in the indigenous population in the early colonial period made it easier for colonists to acquire land from indigenous villages as survivors were left unable to contest land encroachment by elites. In addition, after populations began to recover, landowners had an incentive to acquire additional land around their holdings to deprive workers of outside options, and (during the 19th century) to take advantage of emerging export markets abroad. We provide suggestive historical evidence on these mechanisms in Section 6. As we show empirically in Section 7, Mexico's 20th-century agrarian reform upended these landholding

patterns and transformed rural political economy, but it also kept disease-affected areas on a unique development trajectory that they continue to follow.

3 Data

3.1 Colonial population density

Our measures of colonial population density were digitized from Gerhard (Gerhard 1993*a*,*b*), which contain data on the colonial governorships of Nueva España, Nueva Galicia, Nueva Vizcaya, and Sinaloa and Sonora in central and northern Mexico. The remainder of modern Mexico was either conquered by the Spanish later than the time period that we examine (the northeast and Baja California) or does not have the necessary tree-ring data for us to implement our empirical strategy (the southeast). This extent of the analyzable area is shown figure 1.[3]

Population statistics in Gerhard are reported at the level of 1786 political divisions, outlined with the thin gray lines in Figure 1. We examine a snapshot of population in two periods: around 1570, the first year in which Gerhard's data are relatively complete, and around 1650, the sample year closest to the nadir of Mexico's indigenous population.[4] Of the 191 districts in Gerhard, data are available in both periods for 177 units.[5] Gerhard constructs the population data using a large number of historical sources, relying primarily on the *relaciones geográficas*, a set of comprehensive questionnaires distributed to local officials by King Phillip II beginning in the late 1560s. In consultation with local indigenous leaders,

[3]The Yucatán and northern frontier followed a somewhat different demographic pattern than central Mexico. Though areas of the north were affected by disease, these regions were sparsely populated with sedentary indigenous populations prior to colonial era. The Yucatán experienced a sharp decline in population due to disease, but not until the late 1600s, allowing indigenous holdings to survive in those areas (Garcia Bernal 1979; Gerhard 1993*c*; Knight 2002). The Yucatán is also somewhat unique in the concentration of surviving indigenous population, though our results hold in Oaxaca and Chiapas (both heavily indigenous regions.

[4]Hassig (1985) and Knight (2002) both estimate that the low point in population occurred during the 1630s.

[5]In the online appendix, we show the robustness of the results to using varying bandwidths around these sample years.

Spanish administrators recorded the number of individuals paying tribute to the Spanish Crown (tributaries) in different years. These tribute counts are the primary measures of population available for this time period.

In general, each married male and his immediate family constituted one tributary, and unmarried individuals over 15 and widows were considered "half-tributaries" (Gerhard 1993*a*). Certain groups, such as indigenous nobility and clergy, were exempt from paying tribute. Following Cook and Borah (1960, 1971), we convert any data recorded as "individuals" or "families" to tribute counts using an average of 5 members per family and 2.8 individuals per tributary.[6] After converting units to be comparable, we calculate the municipal-level tribute population density in 1570 and 1650 by overlaying Gerhard's 1786 divisions with 20th-century municipal boundaries[7] and calculating the space-weighted average tributary density. Our analysis is conducted at the level of these 20th-century administrative units, which contain the variation of interest in landholding patterns and our geographic covariates.[8]

Figure 1 shows the density of the tributary population in 1570 and the change between 1570 and 1650 as expressed by the ratio of the latter to the former. The area with the highest population density is the region around Mexico City, the center of the Triple Alliance. Populations are also relatively dense in the Mixtecan region of Oaxaca in the south and in the Tarascan region of Michoacán in the center-west, both of which were major population centers prior to the Conquest (Cook and Borah 1971). The right side of Figure 1 illustrates the precipitous population decline between 1570 and 1650. Tribute densities declined by over 70% on average among the 177 districts for which we have data. Only two districts,

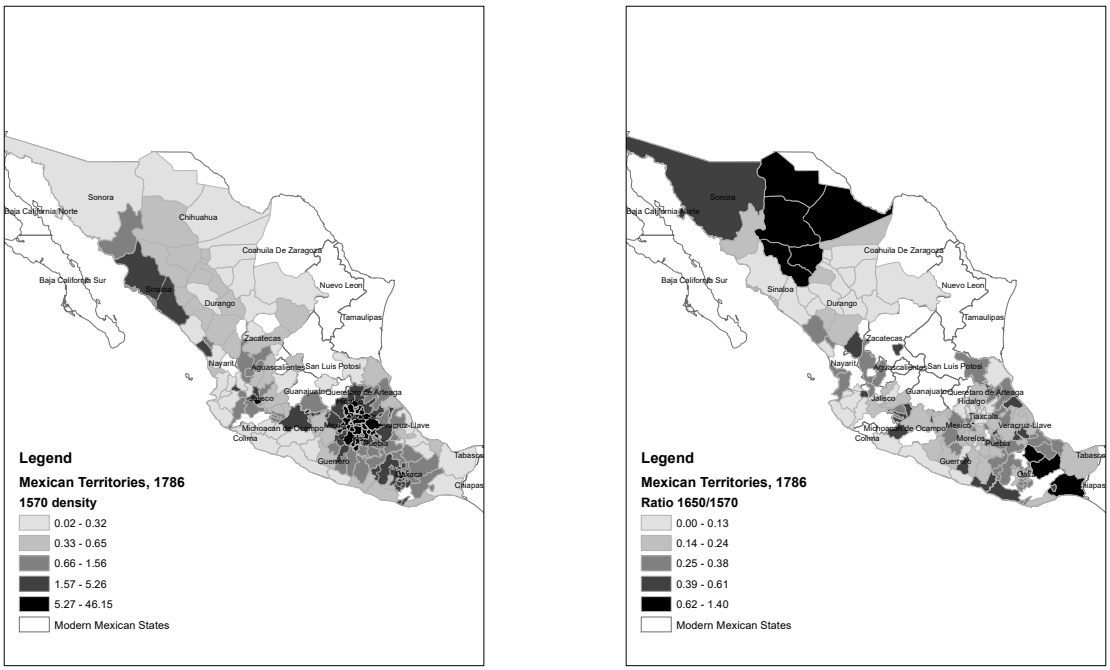
[6]In Appendix A.2, we discuss these conversions and present estimates excluding the regions where the data had to be converted.

[7]Municipalities in Mexico are akin to counties in the United States. Given that municipal boundaries change over time, we use the smallest unit at which all of our data are available. This is typically the municipality as of 1900 except when the 1900 municipality was later absorbed into a larger unit. In Oaxaca, we aggregate the data to the level of the district in 1900.

[8]The "treatment" variable is thus population collapse at the 1786 unit level, conditional on controls calculated at the 20th-century unit level. This is equivalent to examining the impact of village-level variation on household-level outcomes, conditioning on household covariates.

both in Oaxaca, grew between these years.

Figure 1: Tributary Density, 1570 and 1650



(a) 1570 (b) Density 1650/Density 1570

Figures map tributary density, digitized from (Gerhard 1993*a*,*b*). Darker colors indicate greater density in panel a., and less population loss in panel b.

There is some debate regarding the reliability of the tribute data (Borah 1976). While there is evidence that colonists had incentives to systematically over- or undercount population during different times, Cook and Borah (1971) conclude that the Crown was able to limit fraud during the period under study. The late 16th-century numbers are thought to be particularly reliable, as they marked the beginning of Phillip II's efforts to extract resources from his dominion in order to pull Spain out of bankruptcy.[9] Beginning in the 1530s, counts were conducted by a single official for each town and then subjected to a review through which interested parties could challenge the results. The system of tribute collection had

[9]Knight (2002) quipped that Phillip II "liked paperwork almost as much as he disliked rebellion" (p. 57).

been harmonized by the study period, and exemptions from tribute are not thought to vary systematically across administrative units (Cook and Borah 1960). The size of the collapse as measured our data is similar in magnitude to the estimates of Cook and Borah (1979) based on a comparison of village-level data for a subset of towns. However, because measurement error remains a concern, we employ an empirical strategy to address this potential source of bias (Section 4).

3.2 The hacienda

The primary source of data on the hacienda is the Mexican Historical Archive of Localities (AHL), maintained by Mexico's National Institute of Statistics and Geography (INEGI). The AHL records the population of every locality in Mexico in each census or national count beginning in 1900 and continuing at approximately 10-year intervals. The AHL also records the political category of the locality in every census year as determined by the government. We extract the number of localities that are described as "haciendas" (haciendas), "ranchos" (ranches), or "fincas" (farms/estates) and the total population living in these localities in 1900 and 1910 for each municipality. To our knowledge, comprehensive data on land tenure across the country during the colonial period or early Independence period do not exist.

We adopt a relatively expansive measure of the "hacienda" for several reasons. Mexican scholars have highlighted regional and political differences in whether large estates were classified as "haciendas" or "ranchos" and have noted the lack of distinction between the various estate classifications (e.g., González Navarro 1969, p. 120; Gómez-Serrano 2000, p. 464). In our data, a considerable number of localities are classified as haciendas in one census year (either 1900 or 1910) and ranchos or fincas in the other. Southworth (1910) provides a list of large haciendas immediately prior to the Revolution based on tax records, and we were able to match a large number of AHL villages to directory listings by name in the state of Aguascalientes. Among properties listed in both sources, those classified by the AHL as haciendas had an average size of 15,555 hectares, those classified as ranchos

13,000 hectares, those that changed between rancho and hacienda in the two censuses 12,400 hectares, and those in other categories 8,667 hectares.[10] Our estimates of the size of the hacienda population under our expansive definition are somewhat smaller than those of McBride (1923) and Tannenbaum (1929), and they are highly correlated with both of these.

3.3 Control variables

We include a series of control variables in our models to account for geographic factors that could be correlated with population dynamics and land tenure patterns. The main controls are average elevation, average slope, the area of the municipality, distance to Mexico City, distance to railroad line in the late 19th century, an indicator of whether a municipality is in the malarial zone (has a minimum elevation of less than 1000 m), and a measure of potential agricultural productivity. Average elevation and slope are calculated at the administrative level using a digital-elevation model (DEM) with a resolution of 30 meters. Our measure of soil quality is the space-weighted potential low-input maize productivity from the Food and Agriculture Organization's Global Agro-Ecological Zones (IIASA/FAO 2012). The FAO potential productivity measure is calculated using information about local climate, soil types, slope, and rainfall, and comes in a layer with 5-arcminute resolution (about 10-km cells). Maize is the main staple crop in Mexico, and potential maize productivity is highly correlated with that of many other grain and vegetable crops. The means and standard deviations of all of our variables are presented in the Appendix (Tables A2 and A3). In some specifications, we also include control for 1570 population density to account for potential demographic differences between high- and low-collapse areas prior to this time period.

[10] As a comparison, the average size of private property holdings in Mexico was 50 hectares in 1991 (Warman 2001, p. 24).

4 Identification strategy

As discussed above, measurement error and omitted-variables bias are potential concerns in this setting. We therefore adopt an instrumental-variables estimation strategy aimed at leveraging a plausibly exogenous source of variation in the population data. Our strategy is based on the climate conditions associated with a severe epidemic in the 1570s. As discussed in Section 2, drought and disease were major causes of the indigenous population collapse in central Mexico. Three severe epidemics in particular—those of 1519–21, 1545–8, and 1576–80—had catastrophic effects on the native population and on colonial institutions (Gibson 1964; Hassig 1985; Knight 1986*b*). While there is wide agreement that the first of these epidemics was caused by smallpox, recent scholarship has found that these later epidemics, recorded by their Nahuatl name "*cocoliztli*," were caused by an indigenous hemorrhagic fever that was transmitted by rodents (Acuña Soto, Calderon Romero and Maguire 2000; Acuña Soto et al. 2002).[11] Like other rodent-transmitted diseases, such as hantavirus, the spread of *cocoliztli* was climate-related. During severe droughts, disease-carrying rodents congregate around common water sources, allowing the disease to spread in the dwindling rodent population. When climate conditions swing back to heavy rainfall, the rodent population rebounds and comes in contact with the human population, raising the possibility of infection. Consistent with this interpretation, Acuña Soto et al. (2002) show using drought reconstructions from northern Mexico that these *cocoliztli* epidemics emerged in years of above average rainfall after prolonged periods of drought.

Following these authors, we construct instruments for the decline in tributary population between 1570 and 1650 using climate measures of drought, rainfall abundance, and the gap between the two during the 1576–1580 epidemic. We argue that the shock of experiencing severe drought followed by excess rainfall in the 1570s should be orthogonal to possible confounding variables, conditional on our included covariates. To construct our instruments, we

[11] In recent work, Vagene et al. (2018) have proposed that salmonella rather than hemorrhagic fever may be responsible for the *cocoliztli* epidemic. However, because salmonellosis is also spread by rodents through infecting water and food, this would not affect our empirical strategy.

extract subnational data on rainfall and drought using information from the *North American Drought Atlas* (Cook and Krusic 2004). The Atlas reconstructs estimates of historical drought conditions using evidence recovered from a network of tree-ring chronologies using a principal-components framework. Drought is measured using the Palmer Drought Severity Index (PDSI), a normalized measure of soil moisture that captures deviations from typical conditions at a given location. Abnormally dry conditions are recorded using negative numbers, and positive numbers represent abnormally wet conditions. We create a grid (2.5 km cells) of distance-weighted PDSI estimates between points in the Atlas, overlay the 1900 municipal boundaries on this grid, and extract average PDSI for each district and municipality in each year of the 1570s.[12]

There may be concern that this interpolated PDSI measure does not capture meaningful deviations in rainfall. In order to check whether this is the case, we compared data from the Atlas between 1979 and 1990 with rainfall measures from Mexico's National Water Commission (CONAGUA). CONAGUA calculates a rainfall index similar to PDSI for hundreds of stations within the country, 287 of which overlap with our study area. We find that the correlation between our interpolated PDSI measure and that of CONAGUA at the exact points of the rainfall stations is 0.36, and the correlation coefficient between the first differences of the measures is 0.48. This suggests that these data do pick up variation in weather, and are especially responsive to temporal changes.

The excluded instruments used in our analysis are the sum of the two lowest consecutive PDSI values between 1570 and 1575 (more negative numbers indicate more severe, prolonged drought), the maximum PDSI between 1576 and 1580 (which captures excess rainfall), and the difference between the minimum PDSI between 1570 and 1575, and the maximum PDSI between 1576 and 1580. This parameterization is taken from the epidemiology literature on rodent-transmitted disease and specifically from Figures 2 and 3 in Acuña Soto et al. (2002), which emphasize the swing from severe drought conditions to rainfall abundance in

[12]While the Atlas covers most of Mexico, it does not have information on the far southeast and Yucatán peninsula due to the absence of usable tree rings in that area.

the spread of *cocoliztli*. Using these instruments, we estimate the impact of the population decline between 1570 and 1650 on measures of the dominance of the hacienda prior to the Revolution in 1900. Specifically, we estimate the following system of equations:

$$H_{ma} = \alpha_0 + \alpha_1 T_{ma} + \alpha_{\mathbf{2}} \cdot \mathbf{G_m} + \epsilon_{ma} \tag{1}$$

$$T_{ma} = \beta_0 + \beta_{\mathbf{1}} \cdot \mathbf{Z_{ma}} + \beta_{\mathbf{2}} \cdot \mathbf{G_m} + \mu_{ma} \tag{2}$$

The dependent variable (H_{ma}) is as the inverse hyperbolic sine of the percent rural population living in hacienda communities in 1900.[13] We also examine results for an indicator variable for whether the municipality had an above-median number of estates per square kilometer that year as an additional measure of hacienda dominance (see online appendix). The key explanatory variable, the population decline in municipality m within 1786 administrative region a (T_{ma}), is measured using the ratio of 1650 to 1570 density. $\mathbf{Z_{ma}}$ represents the vector of instruments. We also control for a series of geographic features related to population and agriculture ($\mathbf{G_m}$), as discussed in Section 3.3 and include governorship-level fixed effects.[14] To control for correlation across units of analysis within 1786 regions, we cluster standard errors at the level of the 1786 units, of which there are 177. Our results are also robust to using spatially weighted standard errors (see Appendix Table A10).

The key assumption for identification in this setting is that—conditional on agricultural productivity, climatic variation, and other included geographic variables—the climatic shocks between 1576 and 1580 did not independently affect hacienda development. One potential concern could be that areas susceptible to climate shocks might have lower agricultural productivity, independently affecting land concentration. To address the possibility that

[13]The inverse hyperbolic sine transformation can be interpreted similarly to a log transformation and is preferable to it for a variety of reasons (Burbage, Magee and Robb 1988). Results are robust to using $ln(x+1)$ (see online appendix).

[14]We cannot include 1786 district-level fixed effects as numerous districts include only one or two municipal-level observations. In addition, municipalities do not nest neatly into 1786 administrative boundaries.

consistently high variation in rainfall may independently affect both tributary density and agricultural productivity, we control for the standard deviation of PDSI over the entire colonial period. We also control directly for a measure of maize productivity, as well as for various measures of elevation and slope. Elevation and slope are key determinants of agricultural potential, both through their primal determination of soil quality and (in the case of elevation) their importance in determining ecosystem type. It is difficult to imagine a separate channel through which idiosyncratic shocks in the 16th century would affect the development of haciendas in the 20th century after controlling for agricultural productivity other than through the decline in population due to disease.[15] In section 5 we discuss a falsification exercise indicating that similar shocks in other colonial time periods do not predict the fall in population or the presence of haciendas.

5 Results: population collapse and the hacienda

Table 1 presents the first-stage and reduced-form relationships between the climate variables and the decline in population between 1570 and 1650 or measures of hacienda presence in 1900. Columns (1) and (2) show the first-stage results on the effect of the vector of climate instruments on the population collapse. Decreases in the dependent variable indicate a greater decline in population. As expected based on Acuña Soto et al. (2002), the size of the collapse, conditioning on the overall severity of drought and the subsequent size of excess rainfall, is strongly related to the swing between the maximum and the minimum PDSI between 1570 and 1580. This is consistent with the argument that the swing from drought to rainfall was conducive to the spread of *cocoliztli*. The first column shows these results including only governorship fixed effects. In Column (2), we add our vector of purely geographic controls: elevation, slope, maize productivity, municipal area, and the standard deviation of PDSI. The inclusion of these variables makes our estimates much more precise, as would

[15] If these controls were somehow insufficient, and susceptibility to swings in climate were related to hacienda emergence, we would expect this to go in the opposite direction of our reduced-form results. More climate variability should reduce rather than increase agricultural land concentration.

be expected given the relationship between land productivity and population density. The third column introduces a series of additional economic controls: distance to railroad line, distance to Mexico City, the percent of the population in urban areas in 1900, the size of the rural population in 1900, and tributary density in 1570. These variables are intended to further control for alternative channels through which hacienda development might occur. The inclusion of these variables strengthens the relationship between the weather shock and the population collapse.

Columns (4) through (6) present reduced-form estimations of our climate instruments on our measure of hacienda dominance, the inverse hyperbolic sine of percent population on estates in 1900. The signs of all coefficients are the opposite of what they are in columns (1–3), consistent with our argument that landed estates became more dominant in areas that experienced a more severe population loss between 1570 and 1650.

Table 2 reports the instrumental-variables estimates of the impact of the 1570–1650 population collapse on hacienda importance in 1900. These estimations use only the area of the sample for which we have both colonial population density data and reconstructed drought data.[16] The first three columns report OLS estimates of the impact of the population collapse on hacienda populations. As in Table 1, the first column includes only governorship dummies, the second includes geographic controls, and the third includes both geographic and economic controls. Columns (4) through (6) show the instrumental variables estimates using the same sets of control variables.

As expected, the point estimates of the coefficient on the explanatory variable of interest, the ratio of 1650 to 1570 population, are negative. Their magnitudes are large; a one standard-deviation decrease in the ratio of 1650 to 1570 population density is associated with a 0.37 standard-deviation increase in the percentage of the population living on estates in 1900. Interestingly, the coefficient on the control for initial 1570 population density in the IV regressions is negative. One explanation for this would be that land concentration

[16]We lose 43 municipalities from the original tribute density data as there is no climate data for the far southeast of the study area.

Table 1: First-stage and reduced form regressions: climate, population, and 1900 estates

	Dependent variable					
	Ln(1650/1570 density)			IHS(% pop estates)		
	(1)	(2)	(3)	(4)	(5)	(6)
Min($\sum$ 2 years PDSI, 1570-1575)	0.006	0.228	0.202	-0.732	-1.116	-2.680
	(0.155)	(0.162)	(0.151)	(3.431)	(3.200)	(3.238)
Maximum PDSI, 1576-1580	0.162	0.201	0.292	-11.935**	-12.936**	-13.274**
	(0.241)	(0.226)	(0.199)	(5.397)	(5.221)	(5.212)
Max(PDSI, 1576-1580) - Min(PDSI, 1570-1575)	-0.537*	-0.671***	-0.785***	10.294*	11.520**	14.588**
	(0.282)	(0.252)	(0.226)	(5.602)	(5.768)	(5.961)
Ln(Tributary density, 1570)			-0.199***			-1.591
			(0.050)			(1.829)
Geographic covariates	no	yes	yes	no	yes	yes
Additional covariates	no	no	yes	no	no	yes
Observations	1030	1030	1030	1030	1030	1030
Adjusted R^2	0.197	0.307	0.388	0.325	0.347	0.383

Unit of observation is the municipality. Standard errors are in parentheses and are clustered at the 1786 administrative level. * p< 0.10, ** p<0.05, *** p < 0.01. These are partial results. Geographic covariates include elevation < 1000 meters, mean GAEZ maize productivity, mean slope, mean elevation, 1900 municipal area and area squared, and standard deviation of PDSI throughout the colonial period. Additional covariates include 1/km to nearest railroad line, % population in large cities in 1900, ln(km to Mexico City), and ln(rural population in 1900). All estimations include governorship dummies.

Table 2: Population Decline and Hacienda Importance

	Dependent variable IHS(% pop on estates)					
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Tributary density 1650/1570)	-0.173	-0.157*	-0.278***	-0.459	-0.441**	-0.900***
	(0.110)	(0.092)	(0.092)	(0.429)	(0.177)	(0.217)
Ln(Tributary density, 1570)			-0.182			-0.318**
			(0.132)			(0.147)
Geographic covariates	no	yes	yes	no	yes	yes
Additional covariates	no	no	yes	no	no	yes
	OLS	OLS	OLS	IV	IV	IV
Observations	1030	1030	1030	1030	1030	1030
p-value AR-test				0.000	0.000	0.000
Partial R2				0.118	0.206	0.221
F-statistic						

Unit of observation is the 1900 municipality. Standard errors are in parentheses and are clustered at the 1786 administrative level. * p< 0.10, ** p< 0.05, *** p < 0.01. These are partial results. Geographic covariates include elevation < 1000 meters, mean GAEZ maize productivity, mean slope, mean elevation, 1900 municipal area and area squared, and standard deviation of PDSI throughout the colonial period. Additional covariates include 1/km to nearest railroad line, % population in large cities in 1900, ln(km to Mexico City), and ln(rural population in 1900). All estimations include governorship dummies. The AR-test is the p-value of the Anderson-Rubin Wald test of significance of endogenous regressors. The partial R2 is based on Shea's (1997) partial R2 of the first stage instruments, and the F-test is the relevance test based on the first stage with clustered standard errors. The excluded instruments are the sum of the PDSI of the two largest drought years between 1570 and 1575, the difference between the lowest PDSI between 1570 and 1575 and the highest PDSI between 1576 and 1580, and the maximum PDSI between 1576 and 1580.

was more extensive where population density was lower prior to 1570 due to the high cost of acquiring land in high-density areas (see Section 6).

Turning attention to the reliability of our IV estimation strategy, the partial R-squared of the instruments in the first stage ranges from 0.12 without controls to 0.22 with the full set of controls. The F-statistic for the joint significance of excluded instruments in the first-stage regressions ranges from 6.5 without controls to 21 including all controls. The associated p-values for these F-tests are always less than 0.001. We conclude that the instruments are relevant. However, because we have clustered standard errors, which may exacerbate a weak instruments problem (e.g., Cameron and Miller 2015), we also include the p-value on the Anderson-Rubin Wald statistic, which tests the hypothesis that the coefficient of the endogenous regressor is equal to zero, and is robust to the presence of weak instruments.

The IV coefficient estimates are considerably larger in magnitude than those of our OLS models. We believe that the differences in magnitude can be plausibly explained by a combination of omitted-variables bias and measurement error in the OLS estimates. For the difference in magnitudes to be explained by omitted-variables bias, important omitted variables in the OLS estimation would have to influence population density and the establishment of the hacienda in the same direction (i.e., either have a positive effect on both the explanatory and dependent variables or a negative one). One possible omitted variable the could fit this description is agricultural productivity (uncaptured by our geographic and weather controls), which is plausibly positively related to both pre-colonial population density and subsequent incentives for land concentration.

Measurement error would also attenuate the OLS estimates. The OLS estimate is about one-third the size of the IV estimate. For these results to be explained by measurement error alone, there would have to be sufficient noise relative to signal in the early colonial tribute density data. In a simple univariate regression model, where a variable x^* is measured with error v: $x = x^* + v$, it can be shown that the noise-to-signal ratio required to generate a

difference of this magnitude would be: $s = \frac{\sigma_v^2}{\sigma_{x^*}^2} = (1/.7 - 1)^{-1} = 2.3$.[17] In other words, the variance of the population decline measurements would need to be 2.3 times greater than the true variance of the population decline. We believe that this plausible given the peculiar features of the Spanish tribute system.

To increase confidence in our findings, we show in the online appendix that our results are robust to using the swing from drought to rain as a single omitted instrument, conditioning on drought severity and excess rain in both stages. We also show the robustness of our results to weak instrument robust inference using Anderson-Rubin confidence intervals. We present a variety of other robustness checks in the appendix, including examining alternative outcomes, excluding the Mexico City region from our analysis, using the level of 1650 population density (conditional on 1570 density) rather than the size of the drop as our explanatory variable, aggregating the data to 1786 administrative units, and using a spatial error model (Appendix A.3). Our results hold under all of these alternative approaches.

Finally, we conduct a series of randomization inference and placebo exercises to address potential concerns about spatial autocorrelation and endogeneity (Appendix A.4). First, we randomly permute our vector of climate instruments in 20,000 trials, recording coefficient estimates and test statistics for our first-stage, reduced-form, and two-stage least squares models. This exercise shows that our actual estimates fall very far from what would have occurred had the weather shocks been randomly distributed across Mexico. We then repeat the exercise, randomly permuting the ratio of 1650 to 1570 density to provide additional evidence on the performance of our OLS and two-stage-least squares models. Second, we conduct a series of placebo exercises to examine the relationship between our outcomes and weather shocks in the 280 colonial years other than those in the 1570s. We find that coefficient estimates in the reduced-form and first-stage regressions of these placebos are concentrated around zero. The placebo regressions show coefficients of a similar magnitude to our real instruments in 0.7 percent of the cases. A complete analysis of these first-stage

[17] We thank Colin Cameron for pointing this out to us.

and reduced-form placebo exercises can be found in Appendix A.5.

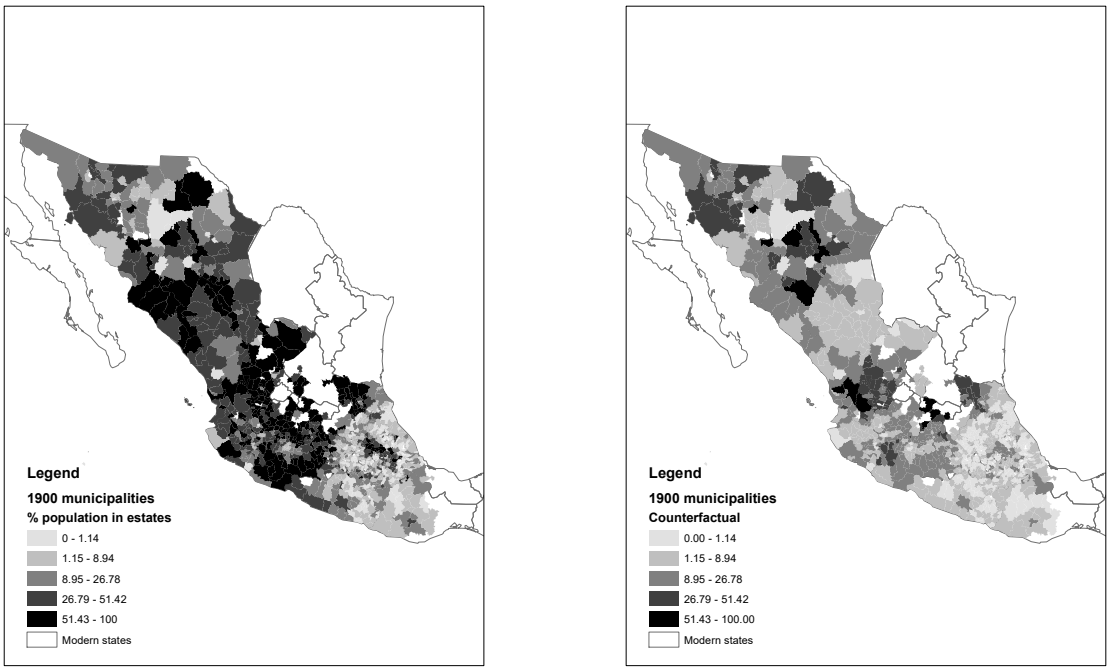
5.1 Counterfactual analysis of the causal impact

What would landholdings have looked like in the absence of the demographic collapse? This section proposes one response. Exploiting the instrumental-variables estimates, we explore what the variation in hacienda distribution in 1900 would have looked like had there been no collapse of the population between 1570 and 1650. To do this, we subtract off the predicted marginal effect of the population change in each municipality from the actual 1900 outcome using the estimates in column (2) of Table 2.[18]. We then transform these estimates back into levels. The distribution of hacienda population changes substantially in this counterfactual. The median percentage of 1900 population living in haciendas in our actual data is 16.7. When we remove the effect of the population collapse, this decreases to 3.9 percent. The change is even larger at the 75th percentile. In actual 1900 levels, 44 percent of the rural population lives on haciendas, while in our counterfactual estimate this is 11.8 percent.

Figure 2 compares the geographic distribution of the hacienda population in 1900 with that implied by our counterfactual exercise. The left panel presents the actual distribution of hacienda population in 1900 (in percent of municipal population). In the right panel, we adjust the hacienda population to take away the impact of the population collapse as above. The shading is coded according to the actual 1900 hacienda population quintiles. The differences between actual and counterfactual hacienda populations are particularly striking across the northwest and just northeast of Mexico City. There is also less variation in the counterfactual estimates of the 1900 hacienda population. The variance is about 17 percent lower. This suggests that the decline in population between 1570 and 1650 had a significant effect in generating the observed distribution of individuals living in hacienda communities in 1900.

[18]This exercise follows the logic of the counterfactual experiment in Bohn and Pugatch (2013). We bound our estimates between 0 and 100.

Figure 2: Change in hacienda population distribution from demographic collapse



(a) Actual (b) Counterfactual

Figures map out intensity in the percentage of the population living on haciendas in 1900, with darker colors indicating higher percentages. Panel a. shows the actual distribution, while panel b. shows the distribution predicted to occur in the absence of the population collapse.

6 Persistence and change in land institutions

The previous section established that more of the population lived on haciendas in 1900 in areas that experienced a greater decline in population between 1570 and 1650. Why did the population collapse encourage land concentration, and why did the hacienda institution persist for centuries? In this section, we discuss some of the mechanisms linking the population collapse in the early colonial era with the rise of the hacienda and its persistence until the Revolution, relying on an extensive qualitative historical literature on this time period.[19]

As noted in Section 2, standard economic theories produce very different predictions about how markets and institutions should respond to a demographic shock of this magnitude. It is not clear that sudden labor scarcity should lead to an increase in land concentration. In neoclassical models, an increase in the land/labor ratio would increase the shadow wage and bargaining power of labor, making it more difficult for landowners to find workers, raising the costs of production, and lowering the benefits of land acquisition (e.g., Acemoglu and Wolitzky 2011). In this context, however, land concentration increased in the wake of the collapse for several reasons.

The most direct channel through which the sharp decline in population enabled the rise of the hacienda was through the depopulation of the countryside and the weakening of indigenous villages. The collapse increased the incentives for acquiring land to feed colonists who could no longer live off of indigenous production (e.g., Borah 1951; Gibson 1964; Knight 2002). The reduction in population simultaneously made it easier for colonists to obtain land in disease-affected areas. Though colonist encroachment was a persistent problem, indigenous villages were not defenseless when faced with the loss of their land. The Crown recognized indigenous property rights, and villages were often able to legally contest attempted land seizures through lengthy disputes that could be very costly for elites (Gibson 1964, Ch. 10; Coatsworth and Tortella Casares 2002, p. 129–30; Knight 2002, p. 120–1; Franco Vivanco 2018). Healthy indigenous villages were somewhat effective in checking the growth of the

[19]We are grateful to an anonymous reviewer for detailed comments on the mechanisms we discuss here.

hacienda, relying at times on the support of the Church and the Crown, who were concerned with indigenous exploitation and with growing elite power (e.g., Chevalier 1963, Ch. 6; Knight 2002, p. 80–1; Franco Vivanco 2018).

By contrast, in areas hit hard by disease, colonists looking to acquire land were met with little resistance (e.g., Florescano 1976; Knight 2002, p. 81). In some places, villages disappeared entirely and survivors were sent to live in larger towns. Mindful of maintaining control of territory, the colonial government offered incentives for land acquisition in these areas, establishing a system through which landowners could request grants of nearby "unoccupied" land that had been depopulated during major epidemics or during the forced resettlements of survivors that followed (Gibson 1964; Hassig 1985; Knight 2002). Even where villages survived, epidemics decimated the informal institutions that had facilitated collective action, making political and social resistance more difficult (e.g., Katz 1988, p. 80). Villages hit hard by disease were also often unable to meet their tribute requirements to the Crown and were thus forced to sell lands to elites to raise revenue (Florescano 1976, p. 52). These factors facilitated colonist land acquisition in disease-affected areas, setting them on a different development trajectory that continued for centuries (e.g., Knight 2002, p. 80).

Though populations eventually recovered from the collapse, early land concentration built on itself over time. The demographic collapse occurred during a critical time of institutional development. By the time the indigenous population rebounded in the 17th century, the formalization of land titles had made it nearly impossible to reclaim lands that had been lost during the collapse (McBride 1923; Gibson 1964; Knight 2002). As agricultural markets began to grow again, the early impacts of the collapse were exacerbated by labor-market effects that encouraged further land acquisition. Though there is little evidence that real wages increased following the collapse, acquiring workers was a persistent problem for landowners. As in many contexts, landowning elites relied on a variety of extra-economic institutions to ration and control labor (e.g., Gibson 1964; Binswanger, Deininger and Feder 1995; Knight

2002). Debt peonage, through which peasant families took up residence in large estates and became tied to them through debt contracts, eventually became synonymous with the hacienda (e.g., Knight 1986*a*; Florescano 1984; Knight 1986*b*, p. 85–9). The manipulation of local land markets offered an additional mechanism for landowning elites to reduce the cost of labor recruitment and retention. By encroaching on nearby indigenous landholdings, landowners were able to drive down the outside economic options of rural laborers, forcing them into employment on large estates (Florescano 1976, 1984; Conning 2007). In addition to being "one of the *hacendados'* principal strategies for acquiring workers" (Florescano 1987, p.267), land encroachment had the additional benefit of reducing smallholder agricultural production, which undermined competition in grain markets and increased the market share of the hacienda (Chevalier 1963; Florescano 1976, p. 95).

Through these interdependent forces, a small number of landowners increasingly dominated local land, labor, and agricultural markets in much of the country with important long-term impacts on the political economy of the countryside (Chevalier 1963; Sokoloff and Engerman 2000; Conning 2007; Knight 2002). By the end of the colonial era, populations had largely recovered, but the legacy of the collapse continued to shape land tenure. After Mexico gained independence from Spain in 1821, ongoing instability prevented the new national government from implementing major reforms to Mexican land and labor markets until the late 19th century. During the rule of Porfirio Díaz (1884–1911), a series of policies were adopted that exacerbated landholding inequality and amplified the political and economic power of landowners. The development of railroads and export-friendly economic policies raised land values and the profitability of agricultural production (McBride 1923; Coatsworth 1974). Limits placed on communal landholding in the mid-19th century provided a legal basis for the seizures of millions of additional hectares of indigenous and public lands by large estates to take advantage of these new economic opportunities (McBride 1923; González Navarro 1957; Meyer 1973; Coatsworth 1974; Knight 1986*b*). The result was a marked increase in land alienation and concentration in the areas around haciendas, though traditional lands

survived in other areas (Knight 1986*b*). This continued the unique path followed by areas affected by the population collapse up to the 20th century.

7 From the hacienda to the ejido

By the eve of the Revolution in 1910, up to four-fifths of rural communities and half of the rural population lived on haciendas and were "subject to the political and social control of the estate" (Knight 1986*b*, p. 96). Agrarian concerns were a central motivation for the Mexican Revolution (1910–21).[20] During the Revolution, the government began a land reform program through which peasant villages could petition the government to expropriate and redistribute land from nearby large estates.[21] This program would last until 1991 and transform the country's landholding patterns (Knight 1986*b*; Sanderson 1984; Markiewicz 1993). By the end of the 1930s, the hacienda sector had been almost completely dismantled and replaced with a unique "ejidal" system, a fusion of collective and individual land tenure where holders were given limited property rights over territory (Sanderson 1984; Knight 1991). According to government statistics, the ejidal sector now contains more than half of the country's surface area, and around half of the rural population lives in an ejidal community (e.g., Gordillo 2014).

Though the hacienda ceased to exist, preexisting land institutions (and thus the legacy of the 16th-century collapse) shaped the implementation of the reform in important ways. Only land in large[22] properties was eligible for expropriation and redistribution, focusing reform in the areas surrounding large plantations (Sanderson 1984; Markiewicz 1993). In addition, the provision of land reform in Mexico depended on citizen pressure for redistribution or

[20]The role of land grievances in the Revolution remains the subject of considerable debate. Jean Meyer among others has argued that early work on the Revolution (e.g., McBride 1923; Tannenbaum 1929) overstated the role of the hacienda in the economy (e.g., Meyer 1973, 1986), though others have contested this claim (Knight 1986*b*, 1991). However, Meyer himself agreed that "the hacienda...directly or indirectly leaves its mark on all of rural life" (1986, p. 483).

[21]See Sanderson (1984) for a detailed description of the design and implementation of the reform.

[22]During the early years of the reform, the cutoff was around 500 hectares, with some adjustments for soil and crop types. This shifted slightly over time. See Sanderson (1984) for a discussion.

restitution of lands. The reform was designed to address agrarian grievances, which were often directly related to hacienda economic domination or encroachment (Sanderson 1984; Knight 1991). Because of this, disease-impacted areas were disproportionately affected by the land reform and thus continue to have unique landholding patterns today.

To illustrate the trajectory linking the population collapse to contemporary land tenure, we extend the analysis of Section 5 to examine the distribution of ejidal lands in Mexico. Our measures of agrarian reform come from several sources. The first of these is Sanderson's (1984) database of every presidential resolution on land reform from 1916 to 1976.[23] We examine the four resolution types involving redistribution of land: outright grants of land to communities, restitutions of lost village lands, expansions of ejidal grants, and grants distributing land for new agricultural villages. We aggregate the resolution-level data to the municipal level, recording the total number of positive land resolutions and the number of reform beneficiaries in each municipality.[24] We also examine two measures of contemporary landholdings in Mexico. The first of these is a measure of municipal land area held in social property[25] in a municipality, as measured using GIS calculations from a layer created after 1992. The social property sector in Mexico is made up of two different institutions: *ejidos* and *comunidades*. While ejidos were created through government expropriation and redistribution of land through the reform, comunidades were preexisting collective landholdings that were later certified by the government and often predated the reform. While there are similarities, ejidos and comunidades are managed somewhat differently, and their historical antecedents are very different (e.g., Warman 2001). Because of these differences, we also construct a second measure of contemporary land tenure in which we examine ejidos only by subtracting the land area that entered the social-property sector as comunidades.[26] Though redistributive land transfers account for most social property in Mexico (80% of the total

[23]The presidential resolution was the final stage of the reform process and finalized land grants.

[24]We use 1900 municipal boundaries to be consistent with the analysis of Section 5. While the reform continued until 1991, the vast majority of land redistribution took place prior to 1976. The time period of most intense redistribution was the 1930s.

[25]Mexican law recognizes three types of property: public, private, and social.

[26]These stem from the resolutions registered as "sector additions" by Sanderson.

land in the GIS layer), comunidades account for a significant portion of land in some areas, such as Oaxaca, where indigenous landholdings were more likely to survive to the 20th century.

We use two estimation strategies to examine the relationship between the population collapse, the hacienda, and the ejido. The first examines the long run effect of the population collapse directly, using the same strategy as in Section 5. For the second, we examine the relationship between the percentage of the population population on estates in 1900 and outcomes related to the land reform and contemporary land tenure. In both cases, we present both the OLS and the IV estimates, using the same covariates and climate instruments as in Section 5. Reported standard errors for all estimations are clustered at the 1786-district level.

The results of these analyses are presented in Table 3. Odd-numbered columns present coefficient estimates from the OLS models, and even-numbered columns contain those of the instrumental-variables models. The top panel presents the results on measures of land redistribution (the number of resolutions and beneficiaries from the reform), and the bottom panel considers the two measures of contemporary land tenure. Examining the effect of the 16th-century population collapse (Columns (1), (2), (5), and (6)), we observe that there were more land grants and more reform beneficiaries in areas experiencing a steeper decline in population between 1570 and 1650, and that disease-affected areas continue to have more land in the ejidal sector. As in the previous models, the vector of instruments performs well, with first-stage F-statistics and partial R^2 values above standard cutoffs for instrument relevance. For the outcomes of resolutions, beneficiaries, and total GIS area in social property, the Anderson-Rubin p-values are all less than 10%. The results are slightly weaker for the area adjusted for titling, with a p-value of 0.129.

In Columns (3), (4), (7), and (8), we show the OLS and IV coefficient estimates on the impact of the percent population in estates in 1900 on resolutions, beneficiaries, and GIS area in social property. As expected, there were more land reform resolutions and more

Table 3: Agrarian Reform, Haciendas, and the Population Collapse

Panel A:	Dependent variable: Resolutions				Dependent variable: Beneficiaries			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln(Tributary density 1650/1570)	-0.271***	-0.668***			-0.280**	-0.510*		
	(0.077)	(0.197)			(0.117)	(0.277)		
% population in estates, 1900			0.014***	0.027***			0.018***	0.027***
			(0.002)	(0.006)			(0.003)	(0.010)
Observations	1030	1030	1030	1030	1030	1030	1030	1030
AR-test		0.000		0.000		0.076		0.076
Partial R2		0.221		0.146		0.221		0.146
F-statistic		21.149		25.537		21.149		25.537

Panel B:	Dependent variable: GIS area				Dependent variable: GIS area - title			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln(Tributary density 1650/1570)	-0.349**	-0.386			-0.313***	-0.396*		
	(0.137)	(0.259)			(0.110)	(0.207)		
% population in estates, 1900			0.020***	0.030***			0.012***	0.010
			(0.004)	(0.011)			(0.002)	(0.007)
Geographic covariates	yes	yes	yes	yes	yes	yes	yes	yes
Other covariates	yes	yes	yes	yes	yes	yes	yes	yes
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Observations	1030	1030	1030	1030	1030	1030	1030	1030
AR-test		0.092		0.092		0.129		0.129
Partial R2		0.221		0.146		0.221		0.146
F-statistic		21.149		25.537		21.149		25.537

Unit of observation is the municipality. All dependent variables are inverse hyperbolic sine transformed. Standard errors are in parentheses and are clustered at the 1786 administrative level. * p< 0.05, ** p<0.01, *** p < 0.001. These are partial results. Other covariates include elevation < 1000 meters, mean GAEZ maize productivity, 1/km to nearest railroad line, mean elevation, mean slope, 1900 municipal area and area squared, standard deviation of PDSI across the colonial period, % population in large cities in 1900, ln(km to Mexico City), ln(population 1900), & governorship dummy variables. The AR-test is the p-value of the Anderson-Rubin test of endogenous regressors. The partial R2 is based on Shea's (1997) partial R-squared of the first stage instruments, and the F test is the relevance test based on the first stage. The excluded instruments are the sum of the PDSI of the two largest drought years between 1570 and 1575, the difference between the lowest PDSI between 1570 and 1575 and the highest PDSI between 1576 and 1580, and the maximum PDSI between 1576 and 1580.

beneficiaries in areas that had a greater percentage of the population in 1900 living in rural communities. These areas also have more land in the social property sector today, though the results using the adjusted land area measure are somewhat weaker. These results are consistent with accounts linking the land reform to the legacy of the hacienda and with the results of Section 5 showing the relationship between the 16th-century population collapse and the emergence of the hacienda.

These findings call attention to the persistent and evolving effect of the demographic collapse. If the hacienda was blamed for Mexican rural underdevelopment at the beginning of the 20th century, the ejido had taken its place by the turn of the 21st. Many studies have described how the peculiar structure of the ejido, through which smallholders were given incomplete property rights over land, has harmed development by inhibiting access to credit, depressing agricultural investment, and increasing reliance on the government (e.g., Yates 1981; de Janvry, Gordillo and Sadoulet 1997; Deininger and Feder 2001; Dell 2012). The process through which land was allocated and regulated on ejidos was subject to political manipulation, and the limited property rights of smallholders left them dependent on government resources and programs that could be used to buy political support and enable corruption (e.g., Gledhill 1991; Fox 2007; Albertus et al. 2015). While efforts to bolster ejidal property rights beginning in the 1990s have arguably mitigated some of the political and economic problems associated with the social property sector (e.g., de Janvry, Gonzalez-Navarro and Saudolet 2014), there has been little privatization or sale of ejidal land (Warman 2001; Gordillo 2014). Though work on the development consequences of the ejido usually focuses on the design and implementation of the land reform, as we demonstrate above, the reform was heavily tied to the legacy of the hacienda. Areas with a strong ejidal sector today tend to be places with highly unequal land tenure in 1900. Informal institutions associated with the hacienda persisted after the dissolution of the estate as long-standing patronage relationships were often recreated in new forms within the ejido (e.g., Gledhill 1991; Fox 2007). It is difficult to discern whether contemporary development patterns can

be traced to the institutional features of the ejido, the contentious process of land reform, the lingering historical legacy of the hacienda, or some combination of these.

It is clear, however, that the reform altered the development trajectory of disease-impacted areas. Though the hacienda and the ejido have both been blamed for rural underdevelopment, they are thought to impede development through different mechanisms with different consequences. While disease-impacted areas were characterized by the political and economic dominance of a small number of large landowners in 1900, land fragmentation rather than concentration is thought to impede development in these areas today (e.g., Warman 2001). Though the institutional structure of the ejido enabled political and economic manipulation, the ejido remains popular among much of the rural population (Warman 2001; Gordillo 2014). For all of its problems, the land reform subverted the long-standing dominance of landowners and facilitated the creation of new institutional actors with different characteristics, interests, and political resources (Gledhill 1991; Knight 1991; Markiewicz 1993). To say that the ejido is merely a continuation of the hacienda is clearly wrong. As one historian of Mexico put it, the story of the land reform is ultimately one of "revolutionary rupture, not conservative continuity" (Knight 1991, p. 103–4). Though the 16th-century population collapse continues to shape development, the character of its legacy has changed dramatically.

8 Conclusion

The population collapse of the early colonial period shaped Mexican land institutions through the present day. A greater proportion of the rural population lived on large estates prior to the Revolution in areas that experienced a steeper population decline between 1570 and 1650. We trace this outcome to the role of the collapse in shaping land concentration in the early colonial period and to forces that built on this early land concentration through the 20th century. Though land tenure arrangements were overhauled after the Revolution,

we show that the collapse continues to shape political economy through its impact on the 20th-century land reform, which determined Mexico's contemporary land institutions and distribution. Our identification strategy uses information on the climate conditions related to a significant epidemic in the late sixteenth century, which sharply reduced the indigenous population in parts of the country.

This paper makes several important contributions to the study of institutions, historical persistence, and the political economy of population. To the best of our knowledge, this represents the first attempt to systematically test the hypothesis of numerous scholars of Mexican history that the sixteenth-century population collapse shaped the emergence of the hacienda. We also contribute to the study of the political and economic consequences of the hacienda by highlighting its relationship with the ejido, arguably the institution that most directly impacts rural political economy today. A great deal of work on Mexico has sought to assess the political and economic effects of this unique common-property institution. By showing that the historical antecedents of the modern ejido date back to the 16th century, we highlight the challenge of assessing which "perverse" institution—the hacienda or the ejido—should be blamed for rural underdevelopment.

More generally, our work contributes to understanding the political impacts of population scarcity. While the relative density of Mexico at the time of the Conquest shaped early institutional development, colonial elites were forced to adapt to the challenges of profound labor scarcity soon after. The catastrophic population decline fueled land concentration and the development of inefficient institutions, which favored small segments of society and misallocated resources for generations. While in some contexts, such as in Western Europe following the Black Plague, sharp reductions in population led to an increase in the bargaining power of labor, the concentration of land and power by elites following the collapse in Mexico set the country on a trajectory of long-term inequality. This illustrates the importance of political and economic context in determining the short- and long-term consequences of demographic shocks. While shocks on the level of *cocoliztli* are rare, these findings can

help us to understand the institutional aftermath of epidemics and civil conflict, which may also produce a concentrated and sharp decline in local populations.

Our findings also have important implications for understanding historical persistence and institutional change. Much of the empirical work on historical dependence examines one or two snapshots in history, often associating a long-ago shock or event with a contemporary outcome. If one had assessed the long-run impact of Mexico's population collapse in 1900, it might have seemed like disease-impacted areas were set on a trajectory of increasing land concentration and economic inequality. The legacy of the collapse looks very different today; disease-impacted areas now tend to have more fractionalized land tenure through the ejido. One should thus be cautious in interpreting studies of historical persistence. Observing a correlation between a contemporary outcome and a long-ago event does not imply that this positive or negative empirical relationship has always existed, nor does it imply that it will persist into the future. Historical legacies evolve, and even well established historical patterns started by significant events—like the centuries-long dominance of the hacienda following the population collapse—may be reversed.

Path-dependent arguments about institutions and development are sometimes criticized as being overly deterministic. As Banerjee and Duflo (2013) argue, the recognition that historical events can have long-run consequences does not imply that current conditions are unchangeable. Areas affected by disease in the early colonial period continue to follow a different development trajectory, but the way in which these areas differ changed dramatically after the Revolution. Demography was not destiny in Mexico. While the modern ejido arguably owed its existence to the legacy of the hacienda, the land reform represented a complete economic and political reorganization of the countryside. This illustrates that even institutions with long historical antecedents may be responsive to policy change. Historical dependence does not imply stasis.

References

Acemoglu, Daron and Alexander Wolitzky. 2011. "The Economics of Labor Coercion." *Econometrica* 79(2):555–600.

Acemoglu, Daron and James A. Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. New York, NY: Crown Business.

Acemoglu, Daron, Simon Johnson and James A Robinson. 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.

Acharya, Avidit and Alexander Lee. 2016. "Path Dependence in European Development: Medieval Politics, Conflict, and State-Building." *Working Paper* . http://stanford.edu/ avidit/europe.pdf.

Acuña Soto, Rodolfo, David W. Stahle, Malcolm K. Cleaveland and Matthew D. Therrell. 2002. "Megadrought and Megadeath in 16th Century Mexico." *Emerging Infectious Diseases* 8(4):360–362.

Acuña Soto, Rodolfo, Leticia Calderon Romero and James H. Maguire. 2000. "Large epidemics of hemorrhagic fevers in Mexico, 1545-1815." *American Journal of Tropical Medicine and Hygiene* 62:733–739.

Albertus, Michael, Alberto Diaz-Cayeros, Beatriz Magaloni and Barry Weingast. 2015. "Authoritarian Survival and Poverty Traps: Land reform in Mexico." *World Development* . Forthcoming.

Austin, Gareth. 2008. "The 'Reversal of Fortune' Thesis and the Compression of History: Perspectives from African and Comparative Economic History." *Journal of International Development* 20:996–1027.

Baland, Jean-Marie and James A. Robinson. 2008. "Land and Power: Theory and Evidence from Chile." *American Economic Review* 98(5):1737–1765.

Banerjee, A. and E. Duflo. 2013. "Under the Thumb of History? Political Institutions and the Scope for Action." MIT Working Paper 14-02, http://economics.mit.edu/files/9658.

Banerjee, Abhijit and Lakshmi Iyer. 2005. "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." *American Economic Review* 95(4):1190–1213.

Binswanger, Hans P, Klaus Deininger and Gershon Feder. 1995. "Power, distortions, revolt and reform in agricultural land relations." *Handbook of development economics* 3:2659–2772.

Bohn, Sarah and Todd Pugatch. 2013. "U.S. Border Enforcement and Mexican Immigrant Location Choice.". IZA Discussion Paper No. 7842.

Boone, Catherine. 2003. *Political Topographies of the African State*. Cambridge, UK: Cambridge University Press.

Boone, Catherine. 2014. *Property and Political Order in Africa: Land Rights and the Structure of Politics*. Cambridge, UK: Cambridge University Press.

Borah, Woodrow. 1976. The historical demography of aboriginal and colonial America: an attempt at perspective. In *The Native Population of the Americas in 1492*, ed. William Denevan. University of Wisconsin Press.

Borah, Woodrow Wilson. 1951. *New Spain's Century of Depression*. Berkeley, CA: University of California Press.

Burbage, John B., Lonnie Magee and A. Leslie Robb. 1988. "Alternative Transformations to Handle Extreme Values of the Dependent Variable." *Journal of the American Statistical Association* 83(401):123–7.

Cameron, A. Colin and Douglas L. Miller. 2015. "A Practioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50(2):317–373.

Chevalier, François. 1963. *Land and Society in Colonial Mexico: The Great Hacienda*. Berkeley, CA: University of California Press.

Coatsworth, John. 1974. "Railroads, Landholding, and Agrarian Protest in the Early Porfiriato." *Hispanic American Historical Review* 54(1):48–71.

Coatsworth, John H. and Gabriel Tortella Casares. 2002. "Institutions and Long-Run Economic Performance in Mexico and Spain, 1800-2000." Paper presented at the XIIIth Congress of the International Economic History Association.

Conning, Jonathan. 2007. "On 'The Causes of Slavery or Serfdom' and the Roads to Agrarian Capitalism: Domar's Hypothesis Revisited.". Working Paper, Hunter College, https://ideas.repec.org/p/htr/hcecon/401.html.

Cook, E. R. and P. J. Krusic. 2004. *The North American Drought Atlas*. New York, NY: Lamont-Doherty Earth Observatory and National Science Foundation.

Cook, Sherburne Friend and Woodrow Wilson Borah. 1960. *The Indian population of central Mexico, 1531-1610*. Vol. 44 University of California Press.

Cook, Sherburne Friend and Woodrow Wilson Borah. 1971. *Essays in population history: Mexico and the Caribbean*. Vol. 1 Univ of California Press.

Cook, Sherburne Friend and Woodrow Wilson Borah. 1979. *Essays in population history: Mexico and California*. Vol. 3 Univ of California Press.

Dasgupta, Partha. 1995. "The Population Problem: Theory and Evidence." *Journal of Economic Literature* 33:1879–1902.

de Janvry, A, Gustavo Gordillo and Elisabeth Sadoulet. 1997. *Mexico's Second Agrarian Reform: household and community responses, 1990-1994.* Center for US-Mexican Studies.

de Janvry, Alain, Marco Gonzalez-Navarro and Elisabeth Saudolet. 2014. "Are land reforms granting complete property rights politically risky? Electoral outcomes of Mexico's certification program." *Journal of Development Economics* 110:216–225.

Deininger, Klaus and Gershon Feder. 2001. "Land institutions and land markets." *Handbook of agricultural economics* 1:288–331.

Dell, Melissa. 2012. "Path Dependence in Development: Evidence from the Mexican Revolution.". Working paper, Harvard University Department of Economics, http://scholar.harvard.edu/files/dell/files/revolutiondraft.pdf.

Domar, Evsey. 1970. "The Causes of Slavery and Serfdom: A Hypothesis." *Journal of Economic History* 30(1):18–32.

Durham, William H. 1979. *Scarcity and Survival in Central America: Ecological Origins of the Soccer War*. Palo Alto, CA: Stanford University Press.

Engerman, Stanley L and Kenneth L Sokoloff. 1997. Factor endowments: institutions, and differential paths of growth among new world economies: a view from economic historians of the United States. In *How Latin America Fell Behind: Essays from the Economic Histories of Brazil and Mexico, 1800-1914*, ed. Stephen Haber. Stanford, CA: Stanford University Press.

Epstein, S. R. 2000. *Freedom and Growth: The Rise of States and Markets in Europe, 1300–1750*. London, UK: LSE/Routledge.

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

Florescano, Enrique. 1976. *Origen y desarrollo de los problemas agrarios de México, 1500–1821*. México, DF: Editorial ERA.

Florescano, Enrique. 1984. The formation and economic structure of the hacienda in New Spain. In *Cambridge History of Latin America, Vol. II*, ed. Leslie Bethell. Cambridge University Press pp. 153–188.

Florescano, Enrique. 1987. The hacienda in New Spain. In *Colonial Spanish America*, ed. Leslie Bethell. Cambridge University Press pp. 250–285.

Fox, Jonathan. 2007. *Accountability Politics: Power and Voice in Rural Mexico*. New York, NY: Oxford University Press.

Franco Vivanco, Edgar. 2018. "Justice as Checks and Balances: Indigenous Claims in the Courts of Colonial Mexico." Stanford University.

Garcia Bernal, Manuela Cristina. 1979. *Población y encomienda en Yucatán bajo los austrias*. Sevilla, ES: Escuela de Estudios Hispano-Americanos.

Gerhard, Peter. 1993*a*. *A guide to the Historical Geography of New Spain: Revised Edition*. University of Oklahoma Press.

Gerhard, Peter. 1993*b*. *The North Frontier of New Spain*. Norman, OK: University of Oklahoma Press.

Gerhard, Peter. 1993*c*. *The Southeast Frontier of New Spain*. Norman, OK: University of Oklahoma Press.

Gibson, Charles. 1964. *The Aztecs Under Spanish Rule: A History of the Indians of the Valley of Mexico, 1519-1810*. Palo Alto, CA: Stanford University Press.

Gledhill, John. 1991. *Casi Nada: A Study of Agrarian Reform in the Homeland of Cardenismo*. Institute for Mesoamerican Studies.

Goldstein, Markus and Christopher Udry. 2008. "The profits of power: Land rights and agricultural investment in Ghana." *Journal of Political Economy* 116(6):981–1022.

Gómez-Serrano, Jesús. 2000. *Haciendas y ranchos de Aguascalientes: Estudio regional sobre la tenencia de la tierra y el desarrollo agrícola en el siglo XIX.* Aguascalientes, Aguascalientes: Universidad Autónoma de Aguascalientes.

González Navarro, Moisés. 1957. El Porfiriato. La Vida Social. In *Historia Moderna de México*, ed. Daniel Cosío Villegas. Vol. IV of *Historia Moderna de México* México, DF: Editorial Hermes.

González Navarro, Moisés. 1969. "La tenencia de la tierra en México." *Caravelle* 12:115–134.

Gordillo, Gustavo. 2014. "El Campo: Contra-Hechos." *La Jornada* August 16.

Hassig, Ross. 1985. *Trade, tribute, and transportation: The sixteenth-century political economy of the Valley of Mexico.* Norman, OK: University of Oklahoma Press.

Herbst, Jeffrey. 2000. *States and Power in Africa: Comparative Lessons in Authority and Control.* Princeton, NJ: Princeton University Press.

Homer-Dixon, Thomas F. 1999. *Environment, Scarcity, and Violence.* Princeton University Press.

IIASA/FAO. 2012. "Global Agro-ecological Zones (GAEZ v3.0).". IIASA, Laxenburg, Austria and FAO, Rome, Italy.

Iliffe, John. 2007. *Africans: The History of a Continent (Second Edition).* New York: Cambridge University Press.

Imbens, Guido W. and Paul R. Rosenbaum. 2005. "Robust, accurate confidence intervals with a weak instrument: quarter of birth and education." *Journal of the Royal Statistical Society, Series A.* 168(1):109–126.

Katz, Friedrich. 1988. Rural Uprisings in Preconquest and Colonial Mexico. In *Riot, Rebellion, and Revolution: Rural Social Conflict in Mexico.* Princeton, NJ: Princeton University Press.

Klein, Alex and Sheilagh Ogilvie. 2017. "Was Domar Right? Serfdom and Factor Endowments in Bohemia." Centre for Economic Policy Research Discussion Paper Series, DP12388.

Knight, Alan. 1986*a*. "Mexican Peonage: What Was It and Why Was It?" *Journal of Latin American Studies* 18(1):41–74.

Knight, Alan. 1986*b*. *The Mexican Revolution*. Lincoln, NE: University of Nebraska Press.

Knight, Alan. 1991. "Land and Society in Revolutionary Mexico: The Destruction of the Great Haciendas." *Mexican Studies/Estudios Mexicanos* 7(1):73–104.

Knight, Alan. 2002. *Mexico: The Colonial Era*. New York, NY: Cambridge University Press.

La Porta, Rafael, Florencio Lopez-De-Silanes, Andrei Shleifer and Robert W. Vishny. 1997. "Legal Determinants of External Finance." *The Journal of Finance* 52(3):1131–1150.

Larreguy, Horacio. 2013. "Monitoring Political Brokers: Evidence from Clientelistic Networks in Mexico.". Manuscript, Harvard Kennedy School of Government.

Levine, David. 2006. Why and How Population Matters. In *The Oxford Handbook of Contextual Political Analysis*. New York, NY: Oxford University Press.

Lockhart, James. 1969. "Encomienda and hacienda: The evolution of the great estate in the Spanish Indies." *The Hispanic American Historical Review* 49(3):411–429.

Mahoney, James. 2010. *Colonialism and Postcolonial Development: Spanish America in Comparative Perspective*. New York: Cambridge University Press.

Mahoney, James and Kathleen Thelen. 2010. A Theory of Gradual Institutional Change. In *Explaining Institutional Change: Ambiguity, Agency, and Power*. Cambridge University Press.

Markevich, Andrei and Ekaterina Zhuravskaya. 2017. "The Economic Effects of the Abolition of Serfdom: Evidence from the Russian Empire." Working Paper, New Economic School.

Markiewicz, Dana. 1993. *The Mexican Revolution and the Limits of Agrarian Reform, 1915-1946*. Boulder, CO: Lynne Rienner Publishers.

McBride, George McCutchen. 1923. *The Land Systems of Mexico*. New York: American Geographical Society.

Meyer, Jean. 1973. *Problemas campesinos y revueltas agrarias (1821-1910)*. México, DF: Secretaría de Educación Pública.

Meyer, Jean. 1986. "Haciendas y ranchos, peones y campesinos en el porfiriato. Algunas falacias estadísticas." *Historia mexicana* 35(3):477–509.

North, Douglass C. 1990. *Institutions, Institutional Change, and Economic Performance*. Cambridge University Press.

North, Douglass C. and Robert Paul Thomas. 1973. *The Rise of the Western World: A New Economic History*. New York, NY: Cambridge University Press.

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

Nunn, Nathan. 2009. "The Importance of History for Economic Development." *Annual Review of Economics* 1:65–92.

Pierson, Paul. 2000. "Increasing Returns, Path Dependence, and the Study of Politics." *American Political Science Review* 94(2):251–267.

Roland, Gerard. 2004. "Understanding institutional change: fast-moving and slow-moving institutions." *Studies in Comparative International Development* 38(4):109–131.

Sánchez de la Sierra, Raúl. 2017. "On the Origins of the State: Stationary Bandits and Taxation in Eastern Congo." Working Paper, University of California.

Sanderson, Susan R Walsh. 1984. *Land reform in Mexico, 1910-1980*. New York, NY: Academic Press.

Sokoloff, Kenneth L. and Stanley L. Engerman. 2000. "History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World." *Journal of Economic Perspectives* 14(3):217–232.

Southworth, John R. 1910. *El Directorio Oficial de las Minas y Haciendas de México/The Official Directory of the Mines and Estates of Mexico*. México, DF: Southworth, FRGS.

Tannenbaum, Frank. 1929. *The Mexican Agrarian Revolution*. Washington, DC: The Brookings Institution.

Thelen, Kathleen. 2004. *How Institutions Evolve: The Political Economy of Skills in Britain, the United States, and Japan*. New York, NY: Cambridge University Press.

Turner, Frederick Jackson. 1920. *The Frontier in American History*. New York: Henry Holt and Company.

Vagene, Ashild J., Alexander Herbig, Michael G. Campana, Nelly M. Robles Garca, Christina Warinner, Susanna Sabin, Maria A. Spyrou, Aida Andrades Valtuea, Daniel Huson, Noreen Tuross, Kirsten I. Bos and Johannes Krause. 2018. "Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico." *Nature Ecology & Evolution* 2(3):520–528.

Voigtländer, Nico and Hans-Joachim Voth. 2013*a*. "Gifts of Mars: Warfare and Europe's Early Rise to Riches." *Journal of Economic Perspectives* 27(4):165–186.

Voigtländer, Nico and Hans-Joachim Voth. 2013*b*. "The Three Horsemen of Riches: Plague, War, and Urbanization in Early Modern Europe." *Review of Economic Studies* 80:774–811.

Warman, Arturo. 2001. *El campo Mexicano en el siglo XX.* México, DF: Fondo de la Cultura Económica.

Yates, P. Lamartine. 1981. *Mexico's Agricultural Dilemma.* Tucson, AZ: University of Arizona Press.

Young, Alwyn. 2005. "The gift of the dying: The tragedy of AIDS and the welfare of future African generations." *The Quarterly Journal of Economics* 120(2):423–466.

Zavala, Silvio A. 1973. *La encomienda indiana, segunda edición.* México, DF: Editorial Porrúa.

Appendix A Supporting information: for online publication only

Appendix A.1 Additional tables and figures

Table A1: Various estimates of Mexico's population collapse

Author	Population (millions) 1519	1595	Percent decline (%)
	"Mexico"		
Rosenblat (1935)	4.5	3.5	22
Aguirre-Beltran (1946)	4.5	2.0	56
Zambardino (1980)	5-10	1.1-1.7	64-89
Mendizabal (1939)	8.2	2.4	71
Cook and Simpson (1948)	10.5	2.1-3.0	71-80
Cook and Borah (1960)	18-30	1.4	78-95
	Central Mexican Region		
Sanders (1976)	2.6-3.1	0.4	85-87
	Valley of Mexico		
Whitmore (1992)	1.3-2.7	0.1-0.4	69-96
Gibson (1964)	1.5	0.2	87
	128 towns		
Kubler (1942)	0.2	0.1	55

source: Robert McCaa. http://www.hist.umn.edu/ rmccaa/noncuant/democatt.htm

Figure A1: Decline of native population of Mexico from Gerhard (1993*a*)

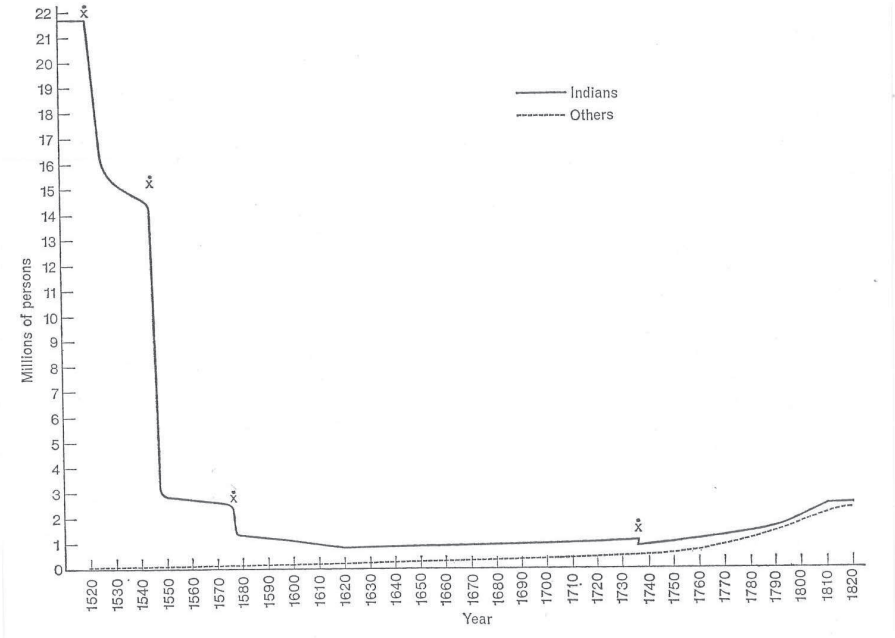


Table A2: Summary Statistics

	mean	sd	max	min
Municipal tributary density, 1570	2.828	4.191	46.155	0.037
Municipal tributary density, 1650	0.622	1.374	35.386	0.005
1650/1570 municipal tributary density	0.320	0.384	7.735	0.022
% population in estates, 1900	25.573	25.855	100.000	0.000
> median estates/km^2	0.521	0.500	1.000	0.000
Mean municipal elevation	1588.885	760.093	3225.858	12.107
Minimum elevation < 1000 m	0.458	0.498	1.000	0.000
Mean municipal maize productivity	0.775	0.649	3.257	0.000
Minimum km to railroad line	74938.280	97441.348	538305.625	0.000
Ln(Municipal area, 1900)	6.011	1.332	10.297	2.617
Ln(Population density, 1900)	2.586	1.359	6.515	-3.960
% population in communities > 5000	0.060	0.182	1.000	0.000
Ln(Km to Mexico City)	5.405	1.020	7.479	1.233
Observations	1030			

Table shows means, standard deviations, maxima and minima of the data sources described in Section 3. The unit of analysis is the 1900 municipality.

Table A3 presents summary statistics for municipalities that experienced above- and below-median population loss between 1570 and 1650. To assess differences between our sample and all applicants, we use normalized differences in means (column (3)), which is a scale-free measure of the difference in distributions recommended by Imbens and Wooldridge (2009). A rule of thumb suggested by Imbens and Rubin (2015) is that normalized differences greater than 0.25 standard deviations should be considered substantial. We also include significance stars to indicate p values lower than the 5th percentile, although we note that these values are calculated without taking into account clustering. Municipalities experiencing a larger decline in population had significantly higher population density in 1570, whereas the density in both groups was approximately equal in 1650. Although there is no apparent difference in the percentage of the rural population in haciendas in 1900 across the columns, the areas with greater collapse have a higher number estates per area on average. These municipalities also tend to be at slightly higher elevation, are less like to be in "malarial" ar-

eas, exhibit higher potential maize productivity, and are closer to 19th-century railroad lines relative to the less-affected municipalities. Given these substantial differences in covariates, it is in fact rather striking that hacienda development is similar across these true groups, and underscores the fact that geography and baseline population are important covariates to include in subsequent estimations.

Table A3: Summary statistics by tributary ratio, 1650/1570

	(1) High collapse	(2) Low collapse	(3) Norm diff
Municipal tributary density, 1570	3.917	1.747	-0.379**
Municipal tributary density, 1650	0.621	0.623	0.001
1650/1570 municipal tributary density	0.158	0.480	0.653**
% population in estates, 1900	25.543	25.603	0.002
> median estates/km^2	0.561	0.482	-0.113**
Mean municipal elevation	1721.742	1457.056	-0.250**
Minimum elevation < 1000 m	0.398	0.518	0.172**
Mean municipal maize productivity	0.846	0.705	-0.154
Ln(Municipal area, 1900)	5.892	6.130	0.127**
Ln(Population density, 1900)	2.762	2.412	-0.183**
% population in communities > 5000	0.055	0.064	0.034
Ln(Km to Mexico City)	5.149	5.660	0.366**
Nueva España	0.795	0.720	-0.125**
Nueva Galicia	0.113	0.101	-0.029
Nueva Vizcaya	0.072	0.085	0.034
Sinaloa y Sonora	0.019	0.095	0.232**
Observations	513	517	1256

Columns (1) and (2) show mean values, while column (3) shows differences (read in terms of standard deviations). ** indicates that a t-test for difference produced a p-value of less than 0.05.

Appendix A.2 Data appendix

In order to generate as complete a sample of tributary data from Gerhard (1993*a*,*b*) as possible, we needed to convert raw population data in several forms into comparable units. In cases where Gerhard reports population in terms of "individuals" rather than tributaries, we converted the figure to tribute counts by dividing the number of individuals by 2.8. This figure is based on the discussion in Cook and Borah (1960, 1971) relating tribute counts to population data. Also following these sources, where the population was reported as "families" or "vecinos" (neighbors), we use these figures as a proxy for the tribute count directly. This second conversion was only necessary in eight cases. We examine whether our main results change if we use only the divisions for which population data were recorded as tribute counts in both 1570 and 1650. Table A4 reports the results from the baseline instrumental-variables regressions using this subset of data. We lose a considerable number of observations but the results remain qualitatively similar and strongly significant.

We also needed to group together tribute data that were recorded in different years. While Gerhard's data are recorded in several waves, the specific year at which data are available for different colonial divisions varies by political division. The most problematic issues would arise if data were reported during the epidemic (1576-1580) as that could lead us to over- or understate the population collapse in some areas. None of our data were recorded in this window. Some data do come from before the period before the first cocoliztli epidemic in 1545. In these cases, we may be overestimating the population collapse by including the fall in population due to the first epidemic. Table A5 shows results from estimations that are limited to municipalities where at least part of the 1570 population data comes from data recorded before 1545. Our results are unaffected by this modification.

One might also think that merging data from time periods that were significantly removed from the dates of interest could drastically affect our outcomes. To test whether this is the case, we limit the data window to ten years around each of 1570 and 1650 (i.e., restrict the data to counts in 1560–1580 and 1640–1660). Table A6 shows that the results are unaffected,

Table A4: IV-Estimation using only data initially reported as tributaries

	Dependent variable			
	% pop on estates		> median estates/km^2	
	(1)	(2)	(3)	(4)
Ln(Tributary density 1650/1570)	-0.737	-1.664***	-0.981***	-0.877***
	(0.499)	(0.552)	(0.323)	(0.235)
Ln(Tributary density, 1570)		-0.537**		-0.297***
		(0.241)		(0.074)
Geographic covariates	yes	yes	yes	yes
Additional covariates	no	yes	no	yes
	IV	IV	IV	IV
Observations	887	887	887	887
Adjusted R^2	0.227	0.105	-1.103	-0.514

Unit of observation is the municipality. Standard errors are in parentheses and are clustered at the 1786 administrative level. * p< 0.10, ** p<0.05, *** p < 0.01. These are partial results. Geographic covariates include elevation < 1000 meters, mean GAEZ maize productivity, mean slope, mean elevation, 1900 municipal area and area squared, and standard deviation of PDSI throughout the colonial period. Additional covariates include 1/km to nearest railroad line, % population in large cities in 1900, ln(km to Mexico City), and ln(rural population in 1900). All estimations include governorship dummies. The excluded instruments are the sum of the PDSI of the two largest drought years between 1570 and 1575, the difference between the lowest PDSI between 1570 and 1575 and the highest PDSI between 1576 and 1580, and the maximum PDSI between 1576 and 1580.

Table A5: IV-Estimation using only data with tributary observations after 1545

	Dependent variable			
	% pop on estates		> median estates/km^2	
	(1)	(2)	(3)	(4)
Ln(Tributary density 1650/1570)	-0.468***	-0.925***	-0.293***	-0.283***
	(0.177)	(0.218)	(0.091)	(0.083)
Ln(Tributary density, 1570)		-0.324**		-0.133***
		(0.148)		(0.034)
Geographic covariates	yes	yes	yes	yes
Additional covariates	no	yes	no	yes
	IV	IV	IV	IV
Observations	1024	1024	1024	1024
Adjusted R^2	0.241	0.231	0.034	0.152

Unit of observation is the municipality. Standard errors are in parentheses and are clustered at the 1786 administrative level. * p< 0.10, ** p<0.05, *** p < 0.01. These are partial results. Geographic covariates include elevation < 1000 meters, mean GAEZ maize productivity, mean slope, mean elevation, 1900 municipal area and area squared, and standard deviation of PDSI throughout the colonial period. Additional covariates include 1/km to nearest railroad line, % population in large cities in 1900, ln(km to Mexico City), and ln(rural population in 1900). All estimations include governorship dummies. The excluded instruments are the sum of the PDSI of the two largest drought years between 1570 and 1575, the difference between the lowest PDSI between 1570 and 1575 and the highest PDSI between 1576 and 1580, and the maximum PDSI between 1576 and 1580.

though the sample size is reduced considerably.

Table A6: IV-Estimation using only data within a 10-year window of 1570 and 1650

	Dependent variable			
	% pop on estates		> median estates/km^2	
	(1)	(2)	(3)	(4)
Ln(Tributary density 1650/1570)	-0.508***	-0.927***	-0.241***	-0.237***
	(0.138)	(0.172)	(0.079)	(0.073)
Ln(Tributary density, 1570)		-0.436***		-0.140***
		(0.133)		(0.037)
Geographic covariates	yes	yes	yes	yes
Additional covariates	no	yes	no	yes
	IV	IV	IV	IV
Observations	672	672	672	672
Adjusted R^2	0.183	0.205	0.130	0.271

Unit of observation is the municipality. Standard errors are in parentheses and are clustered at the 1786 administrative level. * p< 0.10, ** p<0.05, *** p < 0.01. These are partial results. Geographic covariates include elevation < 1000 meters, mean GAEZ maize productivity, mean slope, mean elevation, 1900 municipal area and area squared, and standard deviation of PDSI throughout the colonial period. Additional covariates include 1/km to nearest railroad line, % population in large cities in 1900, ln(km to Mexico City), and ln(rural population in 1900). All estimations include governorship dummies. The excluded instruments are the sum of the PDSI of the two largest drought years between 1570 and 1575, the difference between the lowest PDSI between 1570 and 1575 and the highest PDSI between 1576 and 1580, and the maximum PDSI between 1576 and 1580.

A final issue we consider concerns the reliability of the Gerhard data. The Gerhard volumes are widely respected, and to our knowledge no other comprehensive dataset on subnational population in the early colonial period exists. Cook and Borah (1979) conducted a detailed study evaluating tributary counts on a town-by-town basis for parts of central Mexico. This analysis compares what are considered "complete" tributary lists compiled around 1550 with subsequent tribute and tributary data compiled in 1646. Using these numbers, the authors calculate the ratio of tributaries in the two time periods for locations which appear on both lists (about 54% of towns). They conclude that the average ratio of 1646 to 1548 population for the available cities was 0.268. In our data, the mean ratio of weighted tributary densities in 1650 to those of 1570 is equal to 0.32, and the median is 0.24. Although the boundaries that we use do not overlay exactly with those in Cook and Borah

(1979), the correlation between the figures that we derive from Gerhard and those compiled by Cook increases our confidence in the reliability of our data.

Appendix A.3 Alternative specifications

This appendix discusses the robustness of the findings presented in Table 2 and evaluates the assumptions underlying our empirical strategy. Our reported results are robust to using outcomes measured from the 1910 census in place of 1900, to using above median density of hacienda villages as a measure of hacienda intensity, and using the log transformation of the dependent variable rather than the inverse hyperbolic sine. These results are shown in Table A7. Our findings are also unchanged when using the log of 1650 tributary density as the main explanatory variable and when estimating the model excluding observations from the Federal District, which is the area of highest population density in our sample and potentially not representative of the country as a whole (Table A8). We also aggregate the hacienda outcomes (measured at the level of the 1900 municipality) up to the 1786 division level and re-estimate our system of equations at that larger unit of analysis (Table A9). The results are similar in magnitude and direction as in our baseline estimations, though the instrument loses some power as a result of the significant decrease in the number of observations. As an additional check, we estimate a spatial autoregressive disturbance model to examine whether or not accounting for spatial autocorrelation of the error terms changes the results. The signs of the coefficients remain unchanged, but the magnitude of the effects decreases somewhat in these specifications (Table A10).

In addition to the robustness checks discussed above, we also reestimate the model using a single instrument, the gap between the year of minimum drought and maximum rainfall, including the other two instruments as covariates in both the first- and second-stage estimations. The results are shown in Table A11. The coefficients of interest are larger in magnitude than in the original estimates, and the single instrument has less explanatory power than the vector of three instruments. The first-stage F-statistic on excluded instrument declines to 12.119, and the partial R-squared of the excluded instrument decreases to 0.052. The single instrument is clearly weaker than the vector of three instruments, although the Anderson-Rubin test of significance of endogenous regressor in the main equation clearly

rejects the null of being equal to zero, with a p-value of 0.003 or less in all specifications. Given that the weak instruments problem is exacerbated with clustered standard errors (e.g., Cameron and Miller 2015), we also calculate Anderson-Rubin confidence intervals which are robust to weak instruments. These AR confidence intervals are complete and do not contain zero. They are wide enough to contain the coefficient estimates from the baseline OLS and IV regressions, so we believe that the differences in magnitudes should be interpreted with caution.

Table A7: Estimation using variations on the dependent variable

	Dependent variable					
	> median estates/km^2		Ln(1+% pop on estates)		IHS(% 1910 pop. estates)	
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Tributary density 1650/1570)	-0.072**	-0.276***	-0.242***	-0.804***	-0.274***	-0.722***
	(0.030)	(0.081)	(0.081)	(0.194)	(0.091)	(0.219)
Ln(Tributary density, 1570)	-0.087***	-0.132***	-0.162	-0.284**	-0.199	-0.298**
	(0.026)	(0.033)	(0.116)	(0.129)	(0.127)	(0.140)
All covariates	yes	yes	yes	yes	yes	yes
	OLS	IV	OLS	IV	OLS	IV
Observations	1030	1030	1030	1030	1029	1029
Adjusted R^2	0.220	0.152	0.295	0.239	0.327	0.300

Unit of observation is the municipality. Standard errors are in parentheses and are clustered at the 1786 administrative level. * p< 0.10, ** p<0.05, *** p < 0.01. These are partial results. Geographic covariates include elevation < 1000 meters, mean GAEZ maize productivity, mean slope, mean elevation, 1900 municipal area and area squared, and standard deviation of PDSI throughout the colonial period. Additional covariates include 1/km to nearest railroad line, % population in large cities in 1900, ln(km to Mexico City), and ln(rural population in 1900). All estimations include governorship dummies. The excluded instruments are the sum of the PDSI of the two largest drought years between 1570 and 1575, the difference between the lowest PDSI between 1570 and 1575 and the highest PDSI between 1576 and 1580, and the maximum PDSI between 1576 and 1580. The partial R2 is based on Shea's (1997) partial R-squared of the first stage instruments, and the F-test is the relevance test based on the first stage.

Table A8: Estimation using variations on colonial population and subsamples

	Dependent variable IHS(% pop on estates)			
	Full Sample		Excluding DF	
	(1)	(2)	(3)	(4)
Ln(Tributary density, 1650)	-0.278***	-0.900***		
	(0.092)	(0.217)		
Ln(Tributary density 1650/1570)			-0.274***	-1.039***
			(0.094)	(0.222)
Ln(Tributary density, 1570)	0.096	0.581***	-0.188	-0.229
	(0.147)	(0.212)	(0.132)	(0.145)
All covariates	yes	yes	yes	yes
	OLS	IV	OLS	IV
Observations	1030	1030	1021	1021
p-value AR-test		0.000		0.000
Partial R2		0.221		0.214
p-value F		21.149		20.809

Unit of observation is the municipality. Standard errors are in parentheses and are clustered at the 1786 administrative level. * p< 0.10, ** p<0.05, *** p < 0.01. These are partial results. Geographic covariates include elevation < 1000 meters, mean GAEZ maize productivity, mean slope, mean elevation, 1900 municipal area and area squared, and standard deviation of PDSI throughout the colonial period. Additional covariates include 1/km to nearest railroad line, % population in large cities in 1900, ln(km to Mexico City), and ln(rural population in 1900). All estimations include governorship dummies. The excluded instruments are the sum of the PDSI of the two largest drought years between 1570 and 1575, the difference between the lowest PDSI between 1570 and 1575 and the highest PDSI between 1576 and 1580, and the maximum PDSI between 1576 and 1580. The partial R2 is based on Shea's (1997) partial R-squared of the first stage instruments, and the F-test is the relevance test based on the first stage.

Table A9: Estimation using aggregation to 1786 administrative levels

	Dependent variable			
	Ln(Tributary density 1650/1570)	IHS(% pop on estates)		
	(1)	(2)	(3)	(4)
Max(PDSI, 1576-1580) - Min(PDSI, 1570-1575)	-1.163***	12.138		
	(0.329)	(8.149)		
Maximum PDSI, 1576-1580	0.618**	-13.586*		
	(0.278)	(7.155)		
Min($\sum$ 2 years PDSI, 1570-1575)	-0.008	2.235		
	(0.189)	(4.755)		
Ln(Tributary density 1650/1570)			-1.721	-6.166*
			(2.021)	(3.354)
Ln(Tributary density, 1570)	-0.317***	-2.788	-6.426***	-7.509***
	(0.089)	(1.719)	(1.848)	(1.847)
All covariates	yes	yes	yes	yes
	OLS	OLS	OLS	IV
Observations	177	177	177	177
p-value AR-test				0.000
Partial R2				0.204
p-value F				0.000

Unit of observation is the 1786 administrative division. Standard errors are in parentheses and are robust. * p< 0.10, ** p<0.05, *** p < 0.01. These are partial results. Geographic covariates include elevation < 1000 meters, mean GAEZ maize productivity, mean slope, mean elevation, 1900 municipal area and area squared, and standard deviation of PDSI throughout the colonial period. Additional covariates include 1/km to nearest railroad line, % population in large cities in 1900, ln(km to Mexico City), and ln(rural population in 1900). All estimations include governorship dummies. The excluded instruments are the sum of the PDSI of the two largest drought years between 1570 and 1575, the difference between the lowest PDSI between 1570 and 1575 and the highest PDSI between 1576 and 1580, and the maximum PDSI between 1576 and 1580.

Table A10: Estimation with spatial errors

	Dependent variable			
	1650/1570 density		% pop on estates	
	(1)	(2)	(3)	(4)
Max(PDSI, 1576-1580) - Min(PDSI, 1570-1575)	-0.785***	-0.902***		
	(0.226)	(0.191)		
Maximum PDSI, 1576-1580	0.292	0.346**		
	(0.199)	(0.160)		
Min($\sum$ 2 years PDSI, 1570-1575)	0.202	0.147		
	(0.151)	(0.101)		
Ln(Tributary density 1650/1570)			-0.900***	-0.369***
			(0.217)	(0.133)
Other covariates	yes	yes	yes	yes
	OLS	Spatial errors LS	IV	Spatial errors IV
Observations	1030	1030	1030	1030
Adjusted R^2	0.388		0.237	

Unit of observation is the municipality. Standard errors are in parentheses and are clustered at the 1786 administrative level in columns (1) and (3). Columns (2) and (4) are estimated using generalized 2 stage least squares with spatially autocorrelated disturbances, as described in Drukker, Prucha, and Raciborski (2011). * p< 0.10, ** p<0.05, *** p < 0.01. These are partial results. Geographic covariates include elevation < 1000 meters, mean GAEZ maize productivity, mean slope, mean elevation, 1900 municipal area and area squared, and standard deviation of PDSI throughout the colonial period. Additional covariates include 1/km to nearest railroad line, % population in large cities in 1900, ln(km to Mexico City), ln(rural population in 1900), the sum of the PDSI of the two largest drought years between 1570 and 1575, and the maximum PDSI between 1576 and 1580. All estimations include governorship dummies. The excluded instrument is the difference between the lowest PDSI between 1570 and 1575 and the highest PDSI between 1576 and 1580.

Table A11: IV-Estimation using only drought-rainfall gap as excluded instrument

	Dependent variable			
	% pop on estates		> median estates/km^2	
	(1)	(2)	(3)	(4)
Ln(Tributary density 1650/1570)	-1.173	-1.283*	-0.937*	-0.716**
	(0.626)	(0.507)	(0.406)	(0.247)
	[-3.62,-0.26]	[-3.03,-0.49]	[-2.53,-0.47]	[-1.69,-0.39]
Ln(Tributary density, 1570)		-0.339		-0.203**
		(0.176)		(0.064)
Geographic covariates	yes	yes	yes	yes
Other covariates	no	yes	no	yes
	IV	IV	IV	IV
Observations	1030	1030	1030	1030
AR-test	0.018	0.003	0.000	0.000
Partial R2	0.035	0.052	0.035	0.052
F-statistic	7.093	12.119	7.093	12.119

Unit of observation is the municipality. Standard errors are in parentheses and are clustered at the 1786 administrative level. Anderson-Rubin confidence intervals in []. * p< 0.10, ** p<0.05, *** p < 0.01. These are partial results. Geographic covariates include elevation < 1000 meters, mean GAEZ maize productivity, mean slope, mean elevation, 1900 municipal area and area squared, and standard deviation of PDSI throughout the colonial period. Additional covariates include 1/km to nearest railroad line, % population in large cities in 1900, ln(km to Mexico City), ln(rural population in 1900), the sum of the PDSI of the two largest drought years between 1570 and 1575, and the maximum PDSI between 1576 and 1580. All estimations include governorship dummies. The excluded instrument is the difference between the lowest PDSI between 1570 and 1575 and the highest PDSI between 1576 and 1580. The partial R2 is based on Shea's (1997) partial R-squared of the first stage instruments, and the F-test is the relevance test based on the first stage.

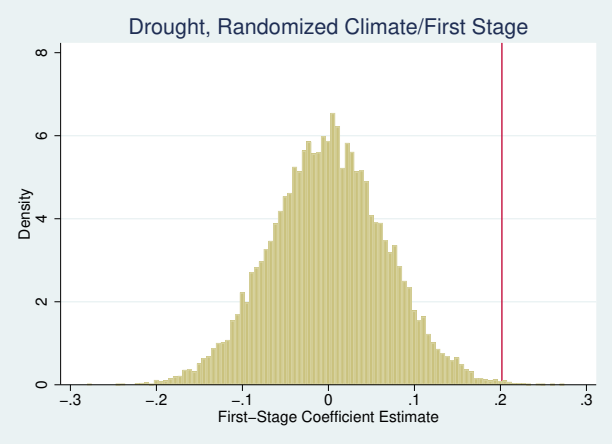
Appendix A.4 Randomization inference exercise

In this section, we conduct a series of randomization inference-style exercises to address concerns about complex spatial correlation in our standard errors. In addition to demonstrating the robustness of our results to estimation using a standard spatial error model (Table A10), we can also calculate p-values for our OLS and first-stage estimates using a randomization inference procedure based on Fisher's exact test (Imbens and Rosenbaum 2005; Sánchez de la Sierra 2017).

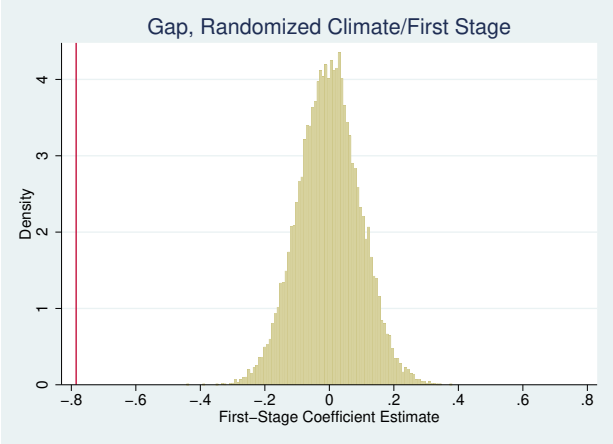
We complete this analysis in two steps. We first randomly permute our vector of climate instruments (severity of 1570s drought, maximum rainfall 1576–80, and the gap between minimum drought and maximum rainfall) 20,000 times and examine the performance of the instruments in first-stage, reduced-form, and instrumental-variables regressions using the permuted data. We then repeat the analysis permuting our explanatory variable of interest, the log ratio of 1650 over 1570 density, in 20,000 trials. The statistics reported here are from the OLS and IV specifications containing the full set of covariates presented in Tables 1 and 2 of the paper.

We begin with the permuted climate data. Figure A2 plots the distribution of coefficient estimates for first-stage regressions (predicting the log ratio of 1650 to 1570 density) using the permuted climate data. The red vertical lines in the figure represent the value of the statistics in regressions using the actual data. As expected, the coefficient estimates using the permuted data are centered around 0. In only 61 out of 20,000 trials was the coefficient estimate on drought as far from 0 as in the true data (0.30%). The corresponding number for the coefficient on maximum 1576–80 rainfall was 48 out of 20,000 (0.24%). In none of the permutations was the coefficient on the gap measure as far from 0 as in the regressions using the true data.

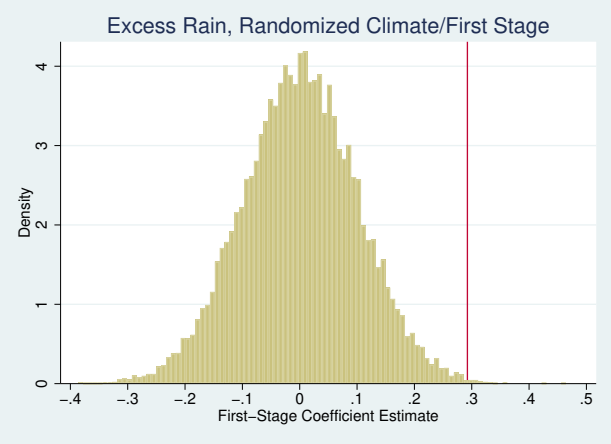
Figure A2: First-stage coefficient estimates, randomized climate



(a) $\hat{\beta}_1$ distribution



(b) $\hat{\beta}_2$ distribution



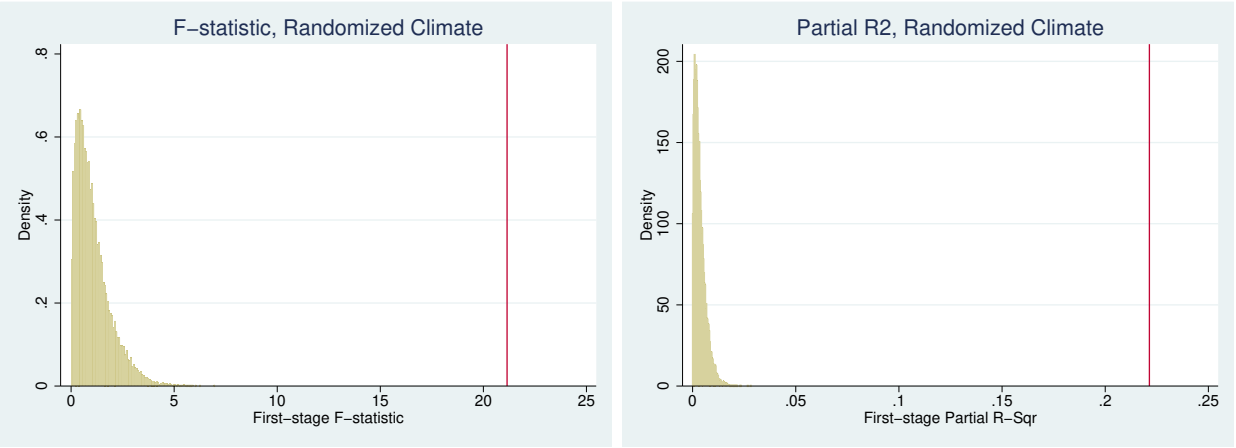
(c) $\hat{\beta}_3$ distribution

Panels show the distribution of the first stage coefficient estimates on drought (panel a.), the difference between rainfall and drought (panel b.), and excess rain (panel c.) using 20,000 random permutations of the climate data.

In Figure A3 we examine the performance of the permuted climate instruments in the IV regressions. We plot the first-stage F-statistics on the excluded climates instruments (left) and the partial R^2 (right) in the IV estimations from these 20,000 trials. As before, statistics using the true data are indicated with vertical lines. Using the permuted data, the distribution of these statistics is clustered around 0 as would be expected under the null. In none of the regressions are either of the statistics as high as in the regressions using the true

data. This provides additional evidence on instrument relevance in the true regressions.[27]

Figure A3: IV First-stage statistics: randomized climate



(a) F-statistic

(b) Partial R^2

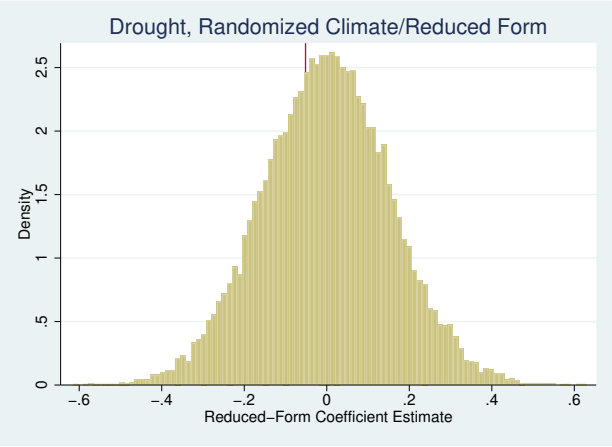
Panels show the distribution of the first stage F-statistic (panel a.) and the first stage partial R2 (panel b.) using 20,000 random permutations of the climate data.

Turning attention to the reduced-form regressions (predicting hacienda presence in 1900), we plot the point estimates of the coefficients on the permuted climate instruments in Figure A4. As before, the coefficient estimates using the permuted climate data are centered on 0. None of the coefficient estimates on the gap measure are as large in magnitude as in the true regressions, and only one of the coefficient estimates on the maximum rainfall measure exceeds the estimate from the actual regression. The coefficient estimates on the drought measure are at least as far away from 0 as in the true regression 74% of the time. As in the true model, the coefficient is thus not statistically distinguishable from 0.

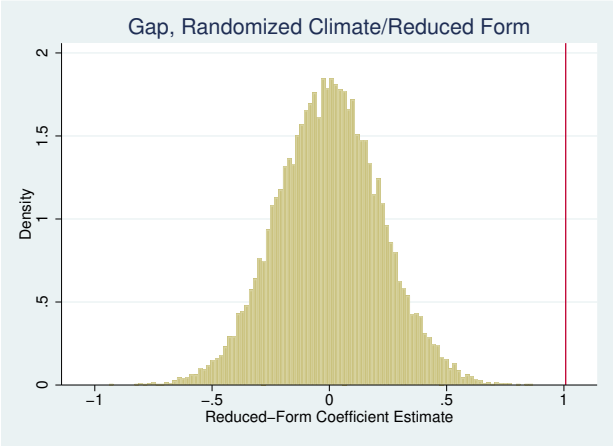
We repeat the analysis using the true climate data and permuting our explanatory variable of interest, the log ratio of 1650 to 1570 population density, 20,000 times. Figure A5 presents the distribution of coefficient estimates and t-statistics from OLS regressions using the permuted density data. We see that both the coefficient estimate and t-statistic are

[27]Because the permuted instruments are not relevant, second-stage regressions are biased and inconsistent. Magnitudes of second-stage coefficient estimates are extremely unstable in these regressions as would be expected using irrelevant instruments.

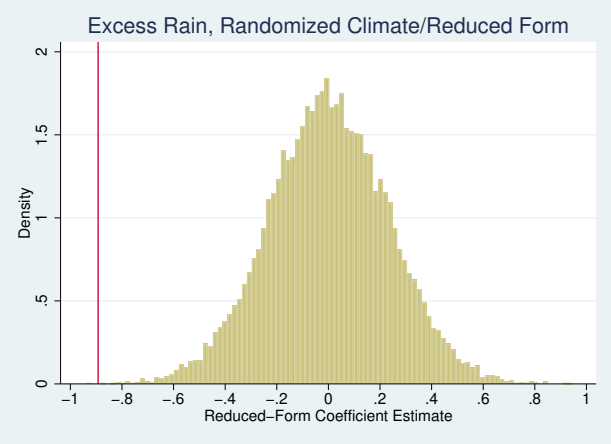
Figure A4: Reduced-form coefficient estimates, randomized climate



(a) $\hat{\beta}_1$ distribution



(b) $\hat{\beta}_2$ distribution



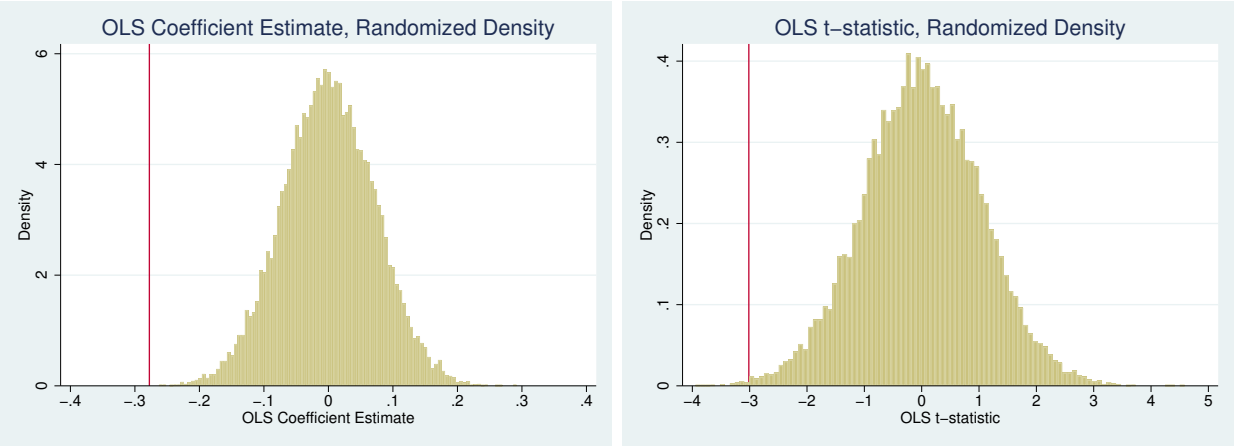
(c) $\hat{\beta}_3$ distribution

Panels show the distribution of the reduced form coefficient estimates on drought (panel a.), the difference between rainfall and drought (panel b.), and excess rain (panel c.) using 20,000 random permutations of the climate data.

centered on 0, as would be expected under the null. Examining the t-statistic, the permuted data generated t-statistics this far or farther away from 0 about 0.39% of the time, so the randomization-based p-value of 0.0039. The coefficient estimate on density ($\hat{\beta}$) was only as large in magnitude as in regressions using the true data in 2 out of 20,000 the regressions using permuted data (0.01%).

In Figure A6 we examine the performance of the instrument using the permuted density data. As before, F-statistics and partial R^2 values are small and close to 0 in almost all of these permuted regressions. No estimations yielded an F-statistic of partial R^2 as high as in the regression using the real data. Together, this analysis increases our confidence in the relevance of the instruments in the true regressions.

Figure A5: OLS statistics, randomized density

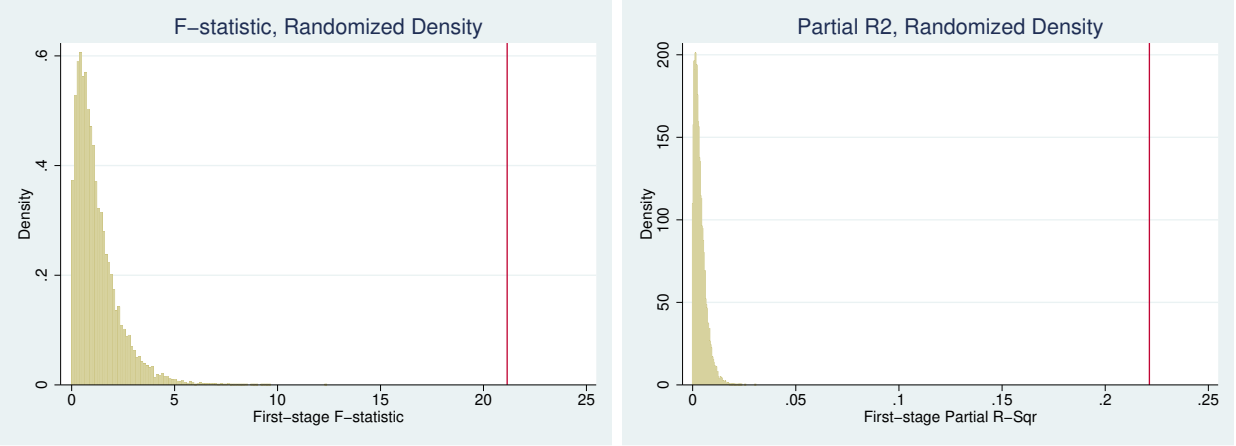


(a) $\hat{\beta}$ distribution

(b) t-statistic distribution

Panels show the distribution of the second stage coefficient estimates of the population drop (panel a.) and the t-statistic (panel b.) using 20,000 random permutations of the population density data.

Figure A6: IV First-stage statistics: randomized population density



(a) F-statistic

(b) Partial R^2

Panels show the distribution of the first stage F-statistics (panel a.) and partial R^2 for the first stage (panel b.) using 20,000 random permutations of the population density data.

Appendix A.5 Falsification exercise using climate data from alternate years

One might be concerned that weather shocks from the 1570s could be correlated with important omitted geographic or economic characteristics that might also affect later land concentration. For example, if climate conditions are highly persistent across time and space, it is possible that areas where climate conditions were conducive to the spread of cocoliztli in the 1570s may have received similar swings between drought and rain at other times. We do not have reason to believe that this is the case, and the alternative mechanism that would link susceptibility to such shocks to land concentration is not clear. However, we consider this possibility more rigorously in this section.

Given that PDSI measures are normalized to account for climatically appropriate conditions in each grid point, the potential for geographic confounds should be minimized in this setting. Furthermore, we are able to condition on average PDSI (to account for susceptibility to drought) and the standard deviation of PDSI (to account for variability of climate) in our estimations to account for potential confounds related to climate. As an additional way of addressing potential concerns about the exogeneity of the climate instruments, we complete a falsification exercise using alternate years of climate data. To conduct this exercise, we extract the smoothed climate data from the North American Drought Atlas for each year of the colonial period (1519–1810) using the identical procedure as described in the paper for the real instruments. We then treat each year of data as a potential "year 0" for a cocoliztli epidemic, calculating our instruments as we do for the epidemic beginning in 1576: we find the two lowest consecutive PDSI values in the preceding five years (to capture prolonged, severe drought), the maximum PDSI in year 0 and subsequent four years (to capture excess rainfall), and the swing between the minimum PDSI in the preceding five years and the maximum PDSI in the five years following year 0.

Once we have calculated the instruments for each year, we complete a series of falsification tests using these alternate years of data. After omitting the 1570s and considering leads and

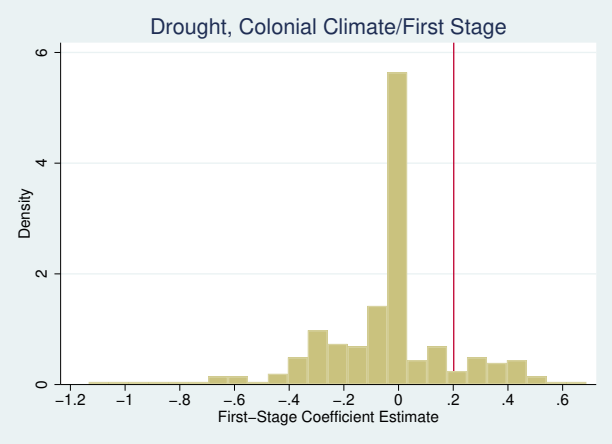
lags necessary to calculate the instruments, we are left with 280 years of data for the placebo exercise. We iteratively replace our true instruments with the instruments calculated for each of the alternate years and replicate our saturated (i.e., including all covariates) first-stage, reduced-form, and two-stage least squares regressions using these placebos. We store the regression coefficients and other test statistics from each of these regressions.

Figures A7 and A9 present histograms of the distribution of first-stage and reduced-form coefficient estimates respectively. The vertical lines in each plot represent the values of the coefficient estimate using the real instruments. As expected, the coefficient estimates from the placebo regressions are concentrated around 0 in both first-stage and reduced form specifications. In the first stage models, the magnitude of the coefficient estimate on the principal instrument (the drought-rain gap) is only as large as that of the actual regressions in 2 out of 280 estimations, and almost all estimates are close to 0 and not significant. The coefficient estimates on the other two instruments are not statistically distinguishable from 0 in the true first-stage regressions (Table 1), but they are still larger in magnitude than those of the majority of the placebo regressions (183 of 280 years for the drought instrument and 226 out of 280 years for the excess rain instrument). Turning attention to the reduced-form regressions, the coefficient estimates are again clustered around 0 and usually not statistically distinguishable from 0. The coefficient estimates smaller in magnitude than those of the true regressions in 269 out of 280 regressions for the gap instrument and 261 out of 280 regressions for the excess rain instrument. The drought instrument is not statistically different from 0 in the estimation in Table 1 and is roughly the same in the placebo regressions.

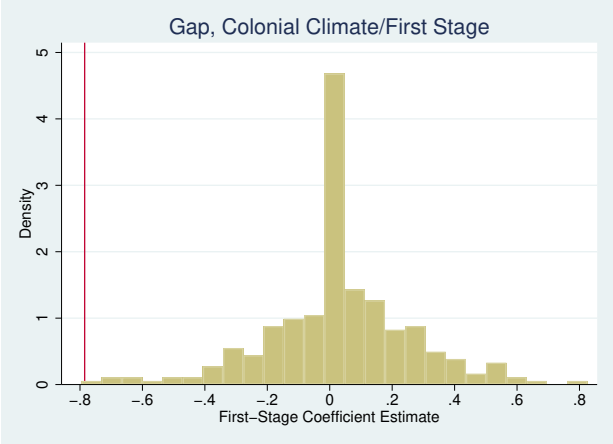
We also examine the performance of the vector of climate instruments in the placebo years in two-stage least squares regressions. In Figure A8 we plot first-stage F-statistics and partial R-squared of the excluded instruments using the alternate years. The distribution of values is roughly what would be expected under the null of non-relevant instruments. The first-stage F-statistic is only as large as in the actual regressions in 6 out of 280 estimations. The partial R-squared is only as large in 1 of the 280 regressions.

Given this evidence, we conclude that it is unlikely that other climate characteristics related to these shocks have direct effects on land concentration or population dynamics other than through the proposed channel.

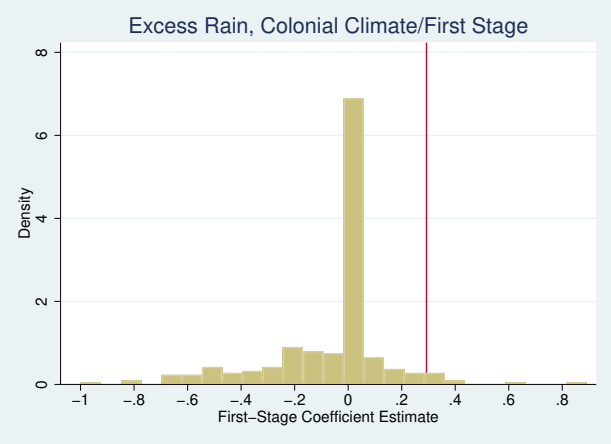
Figure A7: First-stage coefficient estimates, alternate years



(a) $\hat{\beta}_1$ distribution



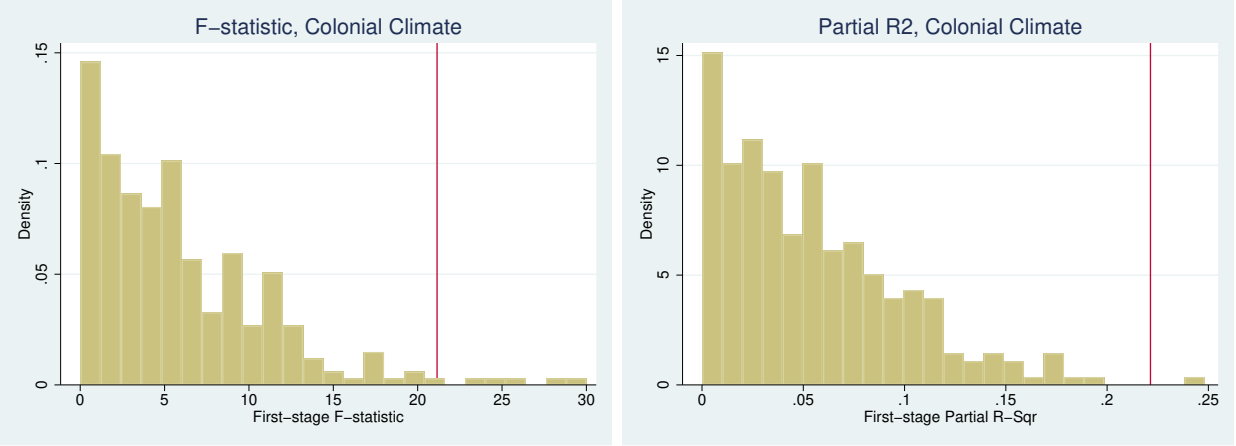
(b) $\hat{\beta}_2$ distribution



(c) $\hat{\beta}_3$ distribution

Panel a. shows the distribution of the estimated coefficients on the variable measuring intensity of drought in placebo years, panel b. the distribution of the estimated coefficients on the variable measuring the gap between rainfall and drought for placebo years, and panel c. the distribution of the estimated coefficient on the variable measuring excess rain for placebo years.

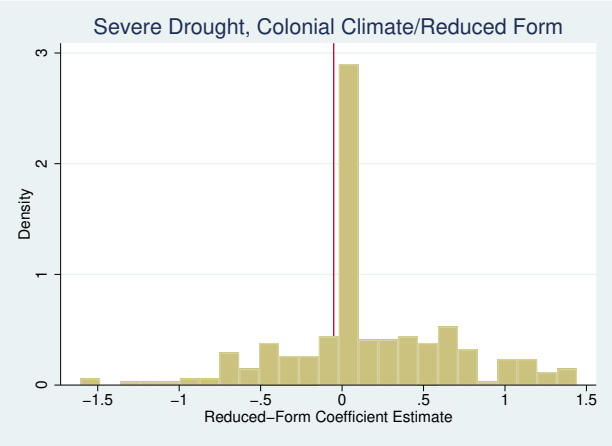
Figure A8: IV First-stage statistics, alternate years



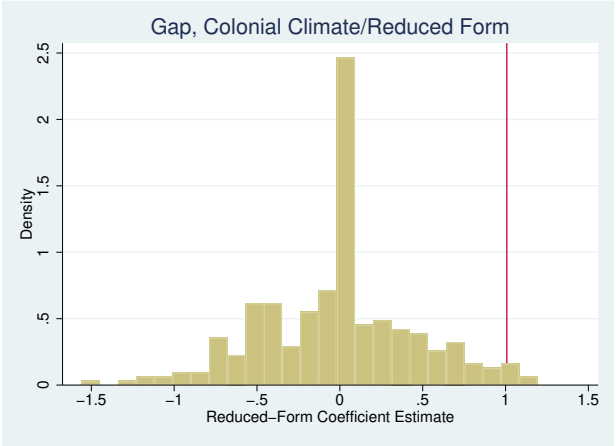
(a) F-statistic (b) Partial R^2

Panel a. shows the distribution of the first stage F-statistics using placebo years of colonial weather and panel b. the distribution of first stage partial R2 using placebo years.

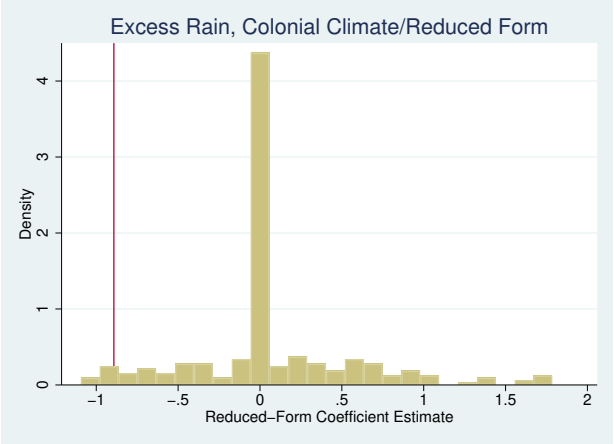
Figure A9: Reduced-form coefficient estimates, alternate years



(a) $\hat{\beta}_1$ distribution



(b) $\hat{\beta}_2$ distribution



(c) $\hat{\beta}_3$ distribution

Panel a. shows the distribution of the estimated coefficients on the variable measuring intensity of drought in placebo years, panel b. the distribution of the estimated coefficients on the variable measuring the gap between rainfall and drought for placebo years, and panel c. the distribution of the estimated coefficient on the variable measuring excess rain for placebo years.