Locational fundamentals, trade, and the changing urban landscape of Mexico*

Jennifer Alix-Garcia† Emily A. Sellars‡

October 21, 2019

Abstract

Geography and history shape urbanization, but the importance of these factors may change over time. We trace Mexico's urban development over 450 years using detailed subnational data on population, locational fundamentals, and trade access. The analysis shows that urbanization patterns remained largely unchanged from the colonial period until the mid-20th century despite major economic shocks, such as the collapse of Mexico's indigenous population, the War of Independence, and the Mexican Revolution. However, the pull of historical population and the nature of geographic advantage were disrupted during the second half of the 20th century as international trade access began to play an important role in the emergence of new cities. The results indicate that while geography and history continue to influence urban development, economic policy can reshape these forces in important ways.

JEL codes: R12, O18, N96
Keywords: urbanization, geography, trade, historical persistence

*Acknowledgments: We are grateful for comments from participants in workshops at the University of Wisconsin, Texas A&M University, Simon Fraser University, the University of British Columbia, and Oregon State University and at the Stanford/LSE/Uniandes Conference on Long-Run Development in Latin America and the Princeton Workshop on Cities, Inequality, and Electoral Politics. We are also indebted to guidance from Leah Boustan and two anonymous reviewers.

†Department of Applied Economics, Oregon State University, email: jennifer.alix-garcia@oregonstate.edu
‡Department of Political Science, Yale University, email: emily.sellars@yale.edu

1 Introduction

History and geography loom large in the discussion of where and why cities emerge and grow. Geographic factors—the richness of soils, access to navigable rivers, the ruggedness of terrain, and other persistent features of the landscape—are thought to determine a location's attractiveness for initial urban and economic development. Once established, areas of early population and wealth concentration may build on themselves over time through the benefits of local increasing returns, expectations of future investment, the positive relationship between economic development and well-functioning institutions, and other factors (Krugman 1991, Davis & Weinstein 2002, Bleakley & Lin 2012, Henderson et al. 2018). However, major economic or policy shifts, such as industrialization or the opening of an economy to trade, may fundamentally alter the nature of geographic advantage. This paper examines how major changes in Mexico's demography, economy, and society over the last 450 years have influenced the roles of historical precedent and geographic fundamentals in the country's urban landscape.

Mexico constitutes an ideal setting to examine sources of persistence and change in urban development. It is a large and diverse country in terms of geography, trade access, history, and contemporary demography. Though some parts of the country have been urbanized for over 500 years, Mexico remained predominantly rural until the 20th century. In 1900, around 12% of the population lived in cities exceeding 15,000 inhabitants. Today, Mexico's urbanization rate (80%) is similar to that of the United States. We use a combination of cross-sectional and panel methods to analyze 15-by-15-kilometer grid cells generated over the territorial extent of the country. We begin by examining the correlates of urbanization at regular intervals from the 16th century until 2010 to identify which geographic features—agricultural productivity, elevation, ruggedness, and proximity to waterways, the northern border, and Mexico City—have been correlated with urbanization in different periods and how patterns of population concentration have shifted over time. We then transform the grid cells into a decadal panel to illuminate the timing of important shifts in urbanization patterns

from 1900 until 2010, and their connection with specific economic policies and events.

The analysis shows that Mexico's regional population distribution has remained largely persistent since the early colonial era, even in the face of major political and economic shocks such as the massive population collapse after the Conquest, numerous civil wars, the Mexican Revolution, and the creation of the modern Mexican state. Covariates associated with historical population, agricultural productivity, proximity to ports and to internal trade centers strongly determined city location from the colonial era to 1900. However, as the 20th century progressed, proximity to the coast and to the U.S. border become driving forces in the emergence of modern cities. The explanatory power of trade-related variables increases more rapidly between 1900 and 2010 than the explanatory power of locational fundamentals or historical population, highlighting important changes in the determinants of urbanization over time. The panel analysis reveals that the correlates of city emergence remained the same from 1900 until around 1950, at which point proximity to the northern border and the coasts began to play a more important role. The timing of these changes is suggestive of the role that economic policies favoring trade and tourism played in shifting population northward and towards the coasts at the expense of the center of the country. It is also probable that advances in transport and health infrastructure over the 20th century reduced the relative importance of agricultural productivity and the disease environment, enabling rapid urbanization in previously disadvantaged coastal and desert areas. While historical persistence has always been an important force in Mexico's urban development, our work illustrates how policy has changed the nature of geographic advantage.

These findings contribute to ongoing debates in urban economics and economic geography over the roles of locational fundamentals and historical precedent in urbanization. Geography has long occupied a central place in the study of urbanization, as it has in the literature on economic development more generally (e.g., Pirenne 1925, Weber et al. 1958, Bairoch 1988, De Vries 1984, Mellinger et al. 2000, Sachs 2001, Nunn & Puga 2012, Henderson et al. 2018). The relationship between urbanization and "first-nature" (resource abundance, agri-

cultural productivity, waterway access, the presence of natural defenses, and the ruggedness of terrain) and "second-nature" (position relative to other areas of population concentration) geographic characteristics remains a central area of study. Recent work has focused attention on whether and why the relationship between geographic fundamentals and urbanization differs across time and space (e.g., Davis & Weinstein 2002, Bleakley & Lin 2012, Henderson et al. 2018). Our paper extends this line of inquiry by tracing the evolving importance of various first- and second-nature geographic factors over a long period time using subnational panel data. This approach allows us to more closely tie observed changes (or lack of change) in the geographic correlates of urbanization to specific policies and events, illustrating the shifting nature of geographic advantage under different levels of industrialization, transport availability, and trade politics.

The paper also speaks to the literatures on history and path dependence in urban economics. Past work has shown that historical precedent can shape the growth and longevity of cities in various ways. Areas of early population concentration may increase in size due to the benefits of increasing returns and agglomeration economies (e.g., Krugman 1991, 1993, Bleakley & Lin 2012, Maloney & Valencia Caicedo 2016), because of natural demographic growth from existing urban populations (Jedwab et al. 2014), or because existing cities serve as focal points for setting expectations about future investment and settlement (Bleakley & Lin 2012, Jedwab et al. 2015). City locations may also be persistent if specific areas remain geographically advantaged over time (e.g., Davis & Weinstein 2002). The notable persistence in Mexican population concentration from the early colonial period until the 20th century is consistent with each of these explanations, but 20th-century shifts in urbanization patterns call attention to the limitations of history-focused theories. While early colonial urbanization patterns proved resilient to some major shocks, such as the collapse of the indigenous population and the political transformations of independence and the Revolution, the rise of northern and coastal areas in the 20th century illustrate that economic policy can alter long-standing patterns of geographic advantage. By altering the benefits and costs of settlement

and investment in different locations, trade and tourism policies increased the attractiveness of previously remote areas and reduced some of the advantages of places that had been well connected. This is consistent with other work on Mexico and other contexts highlighting the importance of trade and tourism in urban development (e.g., Redding & Sturm 2008, Redding et al. 2011, Bosker et al. 2013, Faber & Gaubert 2019).

Our paper is most closely related to three recent papers on urban economics, geography, and historical persistence. Henderson et al. (2018) compare city development in early- and late-developing countries using nighttime lights data, finding evidence that inherent agricultural productivity has declined in importance for urban development over time due to improvements in transport technology. Like Henderson et al., our results suggest that the nature of geographic advantage has shifted due to changes in industrialization and trade, though our empirical approach is distinct. By using detailed panel data on population and geography within a single country over nearly 450 years, we are able to more closely examine the process of urbanization over time and link observed changes in the determinants of urbanization with specific economic and political factors. Our more granular approach also sets this paper apart from that of Maloney & Valencia Caicedo (2016), who document strong historical persistence in regional population concentration in Latin America using estimates of pre-Columbian population aggregated to the state or department level combined with contemporary data. Consistent with this work, we find strong evidence of persistence in population distribution over centuries using our more disaggregated measures of population. However, our work also highlights important changes in population concentration due to industrialization and shifts in trade policy. Our empirical approach allows us to identify the timing of these changes, and it has the additional advantage of not relying on subnational political divisions, the number and size of which have changed endogenously with population concentration (González Navarro 1974, Unikel 1976).

Our methodology is most similar to Bosker & Buringh (2015), who study the changing effects of geography on European city growth from 800 to 1800 using a similar grid-cell

approach. As in that paper, our findings highlight the importance of first-nature geography in urban development over time and draw attention to the ways that the nature of geographic advantage has changed over time. Our paper differs from theirs most directly in our focus on specific policy changes in the modern era, especially those related to the evolution from internal to external trade. Though our focus on geography is similar, the specific correlates of city emergence in Mexico are somewhat distinct from those in Europe, drawing attention to the importance of broader economic and political context in urbanization.

A final contribution of this paper is to provide detailed evidence on the urbanization trajectory of a country as it grew from low- to middle-income, a little-studied path that remains open to the majority of the world but is not yet well understood. Urbanization tends to increase as countries develop, and population concentration raises the prospects for further economic growth (e.g., Overman & Venables 2005, Jedwab et al. 2014). Several scholars have argued that city emergence and growth in countries that urbanized early were distinctive from patterns now being observed in "late developers" in Africa, Latin America, or Asia (Overman & Venables 2005, Jedwab et al. 2014, Henderson et al. 2018). The extent to which lessons from urbanization in medieval Europe or the Middle East can be applied to understanding urbanization in contemporary developing countries is thus unclear. In addition to providing evidence on the geographic factors that influenced urbanization at different points of Mexican history, this work illustrates the role that policy can play in structuring urbanization and regional development in a developing-country setting.

2 Data

2.1 Population and "urbanization"

We use several datasets to trace urbanization over time. We rely most heavily on the Historical Archive of Localities (AHL), which is maintained by Mexico's National Institute of Statistics and Geography (INEGI). This dataset contains the population and political

classification of the vast majority of localities in Mexico in each census from 1900 until 2010. We use text analysis to link localities, many of which have changed names or boundaries since 1900, to their geographic location over time. While the AHL dataset contains all censused localities in later years, it omits many small communities in the early 20th century (see Appendix A). To address this shortcoming, we cross-reference the data with the hard copy of the *División Territorial* for each census, which contains a comprehensive list of localities and their population, and incorporate data on the population of central Mexico City from Unikel (1976). Appendix A provides further detail on adjustments made to this dataset to make it usable for this project.

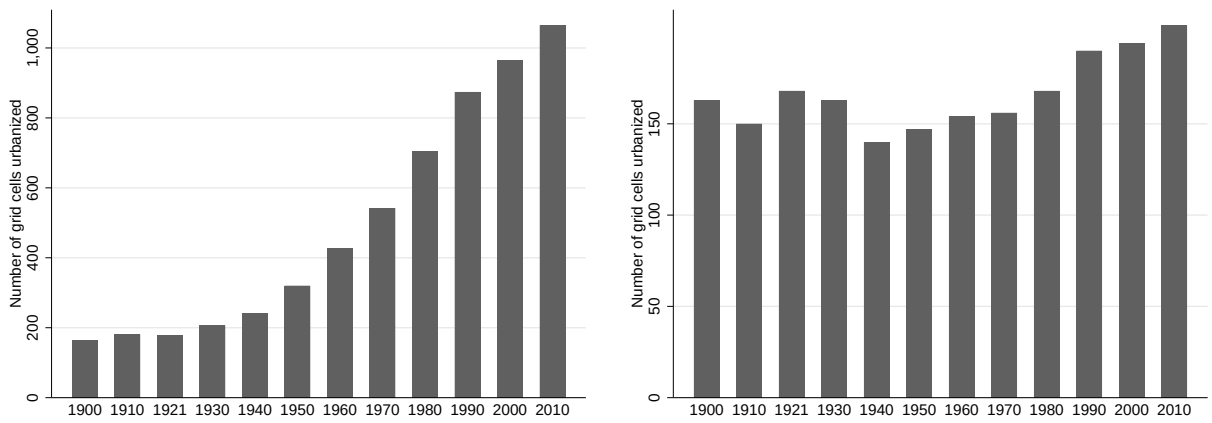
Once localities are linked to a geographic location, we aggregate the data to the level of 225 km^2 (15-by-15 km) grid cells, our primary unit of analysis. The grid approach addresses the problem that both contemporary and past administrative boundaries and metropolitan area distinctions have been determined by urbanization patterns and government policy (González Navarro 1974, Unikel 1976). This grid approach has been used in several other recent papers in urban economics (e.g., Bosker & Buringh 2015, Henderson et al. 2018). The cell size is equivalent to the area of the average Mexican municipality in 2010. We code a grid cell as "urban" if it contains one or more urban localities and zero otherwise.

There is no consensus in the literature regarding the definition of a city (e.g., Henderson 2005, Glaeser 2007). We consider two population thresholds to determine whether a given locality should be classified as "urban": an absolute threshold of 5,000 persons and a time-varying threshold equal to 0.0367% of the total national population (selected to be equal to the absolute 5,000-person threshold in 1900). While the 5,000-person city cutoff might seem small, the Mexican government uses a lower threshold (2,500 people) in its internal definition of urban localities. The 5,000-person definition has also been used in several classic and recent works in urban economics as it captures a minimal definition of population concentration across eras (e.g., Bairoch 1988, Bosker & Buringh 2015). The time-varying urban threshold

increases over time from around 5,000 in 1900 to 41,278 in 2010 due to population growth.[1]

Though less common in the literature, we believe that the time-varying threshold better captures the concept of urbanization in our context. As many others have noted, the nature of what constitutes an "urban agglomeration" has changed over time due to technological change, agglomeration effects, and other factors (e.g., Black & Henderson 1999, Henderson 2005). Using the flat 5,000-person cutoff, over six times as many grid cells are considered "urban" in 2010 as in 1900 (163 vs. 1,065). Not all represent major metropolitan areas or even "urbanized" populations under standard economic definitions (e.g., dominance of industrial/professional over agricultural workforce.). Under the varying threshold, only 204 grid cells are "urban" in 2010, highlighting more important centers of population and settlement. Though we focus primarily on results using the time-varying definition in this paper, we show results of all models using the 5,000-person definition in the appendix to facilitate comparison with other work (e.g., Bosker & Buringh 2015).

Figure 1: Alternate ways of measuring urbanization



(a) > 5,000 (b) > 0.037% population

Figures show the number of grid cells urbanized by decade according to the AHL data, using population cutoffs of 5,000 (a) and 0.0367% of total national population (b)

Figure 1 shows the distribution of the two "urban" indicator variables over time. The

[1]The time-varying thresholds are 5,571 (1910); 5,267 (1921); 6,082 (1930); 7,222 (1940); 9,477 (1950); 12,821 (1960); 17,720 (1970); 24,563 (1980); 29,855 (1990); 35,820 (2000); 41,278 (2010).

number of urbanized grid cells under the 5,000-person cutoff grows dramatically over the 20th century, with the steepest increases occurring between 1940 and 1970, the height of Mexico's demographic boom. Using the relative threshold for city size, by contrast, the size and spread of urbanization declines during the 1930s and increases only slightly over time. This is more consistent with qualitative information about the relative growth of rural and urban populations in this context. Several scholars have documented the de-urbanizing effects of Mexico's agrarian reform, which peaked in the 1930s (e.g., González Navarro 1974, Warman 2001). The temporal pattern under the time-varying definition is also consistent with the concentrated spatial nature of urbanization under import-substitution (e.g., Unikel 1976), where a disproportionate amount of urbanization was concentrated in Mexico City and surrounding areas (thus limiting "new" urbanization in other areas).

In the appendix, we repeat our analyses using grid-cell population density as an alternative measure of demographic and economic concentration, following Davis & Weinstein (2002) and others. However, we note that this measure is potentially problematic in our context as population is undercounted in the AHL relative to more aggregated census data prior to the 1970s (a difference of about 13% or 2 million people in 1900, for example). Moreover, this undercounting varies across space.[2] We also believe that population density and urbanization are somewhat distinct phenomena. As far back as 1900, municipalities (akin to counties in the U.S.-context) containing a locality exceeding our urbanization threshold had much greater numbers of professionals and students than did municipalities without, conditional on population density (Appendix Table C2). The same is true of various education and employment outcomes in 2010 using our time-varying measure of urbanization (Appendix Table C3).

[2]This arises in part because the AHL does not typically trace the populations of localities that ceased to exist by 1970, which are disproportionately rural. See Appendix A.

2.2 Colonial population

Though comprehensive, locality-level data[3] in Mexico are unavailable prior to the 20th century, we examine district-level data on population concentration beginning in 1570. We do this for two reasons. First, we expect that the determinants of population concentration in the colonial era may have been different than would be expected today given differences in technology, economic structure, and politics. Second, population density during this "baseline" time period may continue to shape urban development due to increasing returns, path dependence, or other factors.

Our colonial population data were digitized from Gerhard (1993*a*, 1993*b*, 1993*c*), who uses the *Relaciones geográficas* as his primary source of information.[4] We examine three snapshots of colonial population: in 1570, the first year in which we have data for much of the colony; in 1650, which is close to the nadir of indigenous population following Mexico's demographic collapse[5]; and in 1800, just before the War of Independence. Gerhard's data are reported at roughly the level of 1786 political divisions and are typically recorded in terms of tributaries, or the number of individuals paying tribute to the Spanish Crown. Each married male counted as one tributary, and unmarried adults and widows counted as half of a tributary. The ratio of tributaries to total indigenous population was approximately 2.8 in the late 16th century (Gerhard 1993*a*, Cook & Borah 1960, 1971). In 1800, densities are given in total population rather than tribute units. These data and their reliability are described in Gerhard (1993*a*) and Sellars & Alix-Garcia (2018). Figure 2 in Section 3.1 presents the spatial extent of Gerhard's data. Detailed population information is unavailable for much of northeastern Mexico in the early colonial period, as the region had not yet come

[3]The dataset provided by Buringh (2018) provides a rough estimate of localities across the world that exceeded 5,000 people in size at various historical intervals based on interpolated population data from a variety of official and unofficial sources. According to this source, Mexico had 12 urbanized cities as of 1650 and 24 as of 1750, predominantly in the center of the country. However, these figures should be interpreted with caution. See also Arroyo Abad & van Zanden (2016).

[4]The *Relaciones geográficas* were a series of reports dispatched to Spain from the New World beginning in the sixteenth century with detailed information on population, language, geography, and other information about colonial holdings. See Gerhard (1993*a*), p. 28–31.

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

under solid Spanish control.

2.3 Location fundamentals and trade costs

We create several measures of locational attractiveness for urban development using geographic data. These measures include the standard deviation of elevation (a measure of terrain ruggedness) and an indicator representing whether the average elevation of a grid cell is lower than 1500 meters. This second metric is meant to proxy for the disease environment as malaria, yellow fever, and other diseases inhibited colonization of low-lying areas (e.g., Hassig 1985). Both measures were extracted using a 30-meter digital elevation model. We also include a measure of potential low-input, rain-fed maize productivity, which was extracted from the Food and Agriculture Organization's Global Agro-Ecological Zones (GAEZ) dataset (IIASA/FAO 2012). The GAEZ potential productivity measure is calculated using exogenous information on climate, soil type, slope, and average rainfall calculated at a 5-arcminute resolution (about 10-km cells in Mexico). The potential productivity of maize is highly correlated with that of many other grain and vegetable crops, and maize has been the primary staple crop in Mexico for centuries.

Mexico has no major navigable rivers. However, we include freshwater access among our proxies for geographic advantage, which is calculated as distance to the nearest river or lake using contemporary spatial data from INEGI.[6] Because our study spans a great number of years, direct proxies for transportation costs and internal trade are limited. We calculate two measures of advantageous access to external trade: the Euclidean distance from each grid cell to the US border and to the nearest "anchorage area" (NOAA 2014). Anchorage areas include both ports and natural harbors. Ports are places where infrastructure allowing for trade through shipping already exists, and natural harbors are areas with sufficient depth and protection to allow ships to drop anchor. These points are shown in Figure 2. We include

[6]Though there have been some shifts in the placement of rivers, streams, and lakes over time because of human intervention and other factors, we do not have good data on the extent of surface water prior to the current era.

distance to coast as a control variable to help isolate the effect of ports on city development separate from ocean access. As a proxy for internal trade pull, we calculate the distance from each grid cell to the nearest city with a population greater than 15,000 in 1878 (Unikel 1976). As we discuss below, we also qualitatively examine the effect of changes in trade and transport by comparing trends in city emergence and growth in different time periods.

3 Correlates of urbanization in Mexico

We begin by estimating the correlates of urban concentration over time using a series of cross-sectional regressions of the following form:

$$S_{us} = F(\alpha + \mathbf{X_u}\beta + \epsilon_{us}) \tag{1}$$

The dependent variable is an indicator (S_{us}) capturing whether a grid cell includes a locality with population above the time-varying threshold (or over 5,000) in a given year. The baseline models use a probit specification ($F(\cdot) = \Phi(\cdot)$) to ensure that predicted probabilities are between 0 and 1.[7] The matrix $\mathbf{X_u}$ includes the geographic covariates discussed above.

To control for potential congestion or agglomeration effects that might confound the analysis, we calculate the presence of cities within a 50-km radius of each grid cell and condition on these measures in some specifications.[8] Locations that are too far from other cities may be too remote from trade networks to develop, while those that are too close may be crowded out by the urbanization of neighboring areas (e.g., Krugman 1993, Anas et al. 1998, Fujita & Mori 1996, Duranton 1999, Konishi 2000, Behrens 2007, Bosker & Buringh 2015). Though these spatial spillovers are not a primary focus of the analysis, it is important to include them as our measures of policy, geography, and history are also spatially correlated. We deliberately do not include state fixed effects, as we are interested in

[7]The marginal effects using a linear probability model are nearly identical. We report the probit specification in the text because it is used as the basis for the simulations in Section 4.

[8]We also tested a 25 km radius, but found no interesting differences in results, so we present only the 50 km radius.

measures of geographic access and locational quality across the entire landscape and do not wish to identify these effects off of within-state variation. Unless otherwise noted, standard errors are clustered at the level of modern Mexican states.

3.1 Urban development in Mexico, 1519–1900

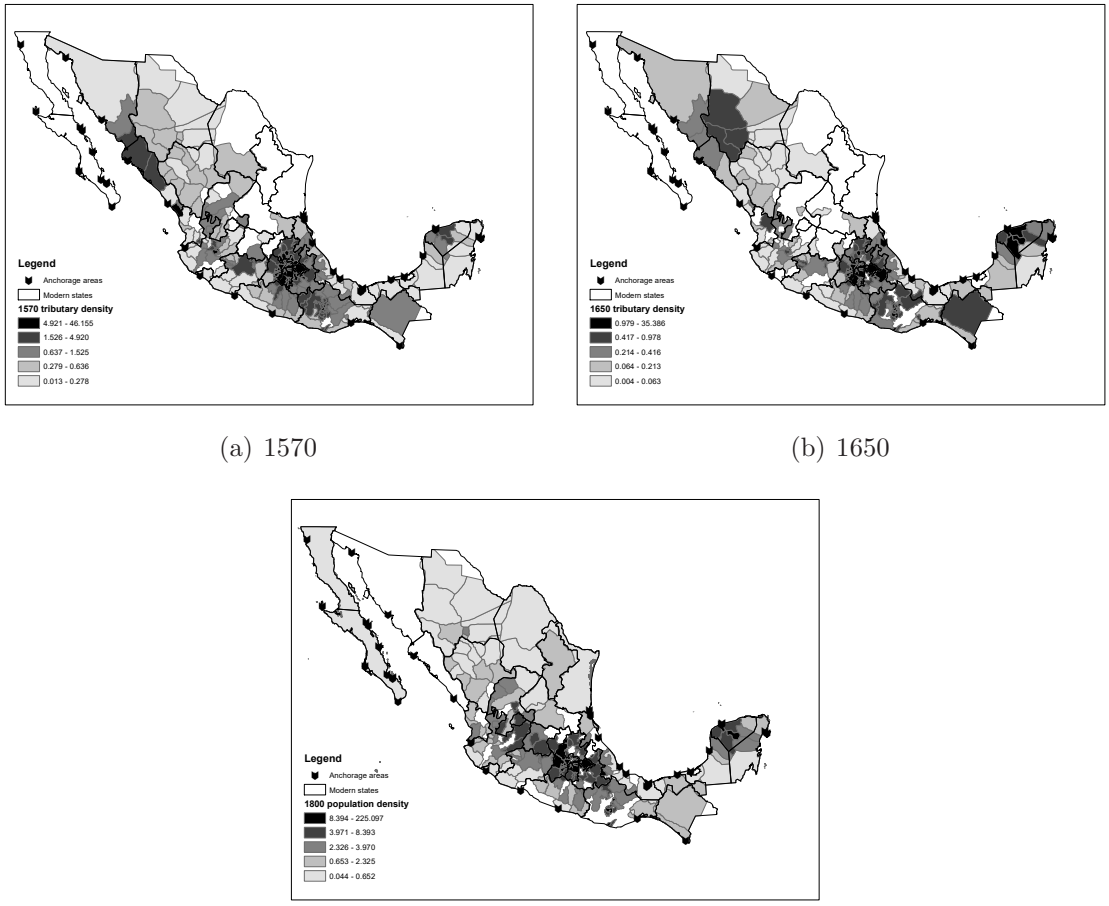
Geographic features played an important role in pre-colonial population concentration. At the time of Spanish contact in 1519, the Triple Alliance (Aztec Empire) was a large and urbanized society with as many as 20 million residents (Gibson 1964, Cook & Borah 1971, Hassig 1985). The nexus of population and power was the altiplano in the center of the country, which had several geographic advantages. The area around Tenochtitlan (now Mexico City) is located outside of the tropical zone in an area where maize productivity is relatively high. Though this area is mountainous, Tenochtitlan was located in the middle of a large lake system, which was used for trade with hinterland areas in a context where overland transit was extremely costly given the mountainous terrain and absence of pack animals (Hassig 1985, Ch. 2).

The Spanish Conquest ushered in numerous political and economic changes. Most notably, it was followed by one of the most dramatic demographic collapses in history. By 1650, a combination of disease, drought, and famine reduced the indigenous population by over 90% by most estimates (e.g., Cook & Borah 1971, Hassig 1985). The magnitude of the collapse has few parallels in history, the closest historical comparison arguably being the depopulation of Europe during the Black Plague. We examine the extent to which this massive demographic shock reshaped the geographic patterns of population concentration.

Figure 2 presents the geographic distribution of population concentration at three points during the colonial period: around 1570 (the first year in which our data are relatively complete), around 1650 (close to the nadir of Mexico's population), and 1800 (at the end of the colonial era). Darker colors indicate areas of greater population/tributary density, as measured by quantiles. While comprehensive population data from the time of the Conquest

do not exist, the distribution in the 1570 map conforms with qualitative descriptions of pre-Columbian settlement patterns from other sources (e.g., Cook & Borah 1971, Hassig 1985, Gerhard 1993*a*).

Figure 2: Colonial era population densities



(a) 1570

(b) 1650

(c) 1800

Shading represents within-year quintiles of density, with darker colors representing higher density. White spaces indicate missing data. 1570 and 1650 density is measured in terms of tributaries to the crown, while 1800 density is measured as total population per square kilometer. Markers indicating anchorage areas are also shown.

A striking feature of Figure 2 is the persistence of population concentration throughout the colonial period. The area around Mexico City retained its relative dominance, even at the height of the demographic collapse in 1650.[9] This is also true of other centers of pre-colonial

[9]Note that the scale of the maps is recorded in quantiles; the 1650 density numbers are of a much smaller

population density, such as the Yucatán Peninsula. Consistent with other historical work (e.g., Unikel 1976, Knight 2002), some shifts towards new mining and agricultural centers in the center-north and west, such as Guadalajara and Guanajuato, are evident during the later colonial period, but these changes were relatively modest. After nearly 300 years of Spanish rule, population density at the end of the colonial period in 1800 remained highly correlated with population density in 1650 ($\rho = 0.70$) and 1570 ($\rho = 0.83$). In the appendix (Table B1), we show that the sizable, positive correlation between density across time periods persists even when conditioning on geographic and other confounds.

Though consistent with other work on historical persistence in Latin America and elsewhere (e.g., Davis & Weinstein 2002, Maloney & Valencia Caicedo 2016), this finding is somewhat surprising given the context. Mexico's unprecedented population collapse and the transition to colonial rule forced a major reorganization of religious, social, and political institutions (e.g., Hassig 1985, Gerhard 1993*a*). Furthermore, important changes to economic geography in the colonial era, such as the draining of Mexico's lake system and the rise of the mining economy, had muted effects on the overall distribution of population. Even in the face of these shocks, however, populations generally rebounded in the same locations that they had occupied prior to the collapse.

This persistence is consistent with classic Malthusian theories about natural population capacity of land (e.g. Davis & Weinstein 2002), but evidence suggests that political and economic institutions played an important role as well. Early colonial institutions were designed to usurp the existing tribute system of the Triple Alliance, encouraging investment and settlement in areas of dense pre-colonial population (Gibson 1964, Hassig 1985, Gerhard 1993*a*). Internal migration was hampered by the tribute system, which kept the indigenous population tied to specific villages, and by the high costs of overland transport, which encouraged population concentration in the grain-producing areas of central Mexico (Unikel 1976, Coatsworth 1978, Knight 2002). Though transformative, the collapse did not alter the

magnitude than those of 1570.

long-run spatial equilibrium of settlement, investment, and political power in the center of the country.

In fact, there was little change in the geographic distribution of population until the 20th century. Following independence from Spain, Mexico experienced stagnant demographic and economic growth for most of the 19th century due to ongoing warfare, economic uncertainty, and political fragmentation.[10] Though the Porfirian era (1876–1910) brought important political and economic changes and a period of sustained demographic growth (e.g. González Navarro 1974, Unikel 1976, Coatsworth 1978), Mexico remained predominantly rural. In 1900, around 65% of the working age population was employed in agriculture (Reynolds 1970), and nearly 90% of the population lived in localities of less than 15,000 people.

Table 1 examines the geographic and historical correlates of city location in 1900.[11] Column (1) presents the baseline specification including the full set of geographic and trade-related covariates described in Section 2. In the subsequent columns, we add various measures of historical persistence and neighborhood effects to examine the role of second-nature geography and historical precedent on city development.

Column (2) adds early colonial (1570) population density to the analysis. This reduces sample size considerably as early colonial population data are not available for the entire country. The specifications in columns (3) and (4) employ two approaches to extend the coverage of the dataset. First, we impute 1570 population density using a vector of geographic and historical covariates. This imputation comes from the estimations in column (1) of Appendix Table B1, which examines the correlates of colonial population density. Second, we replace missing population values with zeros and include an indicator variable for whether the data were missing. The inclusion of historical population density does not substantively change the coefficient estimates of the other covariates, and point estimates are nearly equivalent using either approach to improve geographic coverage. The final two

[10] These factors are described in depth in Coatsworth (1978) and Haber (1992), among others.

[11] Note that the 5,000-person and time-varying urban thresholds are identical in 1900 by design.

Table 1: Correlates of urban grid cells in 1900

	Dep. var: Population > 5,000 = 1					
	(1)	(2)	(3)	(4)	(5)	(6)
Geographic fundamentals						
Elevation < 1500 m	-0.013**	-0.014*	-0.011*	-0.011*	-0.012*	-0.012**
	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
Std grid cell elevation	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ln(km coast)	0.008***	0.012***	0.009***	0.009***	0.009***	0.009***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
Ln(km water)	-0.000	0.000	-0.000	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(maize productivity)	0.012***	0.007**	0.008***	0.009***	0.009***	0.009***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Trade covariates						
Ln(km to 1878 city, Mexico)	-0.006***	-0.006***	-0.006***	-0.006***	-0.008***	-0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(km to border)	-0.001	0.006	-0.003	-0.002	-0.003	-0.002
	(0.003)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
Ln(km anchorage)	-0.013***	-0.017***	-0.013***	-0.014***	-0.014***	-0.014***
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
Historical population						
Ln(Tributary density, 1570)		0.006***				
		(0.002)				
Ln(1570 trib density, imputed)			0.006***		0.006***	
			(0.002)		(0.002)	
Ln(1570 trib density, zeros for missing)				0.006***		0.006***
				(0.002)		(0.002)
= 1 if missing 1570 data				-0.004		-0.003
				(0.003)		(0.003)
City of 15000, t-1, 50 km					-0.016**	-0.015**
					(0.006)	(0.006)
Observations	9342	6491	9342	9342	9342	9342
Adjusted R^2	0.36	0.38	0.40	0.40	0.40	0.40

Unit of observation is the 15 x 15 km grid cell. Clustered standard errors are in parentheses.
Coefficients robust to excluding Mexico City area. Reported coefficients are marginal effects from a probit.
* p< 0.10, ** p<0.05, *** p < 0.01.

columns introduce an indicator for there was a city of 15,000 within a 50-km radius as of 1878 to account for neighborhood effects.

Consistent with standard theories, urbanization in 1900 is positively associated with local agricultural productivity and negatively with distance to anchorage areas (i.e., more urbanization occurs closer to trade access points). Low-lying areas are less likely to be urban, potentially due to the disease environment (e.g., Hassig 1985), as are areas near the coast (conditional on distance to ports/anchorage areas). These geographic patterns are similar to those seen during the colonial era (Appendix B), which is suggestive of persistence in the correlates of population concentration over time. Distance to 1878 urban centers is negatively related to urbanization, indicating a pull of internal hubs that outweighed any "crowding out" effects.[12] By contrast, locations nearer to the U.S. border are not associated with increased urbanization. Finally, as with the population data for the late colonial period, historical population density in 1570 is positively and strongly associated with urbanization in 1900, even conditional on the other covariates.

At the start of the 20th century, Mexican demographic patterns resembled those of the colonial and even pre-colonial era in many respects. Most of the country was rural. Though trade and economic restructuring had shifted populations somewhat toward the coasts and mining areas relative to the 16th century (see Appendix B), settlement and investment remained concentrated in traditional population centers. There were only six cities with population above 50,000 in 1900—Guanajuato, Monterrey, Puebla, San Luis Potosí, Guadalajara, and Mexico City—all of which had been important population centers during the colonial period.

3.2 Mexico's Urban Transformation

The 20th century brought significant changes. There are many more cities in Mexico today than there were in 1900. Figure 3 presents the location of urban centers in contemporary

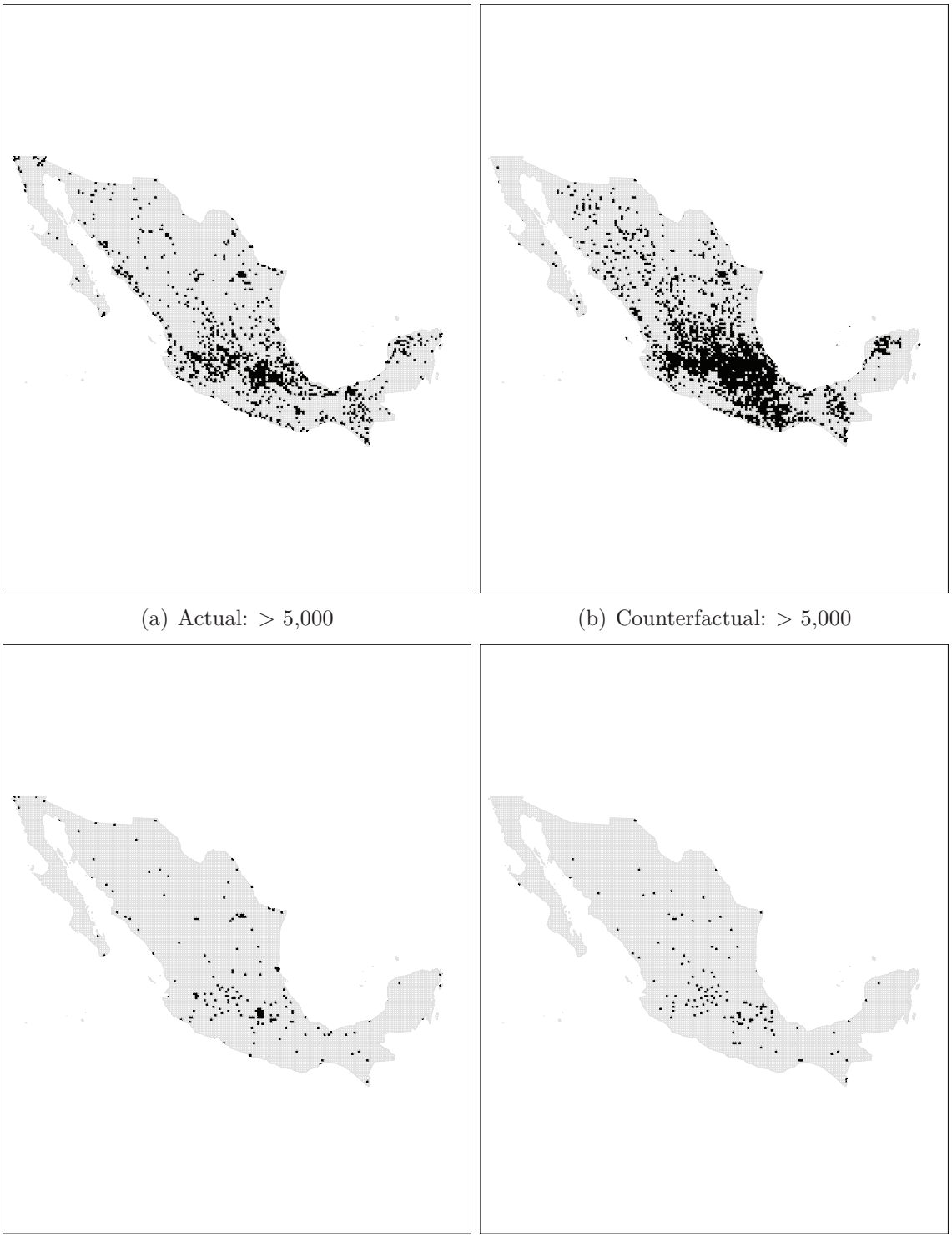
[12]Though the coefficient on the spatial lag term is negative, the important observation for our purposes is that the inclusion of the spatially lagged variables does not substantively change the other point estimates.

Mexico (2010 census). In the left column, dark grid cells indicate the cell contains at least one "urban" locality as of 2010 using either a 5,000 (top) or 50,000 (bottom) population threshold. This latter figure is close to the the time-varying urban definition for 2010 (41,278). Just over 10% of the grid cells contain a locality with a population of at least 5,000 people. These areas cluster in the center of the country, along the coasts, and following the highways leading to border cities. There are over 200 localities of 50,000 people or more in Mexico, but they are clustered in fewer grid cells. While spread throughout the country, there are several notable clusters of large cities around the largest metropolitan areas of the country: Mexico City, Guadalajara, and Monterrey.

Mexico's population grew by a factor of eight between 1900 and 2010. However, population growth alone cannot explain the geography of urbanization. The right column of Figure 3 shows what Mexico's urban system would look like today had all localities in the country grown at the same rate between 1900 and 2010. To create this counterfactual population, we use each locality's 1900 population as the baseline and multiply this by Mexico's 2010 population (112,336,538) over its 1900 population (13,607,259). Under the counterfactual, there would be many more—nearly twice as many—grid cells containing localities of 5,000 people or more as actually exist. The distinction is especially evident in the center of the country, where many small towns failed to grow proportional to population. By contrast, there would be somewhat fewer grid cells today containing large, 50,000-person cities. The actual spatial distribution of large cities also differs from the counterfactual, especially along the U.S. border and around major metropolitan areas.

Table 2 examines the geographic-, policy-, and history-related correlates of 2010 city locations. Reported are the results using the time-varying definition of city location of 0.037% of the national population, or just over 41,000 people as of 2010. Appendix Table C5 repeats the analysis using the flat 5,000-person urbanization cutoff. The results reveal interesting continuities and changes in urban development over the 20th century. Maize productivity and elevation remain important predictors of city location today. Distance to

Figure 3: Mexico's urban distribution: actual and counterfactual



(a) Actual: $> 5{,}000$ (b) Counterfactual: $> 5{,}000$

(c) Actual: $> 50{,}000$ (d) Counterfactual: $> 50{,}000$

Units on map are 15 x 15 km grid cells. Cells are black when population in a grid cell locality exceeds the relevant threshold. Left side panels are the actual city distribution in 2010. Right side panels present counterfactuals under the assumption that 1900 localities all grew at the same rate.

anchorage areas and to 1878 cities remain inversely related to urbanization in 2010, indicating the importance of internal trade access. Moreover, population density in 1570 is strongly correlated with the probability of a grid cell being urban, conditional on other geographic covariates. The magnitude and significance of all coefficients in the regression remain largely the same even in the presence of spatial and temporal lags of city location.

Though the results highlight persistence in urbanization patterns, some shifts in population concentration are evident. Perhaps most striking is the inverse relationship between urbanization and the distance to the U.S. border, which was not present in 1900. Distance to surface water and terrain ruggedness (measured in standard deviation of elevation) are inversely related to urbanization in 2010, where they had no clear relationship to urbanization in 1900. Cross-equation tests of the coefficients of this estimation with that of 1900 reject equivalence (Chi-2 value = 76.42, $p < 0.000$), indicating that the correlates of urbanization are different today than at the turn of the century.

To further explore changes in the correlates of urbanization, in Figures 4 and 5 we map the predicted probabilities of grid cells being urbanized using the locational fundamentals (elevation, maize productivity, distance to coast, and distance to fresh water) and trade-related (distance to border, anchorage area, and 1878 city) covariates respectively.[13] In the left side of each figure, we show the predicted probability of finding a city over the urbanization threshold in a grid cell in 1900 using either the fundamentals (Figure 4) or the trade-related (Figure 5) covariates. The right side presents the percent change in the predicted probability of urbanization in 2010 relative to 1900 using the time-varying city definition and the same variables as above.

Looking first at the maps using locational fundamentals (Figure 4), we can see that geographic predictors of urbanization in 1900 favored fertile highland areas in the center of the country where the risk of malaria and tropical diseases were smaller. Some important

[13]The partial effects come from cross-sectional probit regressions on all of our covariates, where we extract the probability of city presence determined by these key groups of variables, setting all other variables in the regression to their mean values. Map shading is for quintiles within each map.

Table 2: Correlates of urban grid cells in 2010

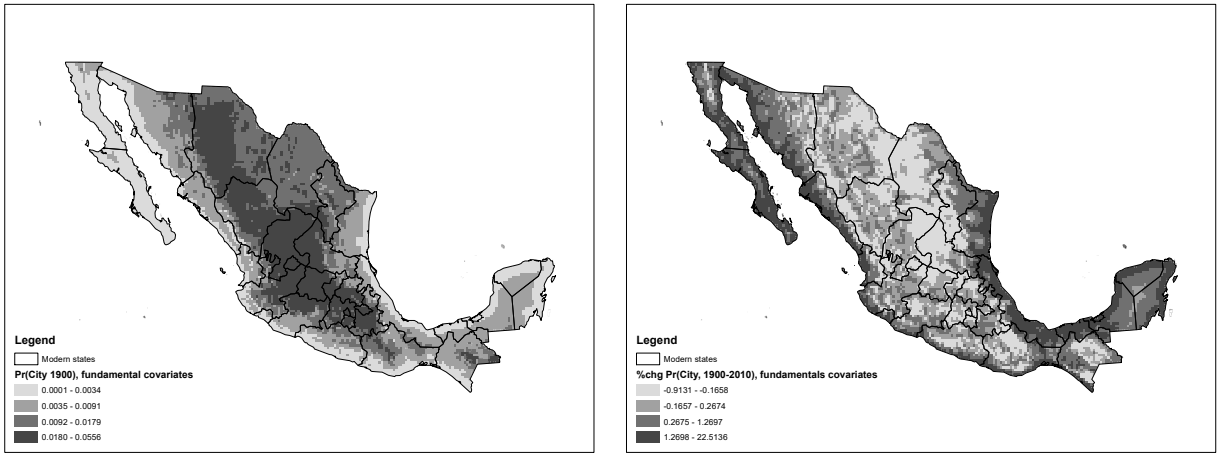
	Dep. var: Population > 45,000 = 1					
	(1)	(2)	(3)	(4)	(5)	(6)
Geographic fundamentals						
Elevation < 1500 m	-0.014*	-0.013*	-0.011	-0.011*	-0.011*	-0.012*
	(0.008)	(0.008)	(0.007)	(0.007)	(0.006)	(0.006)
Std grid cell elevation	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ln(km coast)	0.005*	0.009**	0.005*	0.006*	0.005*	0.006*
	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Ln(km water)	-0.003*	-0.003*	-0.003*	-0.003*	-0.003*	-0.003*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Ln(maize productivity)	0.017***	0.010*	0.012***	0.012***	0.010**	0.011**
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Trade covariates						
Ln(km to 1878 city, Mexico)	-0.008***	-0.006***	-0.008***	-0.008***	-0.007***	-0.007***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Ln(km to border)	-0.010***	-0.002	-0.013***	-0.011***	-0.013***	-0.011***
	(0.003)	(0.005)	(0.003)	(0.003)	(0.003)	(0.002)
Ln(km anchorage)	-0.016***	-0.019***	-0.016***	-0.017***	-0.015***	-0.016***
	(0.002)	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)
Historical population						
Ln(Tributary density, 1570)		0.009***				
		(0.003)				
Ln(1570 trib density, imputed)			0.008**		0.007**	
			(0.003)		(0.003)	
Ln(1570 trib density, zeros for missing)				0.008***		0.008***
				(0.003)		(0.003)
= 1 if missing 1570 data				-0.006		-0.006
				(0.004)		(0.004)
City of 15000, t-1, 50 km					0.014***	0.014***
					(0.005)	(0.005)
Observations	9342	6491	9342	9342	9342	9342
Adjusted R^2	0.16	0.17	0.17	0.18	0.18	0.18

Unit of observation is the 15 x 15 km grid cell. Clustered standard errors are in parentheses.

Coefficients robust to excluding Mexico City area. Reported coefficients are marginal effects from a probit.

* p< 0.10, ** p<0.05, *** p < 0.01.

Figure 4: Predicted cities in 1900 versus 2010 using fundamentals covariates



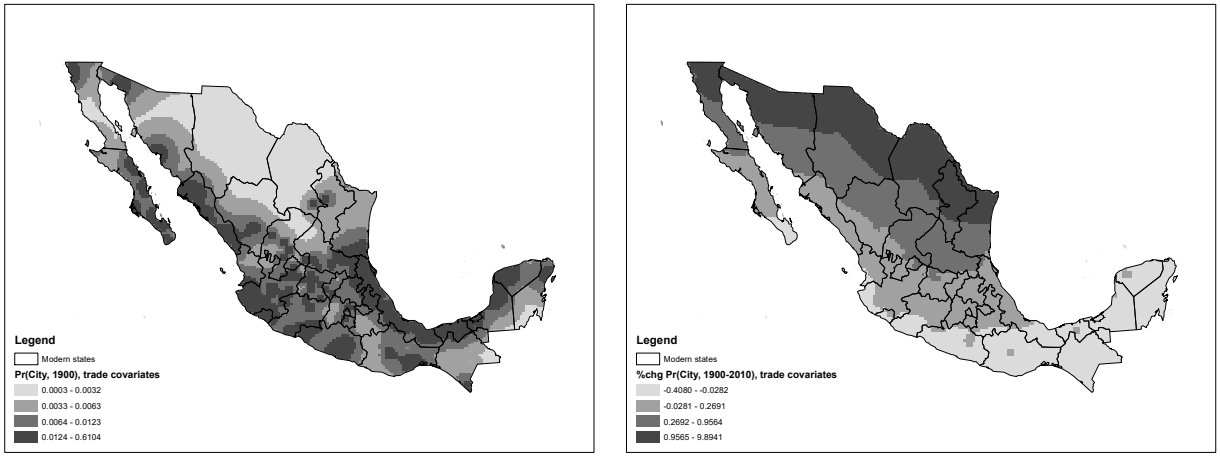
(a) 1900 (b) 1900 v 2010

Map (a) presents the predicted probability of a city in 1900, using only variation from elevation, maize productivity, distance to coast, and distance to fresh water. All other variables are set to their mean values. Map (b) presents percent change of these same predictions in 2010 relative to 1900. Predictions come from the specifications in column (2) of Tables 1 and 2.

changes are evident when comparing 2010 and 1900 predicted probabilities. First, it bears mentioning that the change in the predicted probability of urbanization from locational fundamentals in 2010 is actually negative for over half the grid cells (i.e., less urbanization is predicted in 2010 than in 1900). The highest changes in predicted probabilities of urbanization are found in lowland areas along the coasts and in the Yucatan Peninsula. This likely captures the declining importance of disease environment on urban development and potentially an increase in the importance of coastal tourism (Section 4).

The maps examining the influence of trade variables also demonstrate a substantial change in predicted probabilities at the beginning and end of the period. In 1900, distance to 1878 cities and to anchorage areas were important predictors of urbanization, as reflected in the figure. The Veracruz-Mexico City-Acapulco corridor across the center of the country, a critical nexus of trade since the early colonial era (Unikel 1976, Hassig 1985), has an especially high predicted level of urbanization in 1900 based on trade-related variables. The biggest jump in predicted urbanization between 1900 and 2010 occurs along the

Figure 5: Predicted cities in 1900 versus 2010 using trade covariates



(a) 1900 (b) 1900 v 2010

Map (a) presents the predicted probability of a city in 1900, using only variation from distance to border, anchorage area and Mexico City. All other variables are set to their mean values. Map (b) presents the percent change of these same predictions in 2010 relative to 1900. Predictions come from the specifications in column (2) of Tables 1 and 2.

northern border, consistent with a shift toward economic integration in the United States.

The explanatory power of history, trade, and geography on urbanization also varies over time. Removing the geographic variables from the 1900 regression lowers the pseudo-R-squared by 3.3 percent relative to the unrestricted model, while removing these same variables from the 2010 regressions lowers the pseudo-likelihood by 3.5 percent. For the trade variables, we see much larger shifts. The change in the pseudo-likelihood between the unrestricted regression and one that limits the estimation to non-trade variables lowers the pseudo-likelihood by 8.7 percent in 1900, and by 10.4 percent in 2010. Trade variables have a greater increase in predictive power in 2010 than in 1900, relative to locational fundamentals. The predictive power of colonial population, like that of geography, changes relatively little across the two time periods: in 1900, removing colonial population density from the estimation reduces the pseudo-likelihood by 1.4 percent, while removing it in 2010 reduces it by 1.9 percent.

Mexico's pattern of population concentration remained largely persistent from the early

colonial era until the 20th century. However, the results of this section suggest that technological change and/or policy choice may have played a role in changing long-standing patterns during the country's urban transformation. We assess this possibility in more detail in the next section.

4 Continuities and changes in urban development

This section examines the timing of observed changes in the correlates of urbanization more closely, using a panel approach to identify important shifts in urban development alongside political and economic factors.

4.1 Panel estimation strategy

We construct our cell-level panel dataset using the AHL data for each decennial census between 1900 and 2010.[14] Our outcome variable is the indicator for whether grid cell g in state s and in decade t contains a locality with population above the time-varying threshold. Using this measure of urbanization, numerous cells fall into and out of the urban category over time, so identification comes off of both "positive" and "negative" urbanization. We replicate the analysis in the appendix using both our 5,000 person time-invariant threshold and cell-level population density as alternative measures of population concentration.

This approach is designed to trace shifts in the relationships between the (time-invariant) geographic and historical covariates and the emergence of cities. To accomplish this, we estimate a series of flexible, fixed-effects models interacting these covariates with indicators for different time periods. For ease of interpretation, we divide the 20th century into three eras corresponding with major shifts in political and economic policy (e.g., Unikel 1976, Haber et al. 2008). These are the period of revolution and restructuring (1910–1940), the

[14] Censuses were conducted every 10 years, except for the post-Revolution census, which was delayed until 1921. Our results are robust to excluding the 1921 and 1980 censuses, which have well documented implementation and analysis issues.

period of industrialization and import-substitution (1940–1970), and the era of economic and political liberalization (1970–2010).[15] The omitted category, 1900, can be thought of as the baseline. Because we include unit and year fixed effects, the parameter estimates on the covariate-by-era interactions can be thought of as the conditional effect of that covariate during the specified era, relative to that observed in 1900. The estimating equation is:

$$U_{gst} = \alpha_0 + \sum_m \sum_e \beta'_{me} \mathbf{X}_{mg} T_e + \gamma_g + \gamma_t + \epsilon_{gst} \tag{2}$$

where U_{gst} is the specified urbanization indicator for grid cell g in state s in year t and era e. The matrix of covariates, $\mathbf{X}_{mg}$, includes the same covariates described above: the natural log of distance to the US border, the nearest 1878 city, the nearest water source, the coast, and the nearest anchorage point. It also includes tributary density in 1570, an indicator for elevation <1,500 meters, the standard deviation of elevation, and maize productivity. The indicator T_e takes the value one during era e and 0 otherwise. We include grid cell g and decade t fixed effects. Standard errors are clustered at the level of the state s. We also include, in some specifications, the lagged value of the presence of another city exceeding 15,000 inhabitants within a 50-km radius of the grid cell. The parameters of interest are the β_{me}: the marginal effect of covariate m in era e relative to that observed during the 1900 baseline.

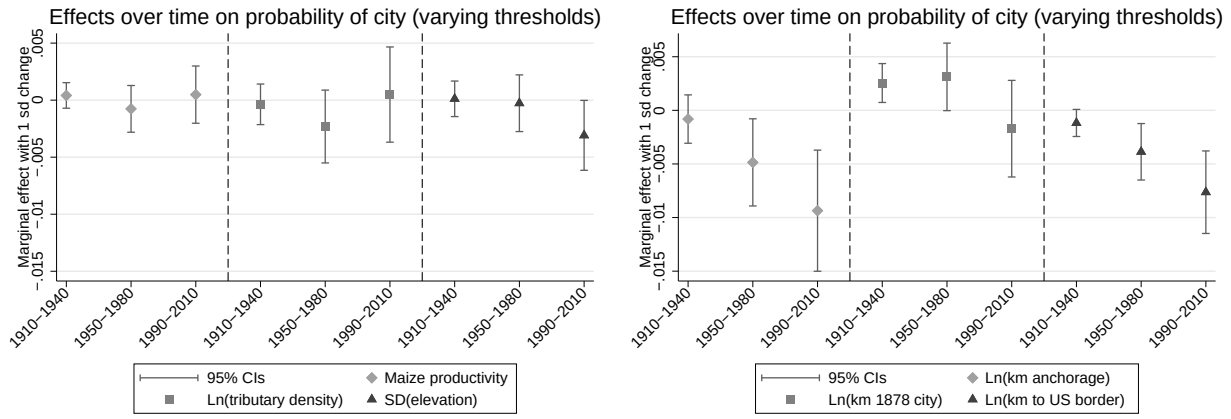
4.2 Results

Figure 6 plots the marginal effects of key explanatory factors on urbanization in each of the different eras outlined above using the fully saturated models with imputed population density for 1570 (column 6 of Appendix Table C6). Point estimates for the same specification using the 5,000-person threshold are available in Appendix Table C7. Effects are calculated for a one standard-deviation change in the covariate. As discussed in Section 3.1,

[15] In an appendix, we replicate the analysis on a decade-by-decade basis using each census year separately (Appendix Table C8).

the significant determinants of city location in 1900 were elevation, distance to anchorage area, distance to 1878 cities, maize productivity, and tributary density.

Figure 6: Evolving marginal effects on city emergence



(a) Geographic variables

(b) Trade variables

Figures show the marginal effects on outcomes of a one standard deviation change in covariates. Estimates come from Appendix Table C6, column (6). These are partial results using the varying-threshold definition of urban.

The left panel shows the marginal effects of two geographic variables – maize productivity and the standard deviation of elevation– and of colonial population density. In almost all cases, these marginal effects are not statistically distinguishable from zero, indicating that their marginal impacts across the eras differs little from their effect in 1900. The exception is a small negative and statistically significant impact of the standard deviation of elevation during the period 1980-2010, indicating less city emergence in areas with high variation in elevation in the modern period relative to 1900. The lack of significance of most of the estimates does not indicate that these variables are not deterministic of city emergence in different eras, but rather that their effects remain relatively constant over time.

By contrast, more change is evident in trade access proxies in the right panel. Distance to U.S. border and distance to natural anchorage area both become increasingly important in the emergence of cities over time. The impact of domestic trade access (measured as distance to 1878 cities), which is quite large in 1900, is attenuated in the periods from 1910 to 1970,

and not statistically different from 1900 in the modern era. These results are consistent with an increase in the relative importance of external trade. Interestingly, while cities were more likely to emerge near the border over time, overall population densities increased more in areas farther from the border (Appendix E). This pattern reflects dense non-urban settlement in the center and south of the country,[16] while population in the northern region remains concentrated in a small number of large metropolitan areas.

How do these shifts coincide with political and economic changes over the century? Our analysis illustrates that the determinants of urbanization from 1910–1940 are nearly equivalent to those of 1900. This is consistent with the economic and political turmoil during this period, which provided a brake on urbanization. After several decades of growth, the population of Mexico declined by about 5% between 1910 and 1921 due to revolutionary violence, disease, and emigration. Economic growth and trade faltered dramatically as major transportation networks—particularly railroads—were destroyed in the fighting (Coatsworth 1978). The 1929 economic crisis provided another shock, halting exports and precipitating the return of hundreds of thousands of migrants from the United States (González Navarro 1974, Knight 1991, Haber et al. 2003, Ficker 2004). Land redistribution under the Cárdenas presidency (1934–1940) drew labor back to the countryside, reducing urban development. This was the last decade in which rural areas accounted for the bulk of the country's population growth (e.g., Unikel 1976, Warman 2001).

Changes emerge in the 1940 to 1970 period, especially for proxies of trade access. This coincides with the period of rapid industrialization, sometimes called the "Mexican Miracle." The urban population increased by a factor of five, driven by both natural growth and rural-urban migration. Following the boost in industrial demand during World War II, Presidents Ávila Camacho and Alemán (1940–46 and 1946–52) introduced a series of tariff, regulatory, and exchange-rate policies designed to spur the growth of domestic industry. Road construction and paving increased rapidly to support industrial growth, improving

[16]Some relatively high-density grid cells occur in agricultural areas in the center and south with only rural localities.

transport linkages in the center of the country and making northern cities more accessible and attractive for settlement (Reynolds 1970, Unikel 1976). Import substitution is often thought to cause centripetal population movement toward capitals and other areas of large population density internally (e..g, Krugman & Elizondo 1996). However, Mexico's Border Industrialization Program and other policies also encouraged development along the northern border in cities like Tijuana, Ciudad Juárez, and Matamoros, each of which grew by a factor of more than eight between 1940 and 1970. Though the development of the northern border pulled some population out of the historical center of population in Mexico, evidence also suggests that urban primacy increased during ISI with Mexico City growing much faster than the country as a whole as might be expected given Mexico's inward-focused development strategy during this time (see Appendix F). The panel estimates reveal the evolving influence of both distance to the US border and distance to anchorage areas during this period of economic expansion and restructuring.

Mexico's centralized political and economic model began to falter in the 1970s and 1980s as a result of falling oil prices and many other factors (Haber et al. 2008). Mexico's debt crisis in 1982 ushered in nearly two decades of economic crises and stagnant growth. Following the crisis, Mexico became an experiment in trade liberalization, deregulation, and privatization. Subsidies and protections for domestic industry were lifted and the government took steps to encourage export-oriented growth (Moreno-Brid et al. 2004, 2005, Haber et al. 2008). During the 1980s, the government signed a series of bilateral trade agreements and became a signatory to the General Agreement on Tariffs and Trade. The implementation of the North American Free Trade Agreement in 1994 was the culmination of this trend. While economic liberalization arguably had a disappointing effect on growth, it succeeded in altering the structure of the economy. Exports increased from 8 percent of GDP in 1970 to almost 30 percent currently. This precipitated changes in Mexico's economic geography. Industrial production increasingly concentrated in the north and north-center regions nearer to the the U.S. border and highway system, as opposed to around Mexico City and domestic consumers

centers as had been the case under ISI (Unikel 1976, Vilalta 2010). Consistent with these trends, the marginal effects of our key international trade variables – distance to US border and distance to anchorage area – show the largest increase in the 1990–2010 period, relative to 1900 levels.

The results of this section illustrate that economic and policy change shifted Mexico's urban geography over the second half of the 20th century. Access to trade became an increasingly important determinant of urban development as the century progressed. Though the central altiplano remains economically and demographically important, economic integration with the U.S. pushed urban development toward the previously geographically disadvantaged border areas. Similar shifts are seen near anchorage areas and along the coast more generally. These trends represent a break from the striking persistence of the 1570–1900 period.

5 Conclusion

Different theories yield different predictions about the role of policy in shaping the urban landscape. If some areas are inherently geographically advantaged over others, as is argued by Sachs (2001) and Davis & Weinstein (2002), there is perhaps little that can be done to alter spatial inequality. If geographic advantage is highly contingent on economic or political factors (e.g., Bosker & Buringh 2015, Henderson et al. 2018), policy choices can alter urbanization patterns in important ways. Finally, if historical persistence is driven by spatial coordination and expectations (e.g., Krugman 1991, Bleakley & Lin 2012, Jedwab et al. 2015), policy and technology may also shift the spatial equilibrium.

In this paper, we examine persistence and change in population concentration in Mexico from 1570 to the present. We demonstrate that population concentration was strikingly persistent for centuries, even in the face of massive shocks like the catastrophic population collapse during the colonial period, dramatic economic restructuring, and numerous civil wars and conflicts. Population density in 1570 was the most substantial determinant of

concentration at the end of the colonial period, and it remained an important correlate of urban growth through the 20th century. The correlates of concentration in the colonial period— agricultural productivity and favorable disease environment —continued to predict city emergence for centuries. At the start of Mexico's century of urban transformation in 1900, the determinants of urbanization were thus similar to those of the early colonial period and strongly correlated with historic population densities.

Historical precedent continued to sow the seeds of urbanization as the 20th century progressed. However, the importance of geography for city emergence shifted over time in response to governmental decisions and changing technology. The role of trade access increased, while disease environment became less important. Geographic variables (elevation, maize productivity, distance to coast, and distance to fresh water) favored the center of the country in 1900, but by 2010 these fundamentals were increasingly driving city emergence along the coasts. Variables measuring trade access (distance to the U.S. border, Mexico City, and anchorage areas) predict cities near anchorage areas and through the traditional trading nexus of Mexico City and Acapulco and Veracruz in 1900, while the 2010 results highlight the importance of economic integration with the U.S. in the north. We also see changes in the relative importance of geography and trade for explaining urban development. Strictly geophysical variables have constant predictive power as the century progresses, while international trade access becomes increasingly important. This illustrates how the value of particular fixed geographic features may change, shifting population to previously undesirable areas.

Our panel analysis highlights the timing of these shifts. The estimations show that cities emerging in the revolutionary and post-revolutionary period (1910–1940) did so largely in similar places to where they had existed in 1900. Starting in 1950, the correlates of urbanization begin to change due to industrialization and economic policy. The role that increasing economic integration with the United States in particular played in remaking Mexico's city system is similar to what has been found in work linking trade patterns and

urbanization in Europe (e.g., Redding & Sturm 2008, Redding et al. 2011, Bosker et al. 2013). Together these findings highlight the role of economic policy in shifting urbanization patterns.

Mexico's urban evolution illustrates the shifting influence of history, geographic fundamentals, and trade. By examining persistence and change in urbanization over 450 years, this paper illustrates both the strong pull of historical precedent and the ways in which policy choices can remake the nature of geographic advantage. The marked persistence of population concentration over several centuries—in the face of transformative economic, political, and demographic changes—suggests that history or perhaps inherent geographic advantage may always play important role in urban development. However, the emergence of cities along the northern border and in the arid regions of the northeast, far from Mexico City and outside of the fertile central plateau, also shows that history and geography are not always determinative. Though the collapse of the indigenous population and the overhaul of political institutions in the 19th and 20th centuries were not sufficient to disrupt the centripetal pull of Tenochtitlan/Mexico City, improvements in transport technology combined with the opening up of trade with the north did reshape patterns of settlement and investment in important ways. Trade and other economic policies shifted the returns to different locational features and the expectations of future development, and in doing so altered the trajectory of urban development.

References

Ades, A. F. & Glaeser, E. L. (1995), 'Trade and Circuses: Explaining Urban Giants', *Quarterly Journal of Economics* **110**(1), 195–227.

Anas, A., Arnott, R. & Small, K. A. (1998), 'Urban spatial structure', *Journal of economic literature* **36**(3), 1426–1464.

Arroyo Abad, L. & van Zanden, J. L. (2016), 'Growth under extractive institutions? latin american per capita gdp in colonial times', *The Journal of Economic History* **76**(4), 1182–1215.

Bairoch, P. (1988), *Cities and Economic Development. From the Dawn of History to the Present*, University of Chicago Press, Chicago, IL.

Bank, W. (1970), 'Mexico - fourth road project (english)', *Transportation projects series; no. PTR 47. Washington, D.C.* .

Beatty, E. (2001), *Institutions and Investment: the political basis of industrialization in Mexico before 1911*, Vol. 42, Stanford University Press Stanford, CA.

Behrens, K. (2007), 'On the location and lock-in of cities: Geography vs transportation technology', *Regional Science and Urban Economics* **37**(1), 22–45.

Black, D. & Henderson, V. (1999), 'A theory of urban growth', *Journal of Political Economy* **3**, 252–284.

Bleakley, H. & Lin, J. (2012), 'Portage and Path Dependence', *Quarterly Journal of Economics* **127**(2), 587–644.

Bosker, M. & Buringh, E. (2015), 'City seeds: geography and the origins of the European city system', *Journal of Urban Economics* .

Bosker, M., Buringh, E. & van Zanden, J. L. (2013), 'From Baghdad to London: Unraveling Urban Development in Europe, the Middle East, and North Africa, 800–1800', *Review of Economics and Statistics* **95**(4), 1418–1437.

Buringh, E. (2018), Urbanisation hub: The clio-infra database on urban settlement sizes: 1500–2000.
URL: *http://www.cgeh.nl/urbanisation-hub-clio-infra-database-urban-settlement-sizes-1500-2000*

Coatsworth, J. H. (1978), 'Obstacles to economic growth in nineteenth-century Mexico', *The American Historical Review* **83**(1), 80–100.

Coatsworth, J. H. & Williamson, J. G. (2004), 'Always protectionist? Latin American tariffs from independence to Great Depression', *Journal of Latin American Studies* **36**(02), 205–232.

Cook, S. F. & Borah, W. W. (1960), *The Indian population of central Mexico, 1531-1610*, Vol. 44, University of California Press.

Cook, S. F. & Borah, W. W. (1971), *Essays in population history: Mexico and the Caribbean*, Vol. 1, Univ of California Press.

Davis, D. R. & Weinstein, D. E. (2002), 'Bones, Bombs, and Break Points: The Geography of Economic Activity', *American Economic Review* **92**, 1269–1289.

De Vries, J. (1984), *European Urbanization, 1500-1800*, Vol. 4, Routledge.

Duranton, G. (1999), 'Distance, land, and proximity: economic analysis and the evolution of cities', *Environment and Planning A* **31**(12), 2169–2188.

Faber, B. & Gaubert, C. (2019), 'Tourism and economic development: Evidence from mexico's coastline', *American Economic Review* **109**(6), 2245–93.

Ficker, S. K. (2004), 'The export boom of the Mexican revolution: Characteristics and contributing factors', *Journal of Latin American Studies* **36**(02), 267–296.

Fujita, M. & Mori, T. (1996), 'The role of ports in the making of major cities: self-agglomeration and hub-effect', *Journal of Development Economics* **49**(1), 93–120.

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

Gerhard, P. (1993*b*), *The North Frontier of New Spain*, University of Oklahoma Press, Norman, OK.

Gerhard, P. (1993*c*), *The Southeast Frontier of New Spain*, University of Oklahoma Press, Norman, OK.

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

Glaeser, E. L. (2007), The economics approach to cities. NBER Working Paper 13696.

Glaeser, E. L. (2014), 'A World of Cities: The Causes and Consequences of Urbanization in Poorer Countries', *Journal of the European Economic Association* **12**(5), 1154–1199.

González Navarro, M. (1974), *Población y sociedad en México (1900–1970)*, Universidad Nacional Autónoma de México, México, D.F.

Haber, S. H. (1992), 'Assessing the Obstacles to Industrialisation: The Mexican Economy, 1830–1940', *Journal of Latin American Studies* **24**(01), 1–32.

Haber, S., Klein, H. S., Maurer, N. & Middlebrook, K. J. (2008), *Mexico Since 1980*, Cambridge Univrsity Press, New York.

Haber, S., Razo, A. & Maurer, N. (2003), *The Politics of Property Rights: Political Instability, Credible Commitments, and Economic Growth in Mexico, 1876–1929*, Cambridge University Press, New York, NY.

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

Henderson, J. V. (2005), 'Urbanization and growth', *Handbook of economic growth* **1**, 1543–1591.

Henderson, J. V., Squires, T. L., Storeygard, A. & Weil, D. N. (2018), 'The Global Spatial Distribution of Economic Activity: Nature, History, and the Role of Trade', **33**, 357–406. The Quarterly Journal of Economics.

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

Jedwab, R., Christiaensen, L., Gindelsky, M. et al. (2014), 'Rural Push, Urban Pull and... Urban Push? New Historical Evidence from Developing Countries', *The George Washington University, Institute for International Economic Policy Working Papers* **4**.

Jedwab, R., Kerby, E. & Moradi, A. (2015), 'History, path dependence and development: Evidence from colonial railroads, settlers and cities in Kenya', *The Economic Journal* .

Knight, A. (1991), 'Popular Culture and the Revolutionary State in Mexico', *The Hispanic American Historical Review* **74**(3), 393–444.

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

Konishi, H. (2000), 'Formation of hub cities: transportation cost advantage and population agglomeration', *Journal of Urban Economics* **48**(1), 1–28.

Krugman, P. (1991), 'History Verus Expectations', *Quarterly Journal of Economics* **106**(2), 651–667.

Krugman, P. (1993), 'On the number and location of cities', *European Economic Review* **37**(2-3), 293–298.

Krugman, P. & Elizondo, R. L. (1996), 'Trade policy and the third world metropolis', *Journal of Development Economics* **49**(1), 137–150.

Maloney, W. F. & Valencia Caicedo, F. (2016), 'The persistence of (subnational) fortune', *The Economic Journal* **126**(598), 2363–2401.

Mellinger, A., Sachs, J. D. & Gallup, J. L. (2000), Climate, coastal proximity, and development, *in* 'The Oxford Handbook of Economic Geography,', Oxford University Press, New York, NY.

Moreno-Brid, J. C., Nápoles, P. R. & Valdivia, J. C. (2004), 'NAFTA and the Mexican Economy: A Look Back on a Ten-Year Relationship', *NCJ Int'l L. & Com. Reg.* **30**, 997.

Moreno-Brid, J. C., Santamaria, J. & Rivas Valdivia, J. C. (2005), 'Industrialization and economic growth in Mexico after NAFTA: the road travelled', *Development and Change* **36**(6), 1095–1119.

NOAA (2014), 'Anchorage areas'. NOAA Office for Coastal Management, US Government Publishing Office.

Nunn, N. & Puga, D. (2012), 'Ruggedness: The Blessing of Bad Geography in Africa', *The Review of Economics and Statistics* **94**(1), 20–36.

Overman, H. G. & Venables, A. J. (2005), *Cities in the developing world*, Centre for Economic Performance, London School of Economics and Political Science.

Pirenne, H. (1925), *Medieval cities*, Doubleday.

Redding, S. J. & Sturm, D. M. (2008), 'The costs of remoteness: Evidence from German division and reunification', *The American Economic Review* **98**(5), 1766–1797.

Redding, S. J., Sturm, D. M. & Wolf, N. (2011), 'History and industry location: Evidence from German airports', *Review of Economics and Statistics* **93**(3), 814–831.

Reynolds, C. W. (1970), *The Mexican economy: twentieth-century structure and growth*, Yale University Press.

Sachs, J. D. (2001), Tropical underdevelopment. NBER Working Paper 8119.

Sellars, E. A. & Alix-Garcia, J. (2018), 'Labor Scarcity, Land Tenure, and Historical Legacy: Evidence from Mexico', *Journal of Development Economics* **135**, 504–516.

Unikel, L. (1976), *El desarrollo urbano de México: diagnóstico e implicaciones futuras*, El Colegio de México, México, DF.

United States Central Intelligence Agency (1990), *The World Factbook (1990)/Mexico*, CIA.

Vilalta, C. (2010), Evolución de las desigualdades regionales, 1960–2020, *in* G. Garza & M. Schteingart, eds, 'Desarrollo urbano y regional, tomo II', El Colegio de México, Mexico, DF.

Warman, A. (2001), *El campo mexicano en el siglo XX*, Fondo de Cultura Económica.

Weber, M., Martindale, D. & Neuwirth, G. (1958), *The City*, Free Press New York.

World Bank (2016), 'World development indicators'.

Appendices for online publication

Appendix A Data Appendix

INEGI's Archivo histórico de localidades (AHL) had to be transformed in several ways for analysis in this paper. The AHL omits some small localities recorded in earlier censuses that had ceased to exist prior to the development of the geographical code system in 1978. In some states (such as Guerrero, México, Tamaulipas, and Yucatán), all localities that shift municipal boundaries prior to 1978 are omitted in the early years of the dataset, though nearly all of the omitted communities are rural. The central *delegaciones* of Mexico City are also omitted from the AHL prior to the administrative reform of the Distrito Federal in 1970. In addition, latitude and longitude values are missing from many communities that ceased to exist or changed census codes prior to 2000, making it difficult to pinpoint their location.

We edit and supplement the data in several ways. First, we replace missing population data from Mexico City using information from Unikel (1976), which contains lists of urban centers over time and complete population counts for central Mexico City (now the four central delegations) up to 1970. Data on city placement in 1878 are also taken from Unikel. We also use information from the *División Territorial*, which contains a list of localities and their population as of each census, to identify other urban localities (those over 5,000 people) that are omitted from the AHL. Because omitted communities are predominantly rural, few observations had to be manually added.

Where geographic data are missing in the AHL (about 15% of localities in the dataset), we use information on the history of a locality to tie it to the location of a contemporary settlement. We do this primarily via text analysis using the AHL's "Movimientos" supplementary text file, which contains information on the status of each locality in each census

year. Using regular expressions, we identify situations where a locality either merged with another locality, split from a parent locality, or transferred to a different municipality (which would entail a change in identifier code). We then link the alternative geographic identifiers that have been connected to a single locality together, which can then be linked to a contemporary location. For the small number of historical localities for which we were unable to identify the geographic location, we place the locality at the position of the *cabecera municipal*, the seat of local government. These localities are almost always small (usually less than 100 in population) rural areas that had been depopulated prior to 1990, which are not used in the baseline empirical analysis in the paper.

Appendix B Colonial population concentration

In this appendix, we examine the determinants of population concentration during the early colonial period. Table B1 presents the correlates of grid-cell population density during the colonial era. Reported are OLS estimates using the log of population density in 1570 (Column (1)), 1650 (Columns (2-3)), and 1800 (Columns (4-6)) as the dependent variables and using the geographic covariates described above. Population density in 1570 is most strongly predicted by maize productivity and elevation, with greater density concentrated in areas of high maize productivity and in higher elevations outside of malarial/disease zones. More rugged terrain is also positively related to population concentration in 1570, which is consistent with the argument that early settlements thrived in areas that were easier to defend militarily. However, terrain ruggedness is negatively related to late colonial population density. This could be due to a decline in the importance of defensive position (perhaps because of the consolidation of Spanish colonial rule) and an increase in the importance of cultivation and construction, which tend to be more difficult in rugged areas (e.g., Nunn & Puga 2012).

Notably, distance to ports and anchorage areas had no discernible relationship with population density in the early colonial period, but it is negatively related to density in 1800. This is in line with the increasing importance of maritime trade to other areas of the Empire (e.g., Knight 2002). Also notable is the increasing and positive relationship between distance to US border and population density in 1800. This is likely due to the growing concentration of population in the center/Mexico City region.

Table B1: Correlates of grid-cell population density, colonial period

	Ln(density, 1570)	Ln(density, 1650)		Ln(density 1800)		
	(1)	(2)	(3)	(4)	(5)	(6)
Elevation < 1500 m	-0.562***	-0.335	0.023	-0.090	-0.405	0.045
	(0.197)	(0.294)	(0.303)	(0.206)	(0.239)	(0.165)
Std grid cell elevation	0.002**	0.001	-0.000	-0.001**	-0.000	-0.001***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Ln(maize productivity)	0.603**	0.479	0.012	-0.147	-0.200	-0.288**
	(0.264)	(0.314)	(0.318)	(0.171)	(0.188)	(0.110)
Ln(km to border)	0.223	0.155	-0.013	0.419***	0.581**	0.315***
	(0.202)	(0.188)	(0.148)	(0.100)	(0.208)	(0.077)
Ln(km anchorage)	-0.173	-0.259	-0.112	-0.182**	-0.301***	-0.227**
	(0.180)	(0.189)	(0.127)	(0.077)	(0.105)	(0.088)
Ln(km coast)	0.023	0.046	0.022	0.039	0.156*	0.033
	(0.099)	(0.083)	(0.067)	(0.055)	(0.082)	(0.056)
Ln(km water)	-0.039	-0.035	0.006	0.033	0.024	0.052**
	(0.037)	(0.070)	(0.055)	(0.028)	(0.025)	(0.024)
Ln(Tributary density, 1570)			0.725***	0.815***		0.793***
			(0.126)	(0.085)		(0.087)
Ln(Tributary density, 1650)					0.528***	0.123
					(0.119)	(0.100)
Observations	6491	5842	5736	4495	4017	3951
Adjusted R^2	0.242	0.132	0.489	0.715	0.618	0.805

Unit of observation is the 15 x 15 km grid cell. Clustered standard errors are in parentheses.
* p< 0.10, ** p<0.05, *** p < 0.01.

Appendix C Additional tables

Table C2: Correlations between presence of professionals, population density, and cities, 1900

	Dependent variable: per capita counts professions in 1900						
	Pharmacists	Doctors	Administrators	Public Servants	Laborers	Teachers	Students
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A:							
IHS(municipal population density)	-0.109***	-0.540**	-0.554***	-3.907***	-614.434***	-3.741***	-0.861***
	(0.038)	(0.242)	(0.197)	(0.935)	(167.390)	(1.226)	(0.328)
Observations	1207	1207	1207	1207	1207	1207	1207
Panel B:							
Has city over threshold	-0.009	-0.137*	-0.494***	-1.472***	-317.061***	-1.252***	0.635
	(0.027)	(0.074)	(0.146)	(0.458)	(55.312)	(0.399)	(0.500)
Observations	1207	1207	1207	1207	1207	1207	1207
Panel C:							
IHS(municipal population density)	-0.140***	-0.678**	-0.631***	-4.820***	-742.446***	-4.644***	-1.238***
	(0.047)	(0.300)	(0.242)	(1.131)	(205.411)	(1.503)	(0.388)
Has city over threshold	0.224***	0.987**	0.552	6.510***	912.327***	6.439***	2.686***
	(0.066)	(0.441)	(0.361)	(1.630)	(298.011)	(2.202)	(0.710)
Observations	1207	1207	1207	1207	1207	1207	1207

Unit of analysis is 1900 municipality. All regression include ln(municipal area in 1900) as a control. Estimates are OLS.

Table C3: Correlations between presence of professionals, population density, and cities, 2010

	Dependent variable: per capita counts outcomes in 2010						
				Employed in			
	Classrooms	High school	Industry	Commerce	Construction	Finance	Services
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A:							
IHS(municipal population, 1,000s)	-4.703***	-0.663	-8.220	-10.879*	-16.030*	-0.531	-7.431
	(1.379)	(0.506)	(6.075)	(6.427)	(9.014)	(0.390)	(7.572)
Observations	1167	1573	1905	1955	534	1013	1941
Panel B:							
Has city over threshold	-10.212***	2.108	-2.523	-8.665	-11.803	1.330	13.288
	(2.450)	(1.510)	(18.948)	(16.795)	(7.943)	(1.230)	(20.786)
Observations	1167	1573	1905	1955	534	1013	1941
Panel C:							
IHS(municipal population, 1,000s)	-5.499***	-1.409**	-11.595	-14.442*	-28.002*	-1.314***	-12.756
	(1.917)	(0.659)	(7.472)	(7.606)	(15.395)	(0.436)	(9.008)
Has city over threshold	5.982	6.267***	32.292	35.238*	50.101*	4.895***	51.943**
	(4.703)	(2.164)	(24.155)	(18.929)	(27.641)	(1.234)	(24.019)
Observations	1167	1573	1905	1955	534	1013	1941

Unit of analysis is 2010 municipality. All regressions include ln(municipal area, 2010). Estimates are OLS. Employment outcomes from 2009.

Table C4: Summary statistics for grid cells with and without city > 15,000 inhabitants, 2010

	(1) No City	(2) City	(3) Norm diff
Km to 1878 city > 15,000, Mexico	231.088	115.734	-0.387
Km to border	53.401	67.864	0.309
Km to nearest anchorage	235.361	192.841	-0.188
Km to nearest river or lake	14.872	9.464	-0.191
Mean grid cell elevation	971.721	1129.229	0.129
Std grid cell elevation	127.210	109.046	-0.117
Mean grid cell maize	0.518	0.985	0.508
Tributary density, 1650	0.233	0.703	0.308
Tributary density, 1570	0.758	2.844	0.450
Population density, 1800	1.238	4.757	0.416
Has one or more localities over 5000 pop	0.067	1.000	3.717
Observations	8876	466	9342

Columns (1) and (2) contain mean values of variables for grid cells containing a city (2) or not (1) in 2010. Column (3) shows the normalized difference of these two means.

Table C5: Correlates of urban grid cells in 2010, urban cutoff of 5,000

	Dep var: City greater than 5,000 = 1					
	(1)	(2)	(3)	(4)	(5)	(6)
Geographic fundamentals						
Elevation < 1500 m	-0.031	-0.025	-0.017	-0.018	-0.022	-0.024
	(0.027)	(0.033)	(0.025)	(0.024)	(0.021)	(0.020)
Std grid cell elevation	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ln(km coast)	0.012**	0.018**	0.013*	0.014**	0.011	0.012*
	(0.006)	(0.009)	(0.007)	(0.007)	(0.006)	(0.006)
Ln(km water)	-0.009**	-0.008	-0.009**	-0.009**	-0.008**	-0.008**
	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Ln(maize productivity)	0.086***	0.068***	0.067***	0.069***	0.059***	0.060***
	(0.016)	(0.016)	(0.014)	(0.014)	(0.012)	(0.012)
Trade covariates						
Ln(km to 1878 city, Mexico)	-0.031***	-0.025***	-0.028***	-0.028***	-0.021***	-0.020***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Ln(km to border)	0.001	0.033	-0.008	-0.002	-0.008	-0.002
	(0.013)	(0.020)	(0.013)	(0.012)	(0.011)	(0.009)
Ln(km anchorage)	-0.045***	-0.050***	-0.044***	-0.046***	-0.036***	-0.038***
	(0.006)	(0.009)	(0.006)	(0.006)	(0.005)	(0.005)
Historical population						
Ln(Tributary density, 1570)		0.037***				
		(0.011)				
Ln(1570 trib density, imputed)			0.032***		0.029***	
			(0.010)		(0.009)	
Ln(1570 trib density, zeros for missing)				0.035***		0.031***
				(0.010)		(0.009)
= 1 if missing 1570 data				-0.034*		-0.028*
				(0.020)		(0.015)
City of 15000, t-1, 50 km					0.085***	0.083***
					(0.014)	(0.013)
Observations	9342	6491	9342	9342	9342	9342
Adjusted R^2						

Unit of observation is the 15 x 15 km grid cell. Clustered standard errors are in parentheses.
Coefficients robust to excluding Mexico City area. Reported coefficients are marginal effects from a probit.
* p< 0.10, ** p<0.05, *** p < 0.01.

Table C6: Correlates of cities: > 0.036% population cutoff

	(1)	(2)	(3)	(4)	(5)	(6)
City of 15000, t-1, 50 km			-0.002	-0.001	-0.001	-0.002
			(0.002)	(0.001)	(0.001)	(0.001)
Ln(km to border), 1910-1940	-0.001	-0.001*	-0.001	-0.001*	-0.001*	-0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(km to border), 1950-1970	-0.002	-0.004***	-0.002	-0.004***	-0.004***	-0.004***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Ln(km to border), 1980-2010	-0.003	-0.007***	-0.002	-0.007***	-0.007***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Ln(km to 1878 city), 1910-1940	0.001*	0.002***	0.002*	0.002***	0.002***	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(km to 1878 city), 1950-1970	0.002*	0.002*	0.002*	0.002*	0.002**	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(km to 1878 city), 1980-2010	0.001	-0.001	0.001	-0.001	-0.001	-0.001
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Ln(km to anchorage), 1910-1940	-0.001	-0.001	-0.001	-0.001	-0.001	-0.004**
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Ln(km to anchorage), 1950-1970	-0.005*	-0.005**	-0.006*	-0.005**	-0.005**	-0.008***
	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Ln(km to anchorage), 1980-2010	-0.012***	-0.010***	-0.012***	-0.010***	-0.010***	-0.014***
	(0.005)	(0.003)	(0.005)	(0.003)	(0.003)	(0.004)
Maize productivity, 1910-1940	0.001	0.001	0.001	0.001	0.001	0.003
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Maize productivity, 1950-1970	-0.003	-0.002	-0.002	-0.002	-0.002	0.000
	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Maize productivity, 1980-2010	-0.000	0.001	-0.000	0.001	0.001	0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Ln(km to coast), 1910-1940	0.000	0.000	0.000	0.000	0.000	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(km to coast), 1950-1970	0.001	0.001*	0.001	0.002*	0.001	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(km to coast), 1980-2010	0.002	0.002	0.002	0.002	0.002	0.003*
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Ln(km to water), 1910-1940	-0.000	-0.000	-0.000	-0.000	-0.000	-0.001*
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Ln(km to water), 1950-1970	-0.002**	-0.002***	-0.002**	-0.002***	-0.002**	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(km to water), 1980-2010	-0.002**	-0.002**	-0.002**	-0.002**	-0.002**	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
SD elevation, 1910-1940	0.000	0.000	0.000	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SD elevation, 1950-1970	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SD elevation, 1980-2010	-0.000***	-0.000**	-0.000***	-0.000**	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Ln(elevation), 1910-1940	-0.001	-0.000	-0.001	-0.000	-0.000	-0.001
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Ln(elevation), 1950-1970	-0.000	-0.001	-0.000	-0.001	-0.001*	-0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(elevation), 1980-2010	0.002	-0.001	0.001	-0.001	-0.001	-0.001
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Ln(1570 trib density), 1910-1940	-0.000		-0.000			
	(0.001)		(0.001)			
Ln(1570 trib density), 1950-1970	-0.002		-0.002			
	(0.001)		(0.001)			
Ln(1570 trib density), 1980-2010	0.001		0.001			
	(0.002)		(0.002)			
Ln(1570 trib density impute), 1910-1940		-0.000		-0.000		-0.001
		(0.001)		(0.001)		(0.001)
Ln(1570 trib density impute), 1950-1970		-0.002		-0.002		-0.003
		(0.001)		(0.001)		(0.002)
Ln(1570 trib density impute), 1980-2010		0.000		0.000		0.000
		(0.002)		(0.002)		(0.002)
Ln(1570 trib density w/zeros), 1910-1940					-0.000	
					(0.001)	
Ln(1570 trib density w/zeros), 1950-1970					-0.001	
					(0.001)	
Ln(1570 trib density w/zeros), 1980-2010					0.001	
					(0.002)	
Observations	77892	112032	77892	112032	112032	93360
Adjusted R^2	0.005	0.006	0.005	0.006	0.006	0.008

Unit of observation is the grid cell, 1900-2010. Robust standard errors clustered at state level are in parentheses. Outcome is a binary variable equal to one when the size of the city exceeds a varying threshold. Column (6) excludes 1921 and 1980 censuses. Marginal effects should be interpreted as differences from 1900, the excluded year.

Table C7: Correlates of cities: $>$ 5,000 population cutoff

	(1)	(2)	(3)	(4)	(5)	(6)
Ln(km to border), 1950-1970	-0.001	-0.003**	-0.001	-0.003**	-0.003**	-0.004**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Ln(km to border), 1980-2010	0.000	-0.004**	-0.000	-0.004*	-0.003	-0.005**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Ln(km to 1878 city), 1910-1940	-0.001	-0.001	-0.001*	-0.001**	-0.001**	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(km to 1878 city), 1950-1970	-0.006***	-0.005***	-0.006***	-0.005***	-0.005***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(km to anchorage), 1910-1940	-0.001	-0.001	-0.001	-0.001	-0.001	-0.002*
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Ln(km to anchorage), 1950-1970	-0.006**	-0.005***	-0.005*	-0.004***	-0.005***	-0.006***
	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Ln(km to anchorage), 1980-2010	-0.014***	-0.011***	-0.012***	-0.010***	-0.011***	-0.009***
	(0.004)	(0.002)	(0.004)	(0.002)	(0.002)	(0.002)
Maize productivity, 1910-1940	0.004**	0.003**	0.004**	0.003**	0.003**	0.005***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Maize productivity, 1950-1970	0.012***	0.013***	0.012***	0.012***	0.012***	0.015***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Maize productivity, 1980-2010	0.016***	0.015***	0.015***	0.015***	0.015***	0.016***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Ln(km to coast), 1910-1940	0.001	0.000	0.001	0.000	0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(km to coast), 1950-1970	0.003	0.002	0.002	0.002	0.001	0.002
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Ln(km to coast), 1980-2010	0.001	0.002	0.001	0.001	0.001	0.002
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Ln(km to water), 1910-1940	-0.000	-0.001	-0.000	-0.001	-0.001	-0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ln(km to water), 1950-1970	-0.002**	-0.002**	-0.002**	-0.001**	-0.001**	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(km to water), 1980-2010	-0.001	-0.001*	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
SD elevation, 1910-1940	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SD elevation, 1950-1970	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SD elevation, 1980-2010	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ln(elevation), 1910-1940	0.000	0.000	0.000	0.000	0.000	0.000
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Ln(elevation), 1950-1970	0.001	0.000	0.001	0.000	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(elevation), 1980-2010	0.003	0.001	0.003*	0.001	0.002*	-0.000
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)

Ln(1570 trib density), 1910-1940	0.001		0.000			
	(0.001)		(0.001)			
Ln(1570 trib density), 1950-1970	0.005***		0.005***			
	(0.002)		(0.002)			
Ln(1570 trib density), 1980-2010	0.010***		0.010***			
	(0.002)		(0.002)			
Ln(1570 trib density impute), 1910-1940		0.001		0.000		0.001
		(0.001)		(0.001)		(0.001)
Ln(1570 trib density impute), 1950-1970		0.005***		0.005***		0.005***
		(0.002)		(0.002)		(0.002)
Ln(1570 trib density impute), 1980-2010		0.010***		0.010***		0.010***
		(0.002)		(0.002)		(0.002)
Ln(1570 trib density w/zeros), 1910-1940					0.001	
					(0.001)	
Ln(1570 trib density w/zeros), 1950-1970					0.005***	
					(0.002)	
Ln(1570 trib density w/zeros), 1980-2010					0.010***	
					(0.002)	
Ln(km to border), 1910-1940	0.000	-0.000	0.000	-0.000	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ln(km to 1878 city), 1980-2010	-0.009***	-0.010***	-0.010***	-0.010***	-0.010***	-0.010***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
City of 15000, t-1, 50 km			0.009***	0.008***	0.008***	0.009***
			(0.002)	(0.002)	(0.002)	(0.002)
Observations	73650	106651	73650	106651	106651	88765
Adjusted R^2	0.031	0.028	0.032	0.029	0.029	0.026

Unit of observation is the grid cell, 1900-2010. Robust standard errors clustered at state level are in parentheses. Outcome is a binary variable equal to one when the size of the city exceeds 5,000. Column (6) excludes 1921 and 1980 censuses. Marginal effects should be interpreted as differences from 1900, the excluded year.

Table C8: Correlates of cities: decadal estimation

	(1)	(2)	(3)
Ln(km to US border) x 1910	-0.004***	-0.000	-0.000
	(0.001)	(0.001)	(0.001)
Ln(km to US border) x 1921	-0.005***	-0.001	0.000
	(0.001)	(0.001)	(.)
Ln(km to US border) x 1930	-0.006***	-0.002*	-0.002*
	(0.002)	(0.001)	(0.001)
Ln(km to US border) x 1940	-0.006***	-0.002	-0.002
	(0.002)	(0.001)	(0.001)
Ln(km to US border) x 1950	-0.006***	-0.003*	-0.003*
	(0.002)	(0.002)	(0.002)
Ln(km to US border) x 1960	-0.007***	-0.004**	-0.004**
	(0.002)	(0.002)	(0.002)
Ln(km to US border) x 1970	-0.008***	-0.005***	-0.005***
	(0.002)	(0.002)	(0.002)
Ln(km to US border) x 1980	-0.009***	-0.005***	0.000
	(0.002)	(0.002)	(.)
Ln(km to US border) x 1990	-0.011***	-0.007***	-0.007***
	(0.002)	(0.002)	(0.002)
Ln(km to US border) x 2000	-0.012***	-0.008***	-0.008***
	(0.003)	(0.002)	(0.002)
Ln(km to US border) x 2010	-0.013***	-0.009***	-0.009***
	(0.003)	(0.002)	(0.002)
Ln(km 1878 city) x 1910	-0.005***	0.002**	0.002**
	(0.001)	(0.001)	(0.001)
Ln(km 1878 city) x 1921	-0.005***	0.002	0.000
	(0.001)	(0.001)	(.)
Ln(km 1878 city) x 1930	-0.005***	0.002*	0.002*
	(0.001)	(0.001)	(0.001)
Ln(km 1878 city) x 1940	-0.004***	0.003**	0.003***
	(0.001)	(0.001)	(0.001)
Ln(km 1878 city) x 1950	-0.003***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)
Ln(km 1878 city) x 1960	-0.004***	0.003*	0.003*
	(0.001)	(0.001)	(0.001)
Ln(km 1878 city) x 1970	-0.005***	0.002	0.002
	(0.001)	(0.002)	(0.002)
Ln(km 1878 city) x 1980	-0.006***	0.001	0.000
	(0.001)	(0.002)	(.)
Ln(km 1878 city) x 1990	-0.007***	0.000	0.000
	(0.002)	(0.002)	(0.002)
Ln(km 1878 city) x 2000	-0.008***	-0.001	-0.001
	(0.002)	(0.002)	(0.002)
Ln(km 1878 city) x 2010	-0.009***	-0.002	-0.002
	(0.002)	(0.002)	(0.002)

Ln(km anchorage) x 1910	-0.003	0.000	0.000
	(0.002)	(0.002)	(0.002)
Ln(km anchorage) x 1921	-0.010***	-0.006***	0.000
	(0.003)	(0.002)	(.)
Ln(km anchorage) x 1930	-0.010***	-0.007***	-0.007***
	(0.003)	(0.003)	(0.003)
Ln(km anchorage) x 1940	-0.009***	-0.006**	-0.006**
	(0.003)	(0.003)	(0.003)
Ln(km anchorage) x 1950	-0.011***	-0.008***	-0.008***
	(0.003)	(0.003)	(0.003)
Ln(km anchorage) x 1960	-0.011***	-0.008***	-0.008***
	(0.003)	(0.003)	(0.003)
Ln(km anchorage) x 1970	-0.012***	-0.009**	-0.009**
	(0.003)	(0.003)	(0.003)
Ln(km anchorage) x 1980	-0.014***	-0.010***	0.000
	(0.003)	(0.004)	(.)
Ln(km anchorage) x 1990	-0.017***	-0.013***	-0.013***
	(0.004)	(0.004)	(0.004)
Ln(km anchorage) x 2000	-0.018***	-0.014***	-0.014***
	(0.004)	(0.004)	(0.004)
Ln(km anchorage) x 2010	-0.017***	-0.013***	-0.013***
	(0.004)	(0.004)	(0.004)
Maize productivity x 1910	0.008***	0.004**	0.004**
	(0.002)	(0.002)	(0.002)
Maize productivity x 1921	0.007***	0.004	0.000
	(0.003)	(0.003)	(.)
Maize productivity x 1930	0.006**	0.003	0.003
	(0.003)	(0.003)	(0.003)
Maize productivity x 1940	0.003	0.000	0.000
	(0.002)	(0.002)	(0.002)
Maize productivity x 1950	0.003	0.000	0.000
	(0.002)	(0.003)	(0.003)
Maize productivity x 1960	0.002	-0.001	-0.001
	(0.002)	(0.003)	(0.003)
Maize productivity x 1970	0.003	-0.000	-0.000
	(0.002)	(0.003)	(0.003)
Maize productivity x 1980	0.003	-0.000	0.000
	(0.003)	(0.003)	(.)
Maize productivity x 1990	0.008***	0.005	0.005
	(0.003)	(0.003)	(0.003)
Maize productivity x 2000	0.006**	0.004	0.004
	(0.003)	(0.003)	(0.003)
Maize productivity x 2010	0.006**	0.003	0.003
	(0.003)	(0.003)	(0.003)
Ln(km to coast) x 1910	0.001	0.000	0.000
	(0.001)	(0.001)	(0.001)

Ln(km to coast) x 1921	0.003***	0.002***	0.000
	(0.001)	(0.001)	(.)
Ln(km to coast) x 1930	0.002*	0.002*	0.002*
	(0.001)	(0.001)	(0.001)
Ln(km to coast) x 1940	0.002	0.001	0.001
	(0.001)	(0.001)	(0.001)
Ln(km to coast) x 1950	0.003***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)
Ln(km to coast) x 1960	0.003**	0.002**	0.002**
	(0.001)	(0.001)	(0.001)
Ln(km to coast) x 1970	0.003**	0.002*	0.002*
	(0.001)	(0.001)	(0.001)
Ln(km to coast) x 1980	0.003**	0.002*	0.000
	(0.001)	(0.001)	(.)
Ln(km to coast) x 1990	0.004**	0.003**	0.003**
	(0.002)	(0.001)	(0.001)
Ln(km to coast) x 2000	0.003*	0.003*	0.003*
	(0.002)	(0.002)	(0.002)
Ln(km to coast) x 2010	0.003	0.002	0.002
	(0.002)	(0.002)	(0.002)
Ln(km to water) x 1910	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
Ln(km to water) x 1921	-0.002***	-0.002***	0.000
	(0.001)	(0.001)	(.)
Ln(km to water) x 1930	-0.002*	-0.002**	-0.002**
	(0.001)	(0.001)	(0.001)
Ln(km to water) x 1940	-0.002**	-0.002**	-0.002**
	(0.001)	(0.001)	(0.001)
Ln(km to water) x 1950	-0.002***	-0.002***	-0.002***
	(0.001)	(0.001)	(0.001)
Ln(km to water) x 1960	-0.004***	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)
Ln(km to water) x 1970	-0.003***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)
Ln(km to water) x 1980	-0.003***	-0.004***	0.000
	(0.001)	(0.001)	(.)
Ln(km to water) x 1990	-0.003***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)
Ln(km to water) x 2000	-0.003***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)
Ln(km to water) x 2010	-0.002**	-0.002**	-0.002**
	(0.001)	(0.001)	(0.001)
SD(elevation) x 1910	-0.000**	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
SD(elevation) x 1921	-0.000***	-0.000**	0.000
	(0.000)	(0.000)	(.)

SD(elevation) x 1930	-0.000***	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)
SD(elevation) x 1940	-0.000***	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
SD(elevation) x 1950	-0.000***	-0.000*	-0.000*
	(0.000)	(0.000)	(0.000)
SD(elevation) x 1960	-0.000***	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)
SD(elevation) x 1970	-0.000***	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
SD(elevation) x 1980	-0.000***	-0.000**	0.000
	(0.000)	(0.000)	(.)
SD(elevation) x 1990	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
SD(elevation) x 2000	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
SD(elevation) x 2010	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Ln(elevation) x 1910	-0.004*	0.002	0.002
	(0.002)	(0.002)	(0.002)
Ln(elevation) x 1921	-0.005**	0.001	0.000
	(0.003)	(0.003)	(.)
Ln(elevation) x 1930	-0.007**	-0.000	-0.000
	(0.003)	(0.003)	(0.003)
Ln(elevation) x 1940	-0.007***	-0.000	-0.000
	(0.002)	(0.003)	(0.003)
Ln(elevation) x 1950	-0.003	0.003	0.003
	(0.002)	(0.003)	(0.003)
Ln(elevation) x 1960	-0.003	0.004	0.004
	(0.003)	(0.003)	(0.003)
Ln(elevation) x 1970	-0.000	0.006*	0.006*
	(0.003)	(0.003)	(0.003)
Ln(elevation) x 1980	-0.002	0.005	0.000
	(0.003)	(0.003)	(.)
Ln(elevation) x 1990	-0.003	0.004	0.004
	(0.003)	(0.003)	(0.003)
Ln(elevation) x 2000	-0.004	0.003	0.003
	(0.003)	(0.003)	(0.003)
Ln(elevation) x 2010	-0.003	0.004	0.004
	(0.003)	(0.003)	(0.003)
Imputed tributary density x 1910	0.002	-0.001	-0.001
	(0.001)	(0.001)	(0.001)
Imputed tributary density x 1921	0.001	-0.001	0.000
	(0.002)	(0.001)	(.)
Imputed tributary density x 1930	0.002	-0.001	-0.001
	(0.002)	(0.002)	(0.002)

Imputed tributary density x 1940	0.001	-0.002	-0.002
	(0.001)	(0.002)	(0.002)
Imputed tributary density x 1950	0.001	-0.002	-0.002
	(0.001)	(0.002)	(0.002)
Imputed tributary density x 1960	-0.000	-0.003*	-0.003*
	(0.002)	(0.002)	(0.002)
Imputed tributary density x 1970	-0.000	-0.003*	-0.003*
	(0.002)	(0.002)	(0.002)
Imputed tributary density x 1980	0.001	-0.002	0.000
	(0.002)	(0.002)	(.)
Imputed tributary density x 1990	0.002	-0.001	-0.001
	(0.002)	(0.002)	(0.002)
Imputed tributary density x 2000	0.003	0.000	0.000
	(0.002)	(0.002)	(0.002)
Imputed tributary density x 2010	0.005*	0.002	0.002
	(0.003)	(0.002)	(0.002)
City of 15000, t-1, 50 km		-0.002	-0.002
		(0.002)	(0.002)
Constant	-0.000	0.124***	0.124***
	(0.001)	(0.028)	(0.028)
Observations	121368	112032	93360
Adjusted R^2	0.014	0.007	0.008

Unit of observation is the grid cell, 1900-2010. Robust standard errors clustered at state level are in parentheses. Outcome is a binary variable equal to one when the size of the city exceeds the varying threshold. Column (3) excludes the 1921 and 1980 censuses.

Appendix D Timeline of Mexican political economy

Table D9: Key features of eras in Mexican history

Year	Era	Population[a]	Trade	Transport
1520	Colonial era	20 m falls to 1.5 m by 1650 2.5 m	Crown centralizes control of trade	*Tlameme*[b]
1821	Independence; Wars with Spain (1829), US (1846–48), France (1862-67)	6 m	First tariff law: 25 ad valorem tax on all imports; ratio of tariffs to imports: 36% [c]	Single road capable of supported wheeled traffic (Mexico City to Chihuahua, via Zacatecas and Durango)[d]
1876			Ratio of tariffs to imports: 46%	1873: First railroad: Mexico City to Veracruz; $<$ 5 km of road passable by four wheeled carts per 10,000 inhabitants[e]
1877	Porfiriato	10.4 (1884); 12.6 m (1895)	2% commodity exports/GDP[f] 1900: 3% commodity exports/GDP	1877: 640 km rr (mule-driven);
1911		11 cities over 50,000 in 1900	3% commity exports/GDP	$>$ 5,000 km rr, freight transport costs fall by 2/3rds
1910	Revolution & restructuring	15.2 m	4% commodity exports/GDP % exports/GDP: 11.0 in 1930[g]	Railroad destruction Total roads: 91,609 km; "good" roads: 17,670 in 1930
1940		19.7 m	% exports/GDP: 6.7	Additional road construction 1935-1940[h]: 3,694 km
1950	ISI	25.8 m	% exports/GDP: 9.9	13,600 km paved road
1970		48.2 m	% exports/GDP: 7.7	1968: 67,000 km roads[i]
1980	Economic restructuring	66.8 m	% exports/GDP: 10.7	14,225 km rr; 1990: 210,000 km roads; 65,000 km paved[j]
2010		112.3 m	% exports/GDP: 29.9	26,704 km rr[k]; 366,095 km roads, 72,577 km paved

[a]Population sources: colonial: Gerhard (1993*a*); 1821: Coatsworth (1978); 1910-2010: INEGI

[b]Source: Hassig (1985)

[c]Source: Coatsworth & Williamson (2004)

[d]Source: Haber (1992)

[e]Source: Beatty (2001)

[f]These calculations use value of commodity exports from Ficker (2004) and GDP values from Mitchell. I am having trouble reconciling exchange rates for this period, so these are likely to be wrong.

[g]Source for 1930-2010: World Bank (2016)

[h]Source: Reynolds (1970)

[i]Source: Bank (1970)

[j]Source: United States Central Intelligence Agency (1990)

[k]Source: World Bank (2016)

Appendix E Correlates of grid-cell population density

Table E10: Correlates of grid cell population density, 1900

	Dep. var: Ln(population in grid cell)					
	(1)	(2)	(3)	(4)	(5)	(6)
Geographic fundamentals						
Elevation < 1500 m	-0.228	-0.133	-0.161	-0.179	-0.157	-0.173
	(0.147)	(0.169)	(0.139)	(0.138)	(0.138)	(0.137)
Std grid cell elevation	0.000*	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Ln(km coast)	0.042*	0.071**	0.042*	0.051**	0.042*	0.051**
	(0.021)	(0.033)	(0.023)	(0.024)	(0.023)	(0.023)
Ln(km water)	-0.023	-0.011	-0.016	-0.016	-0.016	-0.016
	(0.015)	(0.017)	(0.014)	(0.014)	(0.014)	(0.014)
Ln(maize productivity)	0.351***	0.293***	0.261***	0.283***	0.261***	0.281***
	(0.098)	(0.088)	(0.072)	(0.074)	(0.072)	(0.074)
Trade covariates						
Ln(km to 1878 city, Mexico)	-0.231***	-0.203***	-0.209***	-0.207***	-0.190***	-0.185***
	(0.033)	(0.036)	(0.029)	(0.029)	(0.024)	(0.023)
Ln(km to border)	0.006	0.012	-0.024	0.007	-0.018	0.013
	(0.029)	(0.049)	(0.030)	(0.033)	(0.030)	(0.032)
Ln(km anchorage)	-0.062**	-0.053	-0.035	-0.054	-0.034	-0.052
	(0.030)	(0.055)	(0.036)	(0.036)	(0.036)	(0.036)
Historical population						
Ln(Tributary density, 1570)		0.182***				
		(0.060)				
Ln(1570 trib density, imputed)			0.179***		0.179***	
			(0.061)		(0.061)	
Ln(1570 trib density, zeros for missing)				0.175***		0.176***
				(0.059)		(0.059)
= 1 if missing 1570 data				-0.183*		-0.189*
				(0.107)		(0.106)
City of 15000, t-1, 50 km					0.112	0.131
					(0.088)	(0.090)
Observations	9342	6491	9342	9342	9342	9342
Adjusted R^2	0.362	0.375	0.398	0.399	0.399	0.400

Unit of observation is the 15 x 15 km grid cell. Clustered standard errors are in parentheses.
Coefficients robust to excluding Mexico City area. Reported coefficients are marginal effects from a probit.
* p< 0.10, ** p<0.05, *** p < 0.01.

Table E11: Correlates of grid cell population density, 2010

	Dep. var: Ln(population in grid cell)					
	(1)	(2)	(3)	(4)	(5)	(6)
Geographic fundamentals						
Elevation < 1500 m	-0.102	0.079	-0.006	-0.019	-0.061	-0.074
	(0.219)	(0.247)	(0.208)	(0.210)	(0.179)	(0.180)
Std grid cell elevation	-0.000	-0.001	-0.001	-0.001	-0.000	-0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Ln(km coast)	0.094**	0.140**	0.093**	0.097**	0.073*	0.079**
	(0.037)	(0.051)	(0.039)	(0.039)	(0.038)	(0.038)
Ln(km water)	-0.075***	-0.060**	-0.064***	-0.059**	-0.046**	-0.042**
	(0.023)	(0.029)	(0.023)	(0.024)	(0.020)	(0.021)
Ln(maize productivity)	0.735***	0.556***	0.605***	0.603***	0.528***	0.533***
	(0.145)	(0.113)	(0.123)	(0.111)	(0.097)	(0.090)
Trade covariates						
Ln(km to 1878 city, Mexico)	-0.361***	-0.275***	-0.328***	-0.329***	-0.248***	-0.250***
	(0.054)	(0.061)	(0.050)	(0.051)	(0.051)	(0.051)
Ln(km to border)	0.117	0.263**	0.075	0.100	0.088	0.116
	(0.091)	(0.108)	(0.089)	(0.082)	(0.078)	(0.071)
Ln(km anchorage)	-0.294***	-0.289***	-0.255***	-0.270***	-0.183**	-0.201***
	(0.070)	(0.098)	(0.079)	(0.078)	(0.071)	(0.071)
Historical population						
Ln(Tributary density, 1570)		0.284***				
		(0.090)				
Ln(1570 trib density, imputed)			0.257***		0.245***	
			(0.090)		(0.082)	
Ln(1570 trib density, zeros for missing)				0.276***		0.259***
				(0.088)		(0.078)
= 1 if missing 1570 data				-0.420**		-0.369**
				(0.195)		(0.166)
City of 15000, t-1, 50 km					0.670***	0.657***
					(0.108)	(0.102)
Observations	9342	6491	9342	9342	9342	9342
Adjusted R^2	0.369	0.400	0.393	0.400	0.432	0.436

Unit of observation is the 15 x 15 km grid cell. Clustered standard errors are in parentheses.
Coefficients robust to excluding Mexico City area. Reported coefficients are marginal effects from a probit.
* p< 0.10, ** p<0.05, *** p < 0.01.

Table E12: Correlates of cities: ln(population in cell)

	(1)	(2)	(3)	(4)	(5)	(6)
City of 15000, t-1, 50 km			0.165***	0.173***	0.168***	0.167***
			(0.019)	(0.021)	(0.021)	(0.022)
Ln(km to border), 1910-1940	0.018***	0.002	0.015***	0.002	0.003	-0.002
	(0.006)	(0.005)	(0.006)	(0.005)	(0.005)	(0.007)
Ln(km to border), 1950-1970	0.084***	0.011	0.081***	0.018	0.018	0.014
	(0.018)	(0.021)	(0.017)	(0.019)	(0.019)	(0.020)
Ln(km to border), 1980-2010	0.207***	0.071*	0.199***	0.074**	0.074**	0.081**
	(0.033)	(0.039)	(0.031)	(0.037)	(0.037)	(0.041)
Ln(km to 1878 city), 1910-1940	-0.009***	-0.015***	-0.017***	-0.022***	-0.022***	-0.026***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.005)
Ln(km to 1878 city), 1950-1970	-0.040***	-0.062***	-0.048***	-0.068***	-0.068***	-0.072***
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)
Ln(km to 1878 city), 1980-2010	-0.059***	-0.106***	-0.067***	-0.109***	-0.110***	-0.116***
	(0.016)	(0.013)	(0.015)	(0.012)	(0.012)	(0.014)
Ln(km to anchorage), 1910-1940	-0.035***	-0.024***	-0.034***	-0.024***	-0.025***	-0.042***
	(0.007)	(0.006)	(0.008)	(0.006)	(0.006)	(0.008)
Ln(km to anchorage), 1950-1970	-0.113***	-0.091***	-0.101***	-0.082***	-0.083***	-0.100***
	(0.020)	(0.018)	(0.019)	(0.017)	(0.017)	(0.019)
Ln(km to anchorage), 1980-2010	-0.210***	-0.178***	-0.189***	-0.161***	-0.162***	-0.185***
	(0.031)	(0.031)	(0.030)	(0.029)	(0.029)	(0.032)
Maize productivity, 1910-1940	0.010	0.019***	0.013*	0.019***	0.019***	0.038***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.010)
Maize productivity, 1950-1970	0.113***	0.141***	0.110***	0.135***	0.130***	0.155***
	(0.019)	(0.021)	(0.019)	(0.020)	(0.018)	(0.022)
Maize productivity, 1980-2010	0.196***	0.251***	0.186***	0.235***	0.224***	0.261***
	(0.027)	(0.033)	(0.027)	(0.031)	(0.027)	(0.034)
Ln(km to coast), 1910-1940	0.011**	0.014***	0.011**	0.013***	0.013***	0.020***
	(0.005)	(0.003)	(0.005)	(0.003)	(0.003)	(0.005)
Ln(km to coast), 1950-1970	0.035**	0.046***	0.031*	0.039***	0.037***	0.046***
	(0.017)	(0.011)	(0.017)	(0.011)	(0.011)	(0.012)
Ln(km to coast), 1980-2010	0.057**	0.073***	0.049*	0.061***	0.056***	0.068***
	(0.028)	(0.020)	(0.027)	(0.020)	(0.019)	(0.022)
Ln(km to water), 1910-1940	-0.010***	-0.009***	-0.010***	-0.009***	-0.009***	-0.016***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Ln(km to water), 1950-1970	-0.036***	-0.028***	-0.035***	-0.027***	-0.026***	-0.034***
	(0.008)	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)
Ln(km to water), 1980-2010	-0.048***	-0.041***	-0.044***	-0.037***	-0.034***	-0.044***
	(0.013)	(0.009)	(0.012)	(0.009)	(0.009)	(0.010)
SD elevation, 1910-1940	-0.000***	-0.000	-0.000***	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SD elevation, 1950-1970	-0.001***	-0.000*	-0.001***	-0.000**	-0.000**	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SD elevation, 1980-2010	-0.001***	-0.000**	-0.001***	-0.000**	-0.000**	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Ln(elevation), 1910-1940	0.009**	0.000	0.009**	0.001	0.002	-0.004
	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
Ln(elevation), 1950-1970	0.018	-0.015	0.017	-0.010	-0.008	-0.015
	(0.014)	(0.012)	(0.013)	(0.011)	(0.011)	(0.012)
Ln(elevation), 1980-2010	-0.009	-0.054***	-0.004	-0.043***	-0.038**	-0.051***
	(0.023)	(0.017)	(0.022)	(0.016)	(0.015)	(0.017)
Ln(1570 trib density), 1910-1940	0.014***		0.011***			
	(0.004)		(0.004)			
Ln(1570 trib density), 1950-1970	0.046***		0.039***			
	(0.011)		(0.011)			
Ln(1570 trib density), 1980-2010	0.080***		0.075***			
	(0.017)		(0.016)			
Ln(1570 trib density impute), 1910-1940		0.012***		0.009***		0.005
		(0.004)		(0.004)		(0.005)
Ln(1570 trib density impute), 1950-1970		0.040***		0.033***		0.029***
		(0.010)		(0.010)		(0.011)
Ln(1570 trib density impute), 1980-2010		0.067***		0.063***		0.062***
		(0.016)		(0.016)		(0.017)
Ln(1570 trib density w/zeros), 1910-1940					0.011***	
					(0.003)	
Ln(1570 trib density w/zeros), 1950-1970					0.041***	
					(0.009)	
Ln(1570 trib density w/zeros), 1980-2010					0.078***	
					(0.016)	
Observations	77892	112032	77892	112032	112032	93360
Adjusted R^2	0.514	0.458	0.521	0.466	0.469	0.461

Unit of observation is the grid cell, 1900-2010. Robust standard errors clustered at state level are in parentheses. Outcome is the natural log of the total population in the grid cell. Column (6) excludes 1921 and 1980 censuses. Marginal effects should be interpreted as differences from 1900, the excluded year.

Appendix F Urban primacy in Mexico

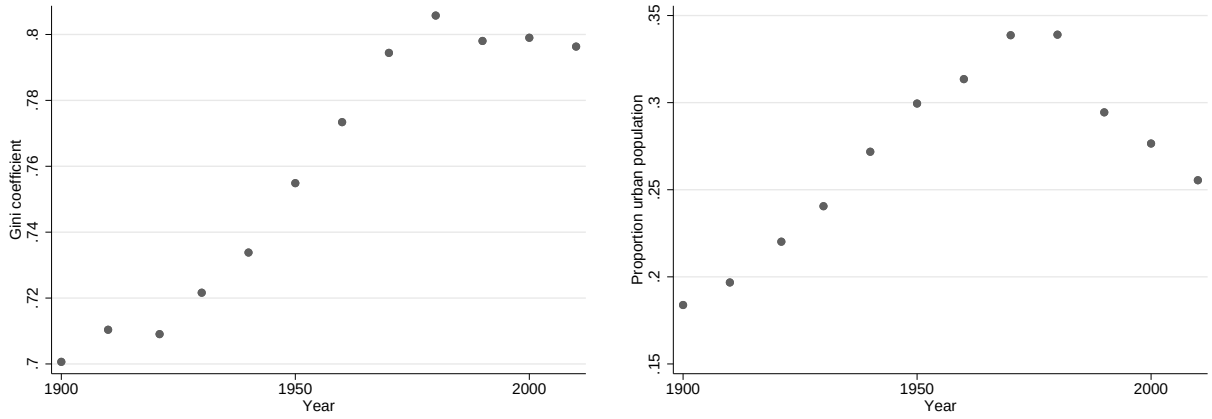
This section examines the timing of changes in the trajectory of urban primacy in Mexico. A major theme of early research in the economic geography of urbanization is the development of "primate cities": cities that dominate a country or region both in terms of size and political influence. Seminal works in this area emphasize the key role of internally-focused trade, high transport costs, and lack of political competition in the development of these cities (Ades & Glaeser 1995, Krugman & Elizondo 1996, Glaeser 2014). One implication of this work is that these metropolises should shrink as a country liberalizes economically and politically. Mexico City is often highlighted as a canonical example of a primate megacity that developed as a result of numerous political (e.g., the centralization of power and one-party rule) and economic (e.g., ISI policies) factors.

In Figure F1a, we plot the Gini coefficient of city size in Mexico by year, using the aggregated AHL population data within metropolitan areas defined by the Mexican government.[17] The Gini coefficient captures inequality in the distribution of urban population across metro areas, specifically how much of the urban population is concentrated in bigger areas. We also plot the percentage of the total urban population of the country living in greater Mexico City for each year in our sample (Figure F1b). The Gini coefficient is relatively flat until 1930, which aligns with the slow growth of urban population up to 1930. It increases steadily until 1980, when it levels off and even slightly decreases post-1990. The proportion of the urban population living in Mexico City increases at an accelerating rate until 1970, and then begins to drop significantly in 1990.

These temporal patterns suggest a correlation between import-led industrialization and ISI in the 1940–1970 period and the concentration of the population in large cities and especially around Mexico City. By contrast, the hold of the Mexico City region begins to weaken as political and economic liberalization progresses. This is in line with other work

[17]This GIS layer, part of the Sistema Urbano Nacional (SUN), was created in 2012 to catalog urban development within Mexico.

Figure F1: Gini coefficients of SUN cities and population in DF



(a) Gini

(b) Proportion urban population in Mexico City

on this subject (e.g., Krugman & Elizondo 1996), and it suggests a role for economic policy in shrinking or amplifying the importance of agglomeration and economic centralization.