The Persistence of (Subnational) Fortune*

William F. Maloney† Felipe Valencia Caicedo ‡

September 15, 2013

Abstract

Using subnational historical data, this paper establishes the within country persistence of economic activity in the New World over the last half millennium, a period including the trauma of the European colonization, the decimation of the native populations, and the imposition of potentially growth inhibiting institutions. We construct a data set incorporating measures of pre-colonial population density, new measures of present regional per capita income and population, and a comprehensive set of locational fundamentals. These fundamentals are shown to have explanatory power: native populations throughout the hemisphere were found in more livable and productive places. We then show that high pre-colonial density areas tend to be dense today: population agglomerations persist. The data and historical evidence suggest this is due partly to locational fundamentals, but also to classic agglomeration effects: colonialists established settlements near existing native populations for reasons of labor, trade, knowledge and defense. The paper then shows that high density (historically prosperous) areas also tend to have higher incomes today, and largely due to agglomeration effects: fortune persists for the United States and most of Latin America.

JEL: J1, N9, R1, O1, O49

Keywords: Persistence, Subnational Growth, Geography, Agglomeration

*Corresponding author Email: wmaloney@worldbank.org. We thank Daron Acemoglu, Laura Chioda, Antonio Ciccone, Ernesto Dal Bo, Carl-Johan Dalgaard, Leo Feler, Claudio Ferraz, Oded Galor, Steve Haber, Lakshmi Iyer, Aart Kraay, Omer Moav, James Robinson, Luis Serven, Hans-Joachim Voth, David Weil, Romain Wacziarg, Noam Yuchtman and participants at the 2012 LACEA-PEG conference, 2013 Warwick Summer School in Economic Growth, and 2013 Economic History Association Meeting for helpful discussions. We are especially grateful to Miriam Bruhn and Francisco Gallego, and to Navin Ramankutty for their kind sharing of data, and Henry Jewell for help processing the HydroSTAT data. Special thanks to Mauricio Sarrias for inspired research support. This paper was partly supported by the regional studies budget of the World Bank Office of the Chief Economist for Latin America and the Knowledge for Change Program (KCP).

†Development Economics Research Group, The World Bank, and Universidad de los Andes, Bogotá, Colombia.

‡Universitat Pompeu Fabra and London School of Economics.

1 Introduction

Tenochtitlán was home to one of the largest concentrations of indigenous peoples in the New World when it was conquered by the Spaniards five centuries ago, and constituted an urban agglomeration rivaling those of Europe. In the words of Hernán Cortés (1522):

> This great city of Tenochtitlán is as big as Seville or Córdoba...It has many plazas where commerce abounds, one of which is twice as large as the city of Salamanca...and where there are usually more than 60,000 souls buying and selling every type of merchandise from every land... There are as many as forty towers, all of which are so high that in the case of the largest there are fifty steps leading up to the main part of it and the most important of these towers is higher than that of the cathedral of Seville. The quality of their construction, both in masonry and woodwork, is unsurpassed anywhere.[1]

México City, erected on the ruins of Tenochtitlán, remains one of the largest and most prosperous cities in Latin America. This paper uses new subnational data from 18 countries in the Western Hemisphere to examine the degree to which such persistence is generally the case: Do rich (high pre-colonial population density) areas before the arrival of Columbus tend to be populous and rich today?

Most similar to the present work, Davis and Weinstein (2002) find persistence in Japanese population concentrations over very long historical spells, and despite massive wartime devastation.[2] Other recent works suggest persistence in economic activity over thousands (Comin et al., 2010) or tens of thousands of years (Ashraf and Galor, 2012). Such persistence is consistent first, with the importance of locational fundamentals such as safe harbors, climates suitable to agriculture, rivers, or concentrations of natural resources that, even if not used for exactly the same purposes, nonetheless retain value over time (Ellison and Glaeser (1999), Rappaport and Sachs (2003), Fujita and Mori (1996), Gallup et al. (1998), and Easterly and Levine (2003)). It may also suggest the importance of agglomeration effects, perhaps arising from increasing returns to scale (see Krugman,

[1] *La Gran Tenochtitlán, Segunda Carta de Relación (1522).* Authors' translation.

[2] In a tragically similar case to that of Hiroshima and Nagasaki, Miguel and Roland (2011) find that heavily bombed areas of Vietnam also recovered almost fully relative to non-bombed areas.

1991, 1993) or Marshallian externalities arising from human capital, infrastructure and technological externalities (Krugman (1992), Comin et al. (2010), Glaeser et al. (1992), Bleakley and Lin (2012), Severnini (2012)) which may lead to path dependence and persistence across time even after the initial attraction of a site has faded in importance.[3] The case of the trauma of the colonization of the New World is especially interesting since, while Hiroshima in Davis and Weinstein (2002)'s work proved resistent to losing a quarter of its population, the American Indian population was literally decimated, falling by up to 90%. Further, the conquest implied the wholesale imposition of distinct culture (preferences), political organization, and technologies. The power of agglomeration effects to preserve the spatial distribution of prosperity in the face of such shocks is far from obvious.

In particular, working against persistence in the context of colonized areas, Acemoglu et al. (2002) argue for what they term a "reversal of fortune:" areas colonized that had large populations of exploitable indigenous populations developed extractive institutions that were, particularly during the second Industrial Revolution, growth impeding. Following Malthus in associating high pre-colonial population density with more productive and prosperous areas in pre-industrial periods (see also Becker et al. (1999), Galor and Weil (1999), Lucas (2004)), they find a negative correlation between pre-colonial population density and present day incomes. This finding has been influential in moving institutions to center stage in the growth debate, and suggests that such forces can more than fully offset agglomeration and locational forces.

This paper revisits the persistence question at the subnational (state, department, region) level for the Western Hemisphere. We focus on the Americas because of the availability of anthropological and archaeological estimates of indigenous population densities before Columbus at a geographically disaggregated level, the near universal colonization by one or more European powers, and the diversity of subsequent growth experiences. We match

[3]For a discussion of the importance of these effects for the ongoing evolution of economic geography among developing countries see World Bank (2008).

the pre-colonial population estimates to new data on present population density and per capita income generated from household surveys and poverty maps. We then incorporate a comprehensive set of geographic controls, including new measures of agricultural suitability and river density, which we show to have explanatory power as locational fundamentals determining pre-colonial settlement patterns. Data at this finer level of geographical aggregation allow us to take a more granular look at the role of locational, agglomeration and institutional forces behind the distribution of economic activity. In particular, using subnational data with country fixed effects mitigates identification problems caused by unobserved country or region specific factors arising from particular cultural or historical inheritances, and national policies, albeit, now asking the question at a level of aggregation where the relative influence of forces for and against persistence may differ.

Our empirical results suggest that, within countries, the forces for persistence dominate. Population density today is strongly and robustly correlated with pre-colonial population density as is current per capita income. Clearly, the puzzle arises of why these findings should diverge so sharply from those found internationally and we examine the forces potentially driving the difference. Both statistical and historical evidence suggests that both locational fundamentals and agglomeration externalities plausibly explain why such persistence should occur despite the violent interaction of cultures of entirely distinct cultural, economic, institutional and technological characteristics.

2 Data

The use of subnational data to explore differential performance along various dimensions is now well established. As noted above, Davis and Weinstein (2002) use regional level data for Japan to document the remarkable persistence of population densities, highlighting the importance of both locational fundamentals and increasing returns to scale. Mitchener and McLean (2003) exploit modes of production and geographical isolation leading to differential de facto institutions as explaining differential growth rates across US states. Banerjee and Iyer (2005) exploit the variation in colonial property rights institutions across India to explain relative performance in agricultural investments, productivity and human development outcomes. Michalopoulos and Papaioannou (2013) find regional economic performance in Africa to be determined more by pre-colonial ethnic grouping than but subsequent national institutions. Within Colombia, present regional development outcomes are shown to be affected by colonial institutions such as slavery, Encomienda, and early state capacity (García-Jimeno, 2005); and slavery (Bonet and Roca, 2006; Acemoglu et al., 2012). Naritomi et al. (2012) analyze how variations in colonial de facto institutions in Brazil led to different public good provision outcomes in modern times. Acemoglu and Dell (2010) examine differences in productivity across Latin American subregions and postulate that the large differences in institutions and enforcement of property rights, entry barriers and freeness and fairness of elections for varying levels of government are important. Dell (2011) uses district level data from Perú and Bolivia to demonstrate the long term impact of the Mita on development through the channels of land tenure and long term public goods provision.[4] In a kindred paper using subnational data across the hemisphere, Bruhn and Gallego (2011) argue that differences in the types of regional colonial activities, whether engendering extractive or inclusive institutions, lead to lower or higher incomes, respectively. Most recently, Gennaioli et al. (2011) use subnational data from 110 countries to argue for the overriding importance of human capital in accounting for regional differences in development.

2.1 Population

We compile subnational data on pre-colonial population densities, contemporary population densities and household incomes for the 18 countries in the hemisphere listed in Table 1, which summarizes the data by country.

[4] Regional differences in institutional arrangements have also been documented in the case of slavery in the US and Brazil (Degler, 1970), and sharecropping and women's rights in Colombia (Safford and Palacios, 1998; Bushnell, 1993).

Pre-colonial Population Density: This measures the estimated number of indigenous people per square kilometer just before colonization. These data draw on a long tradition of academic research dating from the turn of the last century, much of it fuel for the debate over whether the colonial powers encountered a "pristine wilderness" or, alternatively, a world densely inhabited by indigenous peoples subsequently devastated by disease and conquest (Denevan, 1992b). The estimates contributed by the authors in *The Native Population in the Americas in 1492* (Denevan, 1992a) are among the most comprehensive and refined to date and they form the core of the data. The details on the construction of the pre-colonial density measures and their mapping to modern subnational units can be found in Bruhn and Gallego (2011). We expand the sample further using analogous data on Canada from Ubelaker (1988), and Nicaragua from Newson (1982).

Though the project of estimating populations half a millennium past is necessarily speculative, the estimates synthesize the most recent available geographical, anthropological, and archeological findings. In particular, they draw on documentary evidence such as reports by Europeans, actual counts from church and tax records, as well as contemporary and recollected native estimates and counts. Depending on the country, projections across similar geographic areas, regional depopulation ratios, age-sex pyramids, and counts from subsamples of the population (such as warriors, adult males, tribute payers) are used, as well as backward projections from the time of contact with Europeans. These are corroborated by evidence including archaeological findings, skeletal counts, social structure, food production, intensive agricultural relics, carrying capacity, and environmental modification. Importantly, neither modern GDP, climate models nor current population measures are used in the construction of these estimates.

Figure 1 maps these pre-colonial population densities. While some related studies have focused on cities as the unit of observation, such data are not available at our frequency for the pre-colonial period and we work with regional densities. However, as Davis and

Weinstein (2002) note, for numerous reasons in particular related to defining a city over time, estimated regional population densities are arguably preferable. When we aggregate back up to the national level, our estimates correlate highly and significantly with other popular aggregate sources such as McEvedy et al. (1978) used in, for instance, Acemoglu et al. (2002).

Present Subnational Population Density: This measures present population per square kilometer in each subnational unit and is drawn from a highly disaggregated spatial data set on population, income and poverty constructed on the basis of national census data by the World Bank (2008) for the World Development Report on *Reshaping Economic Geography*. Population is aggregated from the census by the present subnational unit and the density is then calculated.[5]

2.2 Income

Subnational Income per Capita: Income in 2005 PPP US dollars is drawn from the same spatial data set.[6] Household income data are preferable to national accounts data as a measure of regional prosperity. In the case of natural resource rich regions, income may or may not accrue to the locality where it is generated and hence may provide a distorted measure of level of development. As an example, the revenues from oil pumped in Tabasco and Campeche, México, are shared throughout the country, although they are sometimes (but not always) attributed entirely to the source state in the national accounts (see Aroca

[5]Censuses: US, El Salvador 2000: Brazil, Panama; 2001: Bolivia, Ecuador; 2002: Chile, Guatemala, Paraguay; 2005: México, Nicaragua, Perú, 2006: Uruguay. All other countries: figures correspond to survey data estimates at the regional level or small-area estimates based on survey and census data.

[6]Household level data sets are combined with limited or non-representative coverage with census data to generate income maps for much of the hemisphere (see Elbers et al., 2003). This approach addresses the problem that in some cases such as México household income surveys are not representative at the "state" level. We thank Gabriel Demombynes for providing the data. See original study for methodological details. We expect that while somewhat more complete, our data is similar to the census based data used by Acemoglu and Dell (2010). For Argentina, Colombia and Venezuela, the spatial data base reports the unsatisfied basic needs index rather than income. We project subnational GDP (production) series on this index to scale it to household income. GDP source: Argentina (CEPAL, Consejo Federal de Inversiones, Colombia (DANE)), Venezuela (Instituto Nacional de Estadística). We expand the sample to include Canada and the United States using the (2005) censuses. The resulting estimates of mean per capita income have been rescaled so that the population-weighted average matches 2007 GDP per capita at 2005 US dollars (PPP adjusted).

et al., 2005). This is a broader issue that emerges wherever resource enclaves are important. For instance, from a national accounts point of view, the richest subnational units in Argentina, Colombia, Chile and Perú, respectively, are Tierra del Fuego (oil), Casanare (oil), Antofagasta (copper), and Moquegua (copper), all of which, with the exception of the last, are average or below average in our household survey measured income. Further, the geographical inhospitability of these locales ensured and continues to ensure relatively little human habitation: Antofagasta is in the driest desert in the world and Tierra del Fuego is the closest point in the hemisphere to Antarctica. This combination can give rise to a negative, although relatively uninteresting, correlation of pre-colonial population density with present income. That said, such correlations still emerge even in our income data due to the selection of the population in these areas: The very small population related to extraction of natural resources has relatively high levels of human capital and remuneration and hence, we may still find that areas which the indigenous avoided are now relatively well-off in per capita terms.

2.3 Locational Fundamentals

To establish the importance of locational fundamentals, we match the population and income subnational data to a comprehensive set of geographical controls. Accounts of 18th century explorers, and anthropological studies confirm the importance to Indian settlements of both arable land and waterways for food and transport, characteristics also attractive to subsequent European settlers and potentially current inhabitants.[7] We incorporate two new measures to capture agricultural suitability and river density.

Suitability for Agriculture: Since agriculture was critical to early settlement, we employ

[7]Denevan (1992b) discusses the extensive evidence on the importance of agriculture throughout the hemisphere in precolonial times. De Vorsey Jr (1986) cites the 18^{th} century explorer William Bartram as noting that "An Indian town is generally so situated, as to be convenient for procuring game, secure from sudden invasion, a large district of excellent arable land adjoining, or in its vicinity, if possible on an isthmus betwixt two waters, or where the doubling of a river forms a peninsula; such a situation generally comprises a sufficient body of excellent land for planting corn, potatoes, squash, pumpkins, citrus, melons, etc." p. 13.

a new measure of agricultural suitability as developed by Ramankutty et al. (2002). This measure uses a combination of three different data sets. First, it calibrates the satellite-based International Geosphere-Biosphere Programme Data and Information System (IGBP-DIS) 1 km land-cover classification data set against a worldwide collection of agricultural census data to capture cultivable land. To derive climatic parameters that may restrict the use of this soil, a second data set captures the mean-monthly climate conditions including temperature, precipitation and potential sunshine hours. Finally, it draws on the IGBP-DIS global soil data sets that contains soil properties such as soil carbon density, nitrogen content, pH, and water holding capacity. Combining these through a model of land suitability, Ramankutty et al. (2002) generate an index of the probability that a particular grid cell will be cultivated. We employ a spatial average of this measure over subnational units.

Waterways and Coasts: For measures of the ubiquity of settlement-suitable waterways, we employ the recently developed HydroSHEDS data that provide globally consistent hydrographic information at high resolution as collected during a Space Shuttle flight for NASA's Shuttle Radar Topography Mission (SRTM). HydroSHEDS generates a mapping of river systems from which we develop a measure of the number of potentially suitable river sites based on the density of rivers in each geographic unit.[8]

Clearly, populations could also be sustained by marine-based economies where farmland and rivers were of less importance. Hence proximity to the coast for saltwater trade, transport, fishing potential and amenities potentially persists in importance, much as it was subsequently for European settlement, and to capture this we employ a measure of whether or not the region is landlocked.

[8]HydroSHEDS stands for Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales. The HydroSHEDS project was developed by the World Wildlife Fund and U.S. Geological Survey among other organizations.Densities were calculated using zonal statistics in ARC-GIS map. Though HydroSHEDS depicts the flow of cells into a given river system, beyond a certain size we do not take into account the flow of the river per se for two reasons. First, settlements are not likely to be proportional to the size of a river, again, beyond a certain threshold. Second, due to the geographical projection, the cells do not map precisely one to one to actual flows.

We also employ several other controls that capture suitability for human settlement as collected by Bruhn and Gallego (2011): average temperature in degrees Celsius, altitude of the capital city of the state in kilometers, and annual rainfall in meters. Some of these clearly overlap the agricultural suitability measure and hence need to be interpreted as capturing effects beyond those on agriculture. Table 2 summarizes the data.[9]

3 Empirical Results

3.1 Locational Fundamentals and Pre-Colonial Densities

Figure 1 and Table 1 present a map and summary statistics of pre-colonial population densities. What is immediately clear is the great heterogeneity of pre-colonial population densities both within and between countries, as well as the substantial overlap of distributions across countries. The Latin American countries span densities averaging from around 0.4 person per square kilometer for Argentina to 1.7 for Venezuela, 2.5 for Brazil to 17 for Perú and 32 for México. Further, Table 1 confirms a large range of variances of initial density within country. México and Perú are not only dense on average, but have much larger variances than, for instance, the US.

However, overall, the US and Canada fit comfortably in the Latin American distribution. With a mean population density of .39, the US is above Uruguay and is roughly the same as Argentina. Canada, at 1.22 is above Argentina, Bolivia, and Uruguay and is just below Paraguay and not so far from Venezuela. Looking at both mean and variance, the US and Argentina are effectively identical: $(.39, 1.34)$ vs $(.44, 1.45)$. A clear divide between hemispheres on the basis of population density is thus not clear.

[9] Ashraf and Galor (2011b) have further suggested as determinants of population time elapsed since the Neolithic revolution, distance from the regional technological frontier, and absolute latitude. As their data is at country level, these effects would be absorbed by the fixed effects and we do not include them.

As a first check on the relevance of our locational fundamentals proxies, Tables 3 report the results of running

$$D_{Precol,ij} = \alpha + \gamma LF_{ij} + \gamma_2 LF_{ij}^2 + \mu_i + \epsilon_{ij} \tag{1}$$

where $D_{Precol,ij}$, is pre-colonial density. LF is a the vector of subnational locational fundamentals and μ_i is a country specific fixed effect. Though we report a specification where fundamentals enter linearly, we also report one with quadratic terms to acknowledge that human utility is not likely to be linear in locational fundamentals, most obviously for temperature and rainfall. With altitude, also, gaining some height may limit the likelihood of disease or invasion, but the benefits potentially decline again after a point.

Table 3 presents estimates both with and without fixed effects (FE). We report the latter despite the fact that the territorial boundaries and corresponding national governments, institutions, and other characteristics clearly were not established at the time for two reasons. First, the generation of the pre-colonial populations was done with the present national boundaries defining the unit of analysis and by different authors and hence there may be subtle differences at that level. Second, in subsequent regressions we will care very much about abstracting from country wide effects and hence the analogous specification is desirable for reference.

For robustness purposes, we also report the corresponding results using the MS (or M-S) estimator (Maronna and Yohai, 2000). Our data combine countries with very different levels and variance of initial population densities as well as often very dramatic spreads within countries. México, for instance, has many states with modest densities, but then Morelos and México City are five to ten times as large. The MS estimator is designed to handle potential good and bad leverage points (explanatory values with extreme values) as well as conventional vertical outliers in a structured way that has been shown superior to other alternatives like quantile or robust regression.[10] The MS FE are our preferred estimates and we base our calculations of maximum and minimum influence on them. R^2 are not calculated for the MS estimator.

Locational fundamentals appear important to where pre-colonial native populations were concentrated. In general, the MS estimators generate more significant coefficients consistent with standard errors that are estimated taking into account outliers and extreme values and we focus mostly on these. Suitability for agriculture enters significantly and positively in all MS specifications (columns 5-8), albeit with diminishing value. The negative quadratic term reflects the low population density found both in the US Midwest and the high suitability areas of both Argentina and Brazil which, for whatever informational or historical reasons, were not heavily settled. But, overall, pre-colonial native populations were attracted to good farming conditions.

The relatively high densities found in often arid relatively unsuitable areas along the coasts (see Figure 1) may reflect the marine, rather than agricultural basis of the local economy. Consistent with Rappaport and Sachs (2003), being landlocked enters negatively in all MS and MS FE specifications although it loses significance when FEs are included. The weakness of this result is less surprising when we consider that the densest concentrations are often found inland: Misiones and Corrientes in Argentina; La Paz in Bolivia; Santiago in Chile: Bogota/Cundinamarca, and Antioquia in Colombia, Mexico City and Morelos, in

[10]The estimator is a combination of M and S estimates. In the case where some variables are categorical (0-1) and some are continuous and random which may contain leverage points, as is the case here, M estimates are not robust and S estimates are computationally intensive. The MS estimator combines both and though less well-known than, for instance, quantile or robust regression for managing potential outliers, it has several advantages. First, it is more robust to bad leverage points. Second, it is likely to provide more efficient estimates of the standard errors than the bootstrapped quantile estimates since it adjusts for outliers. Finally, it attains the maximum breakdown point, being robust to up to half of the observations being contaminated. In practice, a sizable share of our observations in all pooled specifications are identified as outliers and hence a high breakdown point is desirable. The standard "robust" estimator in STATA is a class of M estimator, however it is not robust, in particular, to masked outliers. That is, when calculating whether a observation has a standardized distance above a critical value, it uses the variance calculated using outliers. An upward biased estimate of the standard deviation may therefore allow a true outlier to remain in the sample. The MS estimator obviates this problem.

Mexico. These inland concentrations rely on rivers, instead, for transport and perhaps food. However, our measure of river density enters negatively in the OLS and MS FE specifications (columns 1 and 7). Looking at Figure 1, this result is partly driven by the Amazon which, despite having the richest network of waterways in the hemisphere, has very low indigenous density and which includes a substantial number of subnational units in Venezuela, Colombia, Ecuador, Perú and especially Brazil. A similar pattern is found in the US where, despite relatively (for the US) high population concentrations in the Mississippi watershed, particularly the present day states of Mississippi and Louisiana, even higher concentration are found in Connecticut, California, Massachusetts,and Rhode Island which show lesser river densities.

Average temperature enters positively and significantly in the OLS, MS and MS FE estimates (columns 1,5,7). The MS quadratic specification suggests some convexity although this dissapears with fixed effects. Perhaps for reasons of avoiding disease, altitude enters positively in all specifications although only significantly in the MS and MS quadratic specifications. Rainfall enters negatively in all specifications, although only significantly in the quadratic term of the MS FE specification. Recalling that the agricultural suitability term has already factored in rainfall, this finding is telling us that beyond what is needed for agriculture, rainfall also has diminishing value.

In sum, our locational fundamentals proxies suggest that indigenous populations were concerned with agriculture, transport and fishing, being warm, staying high, perhaps to avoid diseases, and not being too wet.

3.2 Persistence: Overview and Specification

We next explore the correlation of pre-colonial population densities with present population densities and with present per capita income. Again, the summary statistics for all three variables are found in Table 1. As a first look at the data, Figures 2 and 3 offer a striking fact. For the US, both today's state level population density and income per capita are positively correlated with the density of the indigenous population before the arrival of Columbus: economic activity appears to persist. In the next two sections, we document these correlations more rigorously for a broad set of countries of differing pre-colonial densities and present per capita incomes.

The core specification for the pooled data β:

$$D_{2005,ij}; Y_{2005,ij} = \alpha + \beta D_{precol,ij} + \gamma LF_{ij} + \mu_i + \epsilon_{ij} \tag{2}$$

where $D_{2005,ij}$, population density of subunit i of country j, and $Y_{2005,ij}$, present per capita income, are sequentially the dependent variables.[11]

3.3 Current Population

3.3.1 Evidence for persistence in population

Because we do not present individual country graphs for present population, summary Table 4 reports β from the country by country regressions of present on pre-colonial populations density. The estimations are in OLS due to limited observations and hence are subject to the outlier concerns discussed above. However, they confirm that the positive relationship found for the US is significant in both the log-log and level-level specifications and, in fact, they suggest it as an overall stylized fact for the hemisphere. In both regression specifications, the majority, 15 and 11 of 18 respectively, show a significant and positive elasticity. Canada is the only one to show a significant negative coefficient, largely driven by the Arctic Northwest Territories, Yukon and Nunavut which have relatively lower population densities today. 14 of the 18 countries show significant and positive rank correlations, 12 show correlations that exceed .5, and Chile, Guatemala, Nicaragua and Venezuela all exceed .75. Overall, the magnitudes are broadly similar to the .71 found by Davis and Weinstein (2002) for Japan over the period CE 725 to 1998. Consistent with the lower

[11]Davis and Weinstein (2002) report the raw correlation coefficient rather than OLS coefficients. We do not do this for reasons of comparability to the multivariate regressions that follow.

coefficient in the log-log specification, the US is among the lowest of those showing a positive and significant correlation at .37. And again, Canada is the only negative and significant entrant. In general, the Latin American countries show far higher degrees of population persistence than the US or Canada. This may partially reflect the dramatic differences in immigration experiences at the *country* level between the US and Canada on one hand, and Latin America on the other which, with the exception of Argentina, was relatively closed.[12]

Table 5 pools the countries. Here, fixed effects are potentially of greater importance than in the last section because of the desire to control for country level historical effects or policies that would affect the between dimension, and we focus primarily on the FE and MS FE estimates. That said, all specifications, with the exception of the between, are positive and significant. Population density shows strong persistence across time.

We include the fundamentals in quadratic form to absorb as much fundamental influence as possible and this lowers the magnitude of the persistence coefficient somewhat but leaves it strongly significant: the OLS FE estimates do not change (column 5), but they substantially do in the MS case (column 7) where they fall from 3 to .6 reflecting the treatment of outliers. These fundamentals enter broadly similarly to the way they did in the previous exercise, although with some important differences. Agricultural suitability is not significant in any specification confirming that it is not the driver of population agglomeration in the modern world. Rivers emerge with roughly the same degree of significance as previously with a negative quadratic term in the MS specification. Landlocked is negative and significant only in the MS FE specification. Temperature enters convexly, again, although generally more

[12]From the point of view of establishing the particular channels postulated by the reversal of fortune literature, it may be argued that capital cities have a *sui generis* dynamic and should be excluded. From a general point of view of understanding agglomeration effects and persistence, this is less clear- whatever the impetus that established these cities, the existing megalopolises in Latin America are not supported in the main by government activities at present. Precisely the emergence of such "Urban Giants" has been studied by Ades and Glaeser (1995), while Krugman and Elizondo (1996) have focused on México City. In the end, even dropping these overall strengthens the persistence results. The levels levels regression for Argentina, Brazil, and Guatemala become significantly positive and nothing becomes less so. We thank Daron Acegmoglu for bringing this point to our attention.

significantly. Altitude is generally insignificant. Rainfall again appears significant in both terms in the OLS FE and MS FE.

In sum, many of the locational fundamentals emerge statistically significant in determining current populations and with similar sign as they did in explaining pre-colonial population density, albeit contributing modestly to explanatory power. Critically, however, despite a quite complete set of locational controls, the pre-colonial densities themselves appear to be robustly significant. Much of the persistence appears to exist for reasons related to the existence of the populations themselves.

3.3.2 Population Persistence Mechanisms

Beyond the geographical factors discussed above, plausible mechanisms through which populations themselves could drive persistence can be found in the value of the populations themselves, knowledge and information, and complementary technologies brought by the colonists.

Workers, Citizens and Souls. In Latin America, native populations were indeed a source of tribute and labor and hence it is not surprising that Spanish cities would be built near existing population centers, whatever factors drove their initial settlement. In other regions under Spanish colonization, the native populations were valued for otherworldly and strategic reasons. The missions set up along the Alta California (now US) coast-San Diego, Los Angeles, Santa Barbara, San Jose and San Francisco-were established beside major native population centers (as in the Southwest) to recruit souls to Christianity, but also to create colonial subjects to occupy territories perceived under threat of English and Russian encroachment (Taylor, 2001). In these cases, it was the population agglomeration itself, rather than the locational fundamentals per se, that were the attraction and exploitation was not the primary motivation.

Knowledge and Commerce. In the US, pre-colonial native populations were relatively small, topping out at around 2 people per square kilometer, and they were generally, with the exception of South Carolina (Breen, 1984), not exploited for tribute or labor by French, Anglo and Dutch colonizers. This suggests that while the argument that the Spanish and Portuguese located near indigenous populations for purposes of tribute or forced labor through the Encomienda or Mita is compelling, it is not the only mechanism through which pre-colonial agglomerations were perpetuated. To begin, throughout the New World explorers depended on native cartography and knowledge to map the relevant geographical and demographic sites (De Vorsey, 1978). New settlement was likely not to be random, but influenced by the previous "known world". In addition, colonizers needed the knowledge and skills accumulated by the native populations. Cortés employed the stone masons and architects of the pyramids, canals and aqueducts of vanquished Tenochtitlán to remodel Moctezuma's palace into his own, and to raise the most important city in the New World from the ruins of the Aztec capital. The large population of craftsmen and artisans was of world caliber (Parkes, 1969). The conquistadors more fundamentally needed those with a knowledge of plant life, agronomy, and hunting to feed their new towns. Hence, just by virtue of already supporting a civilization in all its dimensions, Tenochtitlán was attractive beyond the brute labor force it offered and in spite of its actually lackluster locational fundamentals.[13]

In non-Spanish North America, the competing colonial powers also established many cities including Albany (Dutch), Augusta (British), New York (Dutch), Philadelphia (British), Pittsburgh (French and British), St. ouis (French) on or next to native population

[13]Tenochtitlán's location was allegedly determined by the god Huitzilopochtli to be established on a small, swampy, island whose chief attraction appears to be that it was uncoveted by the neighboring tribes and was defensible. Parkes (1969) notes that the Mexica (Aztecs) were the last tribe of seven to enter the valley and wandered as outcasts, selling their services as mercenaries to the dominant tribes, and eating reptiles and pond scum to survive. They had the worst pickings of a not entirely favorable locale. The valley of México, and in particular Tenochtitlán, had unreliable weather, with a short growing season and frequent drought. Famine was not uncommon. The lake was subject to storms and a major flood in 1499 caused the loss of much of Tenochtitlán (Thomas, 1993). Simpson (1962) notes that "With the silting up of the lakes and consequent flooding, the city was frequently inundated with its own filth and became a pest hole. Epidemics were a scourge for centuries and were not brought under control until the opening of the Tequixquiac drainage tunnel in 1900" (page 164).

settlements. Partly, the colonists, like the native populations, valued the areas of rich alluvial lands along the major river systems that served as the primary mode of transportation and communication, or the strategic locations. Bleakley and Lin (2012) argue that portage sites around rapids or falls gave rise to agglomerations in commerce and manufacturing that persist today, suggesting path dependence and increasing returns to scale. However, the native populations were critical attractions in themselves as well, again, largely for informational, commercial, and strategic reasons. As Taylor (2001) notes, "On their contested frontiers, each empire desperately needed Indians as trading partners, guides, religious converts, and military allies. Indian relations were central to the development of every colonial region"(p. 49). From Canada to Louisiana, trade and defense led the French to establish their trading posts as nodes of trade and negotiation for securing alliances and food. In the North, French and Dutch Fur traders exploited existing networks of native tribes as suppliers of pelts. Quebec, for example, was located in an area where the local natives were skilled hunters and the nearby and numerous Huron nation served as provisioners and trade middlemen. Similarly, on Vancouver Island and throughout the Pacific Northwest, the British traded extensively with natives in sea otter pelts. Linking geographical fundamentals and pre-colonial populations, Bleakley and Lin (2012) note that the portage site on the Savannah River (now Augusta) was an important collection point for pelts brought by the native hunters. Pre-colonial Indian population concentrations offered benefits to colonizers along many dimensions, and those of trade in goods and information are classic positive externalities associated with agglomerations.

Complementary Technologies. As a final effect working in the opposite direction, for some indigenous agglomerations, contact with European culture and technology may have perpetuated their dominance after an initial period of trauma, particularly given the proximity to the Industrial Revolution. Comin et al. (2010) for instance, document an association between technology in 1500 AD and present income, roughly our period. Ashraf and Galor (2011a) argue that at the moment of transition between technological regimes, more cultural diffusion facilitates innovation and the adoption of new technologies. As one suggestive example, Steckel and Prince (2001) argue that one reason that the US plains Native Americans were the tallest people in the world in the mid-19th century was the buffalo and game made more accessible with the introduction of horses, metal tools, and guns by Europeans (see also Coatsworth (2008)). Our documented patterns of persistence may therefore be partly driven by the degree to which the conquest transferred the old world technological endowment. The population centers of the earlier Maya, Anasazi, and Toltec civilizations have vanished. Perhaps partly because of their contact with the Spanish, the Aztec population center persists.

Taken together, locational fundamentals, agglomeration externalities, and technological transfer may plausibly contribute to an explanation of why pre-colonial densities, even after the decimation of the local population, mapped to early colonial densities which, in turn, have persisted to this day.

3.4 Current Income

3.4.1 Evidence for persistence of income

The previous section confirms for the Americas Davis and Weinstein's (2002) finding that population density is persistent over very long periods of time. A large literature argues from Malthus that high population densities in pre-industrial periods signal higher productivity and prosperity (see, for example, Becker et al. (1999), Galor and Weil (1999), Acemoglu et al. (2002), Lucas (2004)). The relationship between present population and present income may be expected to be less tight than historically was the case for at least two reasons: as Ashraf and Galor (2011b) and Galor (2011) note, the traditional Malthusian relationship between population and wealth weakens with technological progress, and the natural resource endowment effects discussed earlier. Hence, we do not interpret our population persistence results as more than just that. However, Acemoglu et al. (2002)

precisely argue at the country level, that high prosperity areas in pre-colonial times, measured by population density, became low prosperity areas today as measured by GDP per capita. The causality, again is that higher indigenous population densities led to extractive institutions that subsequently depress growth. We investigate the Acemoglu et al. (2002) reversal at the subnational level. Because the relationship of pre-colonial populations density is more heterogeous and complex than that with population, we reproduce figures for eight informative countries in Figure 4 which again, with two exceptions, suggest persistence.

Table 6 shows a negative, but insignificant relationship between pre-colonial densities and present income in the OLS and between estimators, broadly consistent with Acemoglu et al. (2002), but a strongly significant positive relationship in all of the within estimates.[14] Including the locational fundamentals complicates the picture of the relative contribution of geography and agglomeration externalities to these results. Generally, the influence of locational fundamentals appears attenuated somewhat: agricultural suitability, river density and temperature never appear significantly. In modern economies, farming areas are no longer the richest areas of the economy; river density is not essential for fishing or transport; air conditioning allows living in desertic areas such as the US Southwest. Landlocked enters negatively and significantly and altitude and rainfall enter positively with diminishing effect as in the pre-colonial regressions.

The OLS FE estimates remain unchanged at .09 and become somewhat more statistically significant. The MS FE freestanding estimates are lower than the OLS FE case (.06), but actually experience a statistically significant *rise* to .1 when the fundamentals are included. Whether due to the fact that altitude and rainfall now enter with reversed sign relative to the pre-colonial density regressions, or other factors, locational fundamentals appear to *decrease* persistence in this case. The specifications strongly support persistence of fortune

[14]We also run log-log specifications which are less robust but in which no significant relationship for either persistence or reversals emerges. Despite leaving unattentuated some quite extreme values, we find the persistence findings from the log-level specification with the MS corrections for outliers and bad leverage points more defendable.

that is *not* related to locational fundamentals: high density pre-colonial density areas appear richer now because they were dense.

These results raise the question of why, at the subnational level, we find evidence for persistence, while at the national level Acemoglu et al. (2002) find reversals. Several hypotheses come to mind. First, it may be that national institutions intrinsically dominate those subnational. This cannot be taken as axiomatic, however. Colombia, for instance, is famously fragmented, the titles of the two principal English-language histories emphasizing precisely the lack of national integration: *Colombia, Fragmented Country, Divided Society* by Bushnell and *The Making of Modern Colombia: A Nation in Spite of Itself* by Safford and Palacios. As Safford and Palacios (1998) note, "Provincial government remained effectively independent of the Audiencia [the local Spanish seat of control], and Santa Fe de Bogotá lacked formal authority of what is now western Colombia"(p. 55). This means that national conditions and institutions were often relatively less important than those local compared to other countries, and, as noted earlier, a variety of local institutional structures coexisted (see again García-Jimeno (2005); Bonet and Roca (2006)) and affected local development outcomes. Several regions employed both native and African slaves and evolved extractive institutions to manage them, others far less. Similarly, the Mita, Resguardo, and Encomienda are found to varying degrees in different departments. Independence saw several (repressed) attempts at regional succession, and the construction of a strong national state was effectively resisted.[15]

But even countries considered more centrally consolidated show a high degree of fragmentation with a somewhat tenative reach of the central institutions. To quote a 19th century observer from southern Chile:

[15] As an example, Bushnell (1993) recounts that the 1863 Constitution created nine United States of Colombia, but with far more restricted central power than was the case in the North American analogue. For instance, states issued their own stamps; the national government had responsibility only for "interoceanic" transport routes (that is, pertaining to the Panama railroad) thereby weakening any integrative national infrastructure project; and the upper house of the national Congress was called the Senate of Plenipotentiaries "as if its members were the emissaries of sovereign nations" (p. 122).

> Not so many years ago the inhabitants of the central region spoke of those in the South as distant and unknown, without the beneficial influence of civilization or the protection of the Government. Those in the upper altitudes considered those of the valleys as residents of another hemisphere, to which one could not travel except at great risk and financial sacrifice. Anyone who claimed that he had gone to Santiago or come from the South was treated as a dreamer or a liar; if he could prove with documents and irrefutable witnesses that he had made the journey, he was celebrated as if he had gone to China. This was because, to go to or leave Santiago, one had to arm and equip himself as though on an expedition.– Pedro Ruiz Aldea, *La Tarantula* (Concepcion) August 9, 1862, cited in Monteón (1982), page 3.

The seeming absence of a blanketing national "civilization" or [institutions of] government protection likely allow local analogues to emerge, and potentially permits the domination of the kind of subnational institutional dynamics the literature has identified. These could take the form of *de jure* regional differences, such as slavery in the North and South of the U.S. or less observable *de facto* power structures (see, among others, Acemoglu and Robinson (2006), García-Jimeno (2005), Naritomi et al. (2012)). If, in fact, differing concentrations of indigenous peoples appreciably affect institutional structures, it is not clear that the local effects should not be as or more powerful than the national. Michalopoulos and Papaioannou (2013), for instance, find this to be the case in Africa where pre-colonial tribal patterns are more correlated with present income than subsequent national institutions.

A complementary explanation is that effects relating to locational fundamentals or agglomeration effects emerge more strongly at lower levels of aggregation as suggested by Davis and Weinstein (2002). Indeed, studies that have tried to measure agglomeration economies carefully (Ciccone and Hall (1996), Duranton (2005), Ellison and Glaeser (1997) and Greenstone et al. (2010)) have all done so using highly disaggregated subnational data. Summerhill (2010) finds that in São Paulo Brazil, a "potentially coercive" colonial institution, the *aldeamento* that regulated indigenous populations is *positively* correlated with income per capita at the end of the twentieth century. He argues that there were both extractive and settler (effectively agglomeration effects), and the net effect was positive. It is possible that at this level the net effect yields persistence, while at the national level where local agglomeration and geographical effects are diluted, national institutions dominate. For example, for all the reasons discussed in the previous section, colonists settled where there were previous indigenous concentrations and greater agglomerations prospered relative to lesser. However, the institutions that emerged at the level of the nation may have reflected the extractive dimension of the agglomerations and have led, overall, to lower growth of the country.[16]

Finally, it is entirely possible that the link between national indigenous concentrations and institutions has been overdrawn and the reversal at that level, which we find present, but not significant in our data, is capturing something else besides extractive institutions. Certainly, the fact that Argentina, Chile and Uruguay are far closer to the US and Canada in aggregate density than to Perú or México, yet share Latin America's overall dissapointing growth trajectory suggests that there other factors at play in addition to institutions, or that the mapping from precolonial population density to institutionsis not so tight in the region. That said, Annex 1 also presents evidence, at the subnational level, for the negative institutional effects associated with pre-colonial agglomerations postulated by Acemoglu et al. (2002) at the country level. Using the share of slaves in the population as a proxy for extractive institutions, we find that regions with a higher incidence of slavery show both lower incomes and less (although still strongly positive and significant) persistence. Hence, persistence is likely to have been stronger were such institutions not a feature of the colonization. However, these effects do not appear strong enough to cause reversals at the subnational level.

In the next section, we first take a more careful historical look at individual low, medium and high density country cases to get a clearer view of how the different forces interacted to yield the patterns we see at the subnational level, and the two low density countries which offer our only negative evidence for reversals.

[16]Our thanks to Noam Yuchtman for suggesting this interpretation.

3.4.2 Low Density and Persistent: The US

Persistence holds strongly in the US. California, Massachusetts, and Rhode Island again, show the highest pre-colonial density and above average incomes. Among the mid-level pre-colonial density states, New Jersey, Connecticut, Delaware, are also among the richest, and Washington and Oregon are solidly above average. This mass of points on the two coasts drives the upward sloping relationship while a diffuse mass of largely southern and mountain states anchors the low pre-colonial density-low current income nexus.

As noted earlier, higher incomes plausibly find their roots both in the initial native agglomerations and locational fundamentals that attracted both native populations and Europeans. Both effects continued to be important across the centuries. For example, New York, Boston and Chicago have all played to their locational particularities, especially in transport, but they have also built on their strengths in accumulated human capital and information (see, for example Glaeser (2005)). There is also an argument for poor institutions driving the poorer regions, although perhaps more in line with Acemoglu et al. (2001). The adverse disease environment and climate of much of the South discouraged settlement and, in the end, colonization required the importing of African slaves.

The state most likely to capture the colonization-driven inversion dynamic might have been Mississippi since it incorporated the third largest native civilization in North America, was abused by the Spaniards, and is now the poorest state. However, the reversal of the state's fortune from a rich cotton center in the 19th century is likely more due to the institutional, demographic and education legacy of African slavery than the long vanished native population (We explore in a more structured fashion, the impact of slavery in Annex I). As Taylor (2001) notes, the Spanish conquistador Hernando de Soto arriving in the fertile Mississippi river valley in 1540-1542 was impressed by size of native populations, the expansive maize fields, the power of their chiefs to command large numbers of well trained warriors, even the pyramids, one of which was the third largest in North America after those of central México (The pyramid at Cahokia was near present day St. Louis.) De Soto died on the banks of the Mississippi, frustrated at finding no gold, and the Spaniards withdrew to México City, but not before widespread pillaging and infection decimated the native population. When the French returned a century later, only the Natchez people near present day Natchez, Mississippi remained in strength and organization. French encroachments on Natchez territories in 1729 led to massacres by the French and their Choctaw allies and dispersion and sale into slavery in the French West Indies of the surviving population. With the passage of another century, Natchez and Mississippi would emerge very prosperous at the height of the cotton boom.

Overall, for a moderate range of pre-colonial densities, the US suggests the persistence of economic activity.

3.4.3 Low density Reversals: Argentina and Chile

Argentina and Chile, provide the only two examples of statistically significant "reversals" (Figure 4). Hence, understanding the cause of their negative relationship is of particular interest. For Argentina, the evidence supports an idiosyncratic geographical fundamentals story rather than an institutional one. The richest areas in Figure 2-the Province of Buenos Aires, La Pampa, Córdoba, Santa Fe and Entre Rios surround Buenos Aires City-tend, in fact, to be in areas of low pre-colonial population density. The other richer departments, Santa Cruz and Chubut, are relatively undiversified mineral producers in relatively unattractive climates and hence show the "resource inversion" discussed earlier. At the other extreme, Corrientes and Misiones are relatively underdeveloped humid semi-tropical areas that were traditionally isolated and show the highest pre-colonial density and, hence, potentially extractive institutions. But it must be kept in mind that these densities map in both absolute and relative magnitude to those of Massachusetts and California within an overall distribution that, again, is remarkably similar that of the US. Hence, a theory of institution-driven inversion would need to explain why the endogenously emerging institutions would be so different from the US. In addition, Buenos Aires may well not have been such a paragon of inclusionary institutions that would account for its unusual growth. It was a major port of slave desembarcation in the New World and, in the last years of Spanish domination, it was 30% black (Andrews, 1980).[17]

It seems more likely that the present distribution of income arises largely from Buenos Aires' status as the principal Atlantic port of the Spanish empire. This was not always the case. Despite the evolution of the surrounding pampas economy, prior to the mid 18th century Buenos Aires was a backwater, surviving on smuggling contraband silver and slaves. This was largely due to the repression of natural locational advantage. By Spanish law, the production of silver and other products of the interior towns were directed over the Andes to Lima on the Pacific, where they were loaded on convoys passing through the Isthmus of Panama and then to Spain. The more logical route-through the Atlantic port of Buenos Aires, and then directly to Spain-was forbidden. However, largely for geostrategic reasons arising from the emergence of the North American colonies as a potential Atlantic power, the policy was reversed in 1776 when Spain established Buenos Aires as the capital of the new Viceroyalty of Rio de la Plata. Trade was now mandated through Buenos Aires and forbidden through Lima, leading to an abrupt reorientation of the country's economy away from the traditional interior towns, and towards the emerging coastal economy (Scobie, 1964). Hence, by royal fiat, locational fundamentals went from being repressed to dominant. From here, the agglomeration related externalities arising from becoming the dominant Atlantic port can explain the pattern we see.

Chile also shows a significant negative relationship between pre-colonial densities and present income but one which, again, does not appear driven by the institutional story for three reasons. First, several observations at the highest end of the country's relatively low

[17]As a final point, Ades and Glaeser (1995) note argue that industry did not play a prominent role in the rise of Buenos Aires so that a case for it being more suited to the second wave of the Industrial Revolution seems unlikely. Even by 1914, only 15 percent of the labor force was in manufacturing and the government displayed "hostility toward manufacturing and innovation" p 221.

density (4.7 per square kilometer)-Bio Bio, Maule, O'Higgins, Los Lagos, and Araucania-are among the poorest. However, these form a contiguous region, with the area below the Bio Bio River that includes them dominated by the Mapuche Indians and conquered only very late in the 19th century and hence never experienced Spanish rule. More likely, the technological complementarities discussed earlier were at play- the institutional case is not as compelling, perhaps, as one stressing the costs of being out of the global technological loop. In fact, the eventual conquest had to wait for the Chileans to import recent advances in weaponry to which the Mapuches did not have access. The capital, Santiago, offers a counter example of a regression discontinuity flavor: it has the same density and is contiguous to this region, but it was conquered several hundred years earlier and is much more prosperous. Second, the country is one of extremes with extractive industries in some of the driest and coldest areas of the planet which were not attractive to native populations. This implies a relatively uninteresting correlation of relatively low pre-colonial densities, and moderately high incomes (for a very few people) today. Finally, as in Argentina, the fact that the overall density is roughly equivalent to that of Canada raises the question of why such effects could be so so much stronger in Chile.

3.4.4 Medium Density and Persistent: Colombia

Colombia, a middle density country, is an important case for understanding the relative import of the different forces for and against persistence and, in particular, that of extractive institutions (see Acemoglu et al. (2012); García-Jimeno (2005)). First, though it is not among the countries with the highest pre-colonial density, it is a classic example of Spanish conquest with the usual attendant institutions. Hence, while we might argue that something about Anglo or French colonists led to different colonizer-native dynamics, this would not be the case in Colombia. Second, as noted above, it is a perhaps the case which offers the greatest independence of subnational observations from national conditions and institutions. As noted earlier, the country is highly geographically fragmented and its regions have shown a fierce autonomy, long resisting centrally imposed rule. Yet, despite the relative strength of local instiutions, Colombia still shows one of the cleanest examples of persistence in the sample (Figure 4). Not only the capital, but other areas of high pre-colonial density-Valle de Cauca, Santander, and Antioquia-have among the highest present day incomes. Hence, again, local agglomeration and locational forces appear to be dominant.

One particular reversal within the country is illustrative of the relative strength of locational fundamentals in particular. Although understated in the figures, Cauca department and its principal city Popayán fell from one of the two most important regions in Colombia–a major provider of early Colombian presidents and possessor of one of the country's two mints–to one of the poorer regions. The Spaniards favored it for the availability of indigenous labor to extract its mineral wealth, and its subsequent use of imported African slaves defined its culture in fundamental ways. However, the city that it lost market share to, Cali, in Colombia's now second richest department, Valle de Cauca, had an indigenous population density 30% larger and only 10% fewer slaves per capita than Cauca. In fact, it had the largest number of slaves of any department in Colombia.[18] The period critical to the reversal appears to be 1878 to 1915 with the construction of the Pacific Railroad connecting Cali with Buenaventura, Colombia's largest Pacific port, and through the Panama Canal (finished in 1914) to the rest of the world, while Popayán remained relatively isolated (Safford and Palacios 2002). It is likely that the location of the railroad, while importantly dictated by Cali's proximity to the Cauca River, is partly due to political economy considerations. However, a story related to initial populations or slavery does not appear clearly. It seems more likely that a permanent shock to locational fundamentals altered the relative attractiveness of the two regions.[19]

Of interest is that Cali's new locational advantage did not also diminish Antioquia and

[18] According to the 1843 Census of Colombia, 7.1% of the population was slaves in Cauca and 6.4% in Valle; in 1851 4.7% and 4.3% respectively. Initial indigenous density was 7.1 and 9.2 respectively.

[19] A similar story is the rise and fall of Mompóx, Colombia. This affluent port in the Magdalena River saw its demise when the river shifted course, allowing the development of Magangué. Since then, this UNESCO World Heritage Site has virtually remained stuck in time.

the Bogotá/Cundinamarca agglomeration as industrial centers. In the colonial period, both had a locational advantage in terms of climate and soil suitable for agriculture, proximity to mineral reserves, and disease inhibiting altitude. Yet, none of these are important to explaining the overwhelming dominance of both areas in the manufacturing and service sectors, while the need to cross, especially in Bogotá's case, several mountain ranges to access world markets is a major drag on competitiveness. This is suggestive that, as as in the US case, agglomeration effects, in particular, the availability of talent and knowledge- are critical to the continued dominance of these areas.[20] In sum, then, in a country where local institutions were relatively important, agglomeration and locational fundamentals appear to dominate.

3.4.5 High Density and Persistent: El Salvador, Nicaragua, Perú and México

El Salvador, Nicaragua, Perú and México have among the highest densities in our sample and, despite presumably the strongest agglomeration and negative institutional effects, they show support for persistence. In figure 4, Both El Salvador and Nicaragua show clear and statistically significant positive slopes. México and Perú are the emblematic examples of the colonization of the New World and show the highest values. Though less clearly significant than the other two, both offer support for the importance of the forces of persistence, albeit contaminated by changes in locational fundamentals.

[20]The Bogotá/Cundinamarca agglomeration dominates the country in most modern services and manufactures. The capital city, Bogotá, has revealed comparative advantage (participation of sector in value added relative to country average greater than 1) in non-food manufacturing (12% of value added), commerce (14%), financial services (10%), real estate services (10%), services to firms (7%), air transport services (1%), few of which are tied to locational fundamentals. In these areas and industry in general, it is the largest single producer in the country. In particular, it accounts for 50% of all financial services. Emphatically, it has neither comparative advantage or much production in the agriculture (0%) or minerals (0%) sectors which first attracted the colonists. As capital city, it also shows a comparative advantage in public administration, but this is not dominant or unusually large (9% of value added relative to 7% on average for the country as a whole). The enveloping department of Cundinamarca, maintains a comparative advantage in agricultural production, but also in both agricultural and non agricultural manufacturing and is the third and fourth largest national producer respectively. The growth of Antioquia historically was driven by mining and then by coffee. It maintains a comparative advantage in both, but each accounts for roughly 1% of departmental value added. It has a comparative advantage in manufacturing (18%) and commerce (11.5%) and is the second largest producer of manufactures, commerce, and financial services after the Bogotá/Cundinamarca agglomeration.

For Perú, Figure 4 suggests that Lima, La Libertad, Ica and Piura all correspond to very high pre-colonial density areas that remain among the better off regions today. However, Lambayeque province undermines the statistical relationship by showing the highest density observation but below average current income. Lambayeque's decline appears largely driven by compounding natural disasters-negative locational fundamental shocks. In pre-colonial times, the region was a major center of the Chimor and then Inca cultures. The Spanish colonizers subsequently built a livestock industry on appropriated native land and irrigation systems, as in Tenochtitlán, taking advantage of the infrastructure and knowledge of the previous civilization. From 1650 to 1719, a dynamic sugar based hacienda economy emerged and generated numerous fortunes. However, after 1720, the economy collapsed into a century long period of stagnation. While this was partly due to competition from other Peruvian (including local native) and Caribbean producers, a plague of cane-eating rats in 1701 followed by two devastating floods in 1720 and 1728 constituted idiosyncratic but very long lived shocks which caused widespread foreclosures and the bankruptcy of the traditional producing class. Only in the late colonial period did the regional economy recover somewhat to a now average level income as the new owners shifted from sugar to livestock and tobacco (Ramirez, 1986).[21] Since the shocks driving Lambayeque's fate seem idiosyncratic and dropping the region causes Perú to join countries showing persistence with lower mean densities, Perú, should probably be seen as confirming persistence across a wide range of initial densities.

México appears to combine two distinct sets of growth dynamics that interact to obscure any clear relationship. The first is the persistence effect. The Mexican Federal District (city) is the highest density region in our sample and it is one of the richest regions in all of Latin America. Morelos, the second densest region in our sample, has above average income.

[21] Lambayeque did differ in its continued heavy reliance on Indian labor as competitor sugar growing areas shifted more toward African slaves, although it is not clear whether this should have generated more or less toxic extractive institutions.

Both suggest persistence in the most native intensive regions of the hemisphere. Tlaxcala, the third most dense area in México ranks among the lower levels of prosperity. However, it seems unlikely that we can attribute it to especially extractive institutions since, in exchange for being the principal allies of the Spaniards and sheltering them in a particularly dire moment in the conquest of Tenochtitlán, the Tlaxcalans were granted "perpetual exemption from tribute of any sort," a share of the spoils of conquest, and control of two bordering provinces, an agreement that was substantially respected for the duration of Spanish rule (see Marks, 1994, p. 188).[22] Among the very highest pre-colonial densities in our sample, agglomeration effects again appear dominant.

However, there is a second dynamic. The present high income of the low pre-colonial density states of Baja California Sur, Nuevo León, Baja California Norte, Chihuahua, Sonora, and Coahuila provide a strong countervailing "reversal" that offsets the persistence effects. The proximity of these states to the increasingly dynamic US border makes it difficult to disentangle the influence of various types from the North (proximity to markets, knowledge spillovers), where it was in large part an appendage of the US economy. At the time of the establishment of the border at the Rio Grande, it was linked by population flows and contraband; during the civil war, it was a significant Southern export outlet; and by the turn of the century, it had received substantial US investments in railroads and mining that gave the impetus to the development of capitalism in the North (Mora-Torres, 2001). For instance, US firms operated mines in the North for export to US foundries (e.g. Consolidated Kansas City Smelting in Chihuahua). The three large foundries that formed the basis for the future dynamism of the principal industrial city in northern México, Monterrey, Nuevo Leon (with spillovers to much of the north of México) were primarily oriented toward the US market, and the largest was established by the Guggenheim interests with US capital

[22] In fact it may have been the opportunities for adventurism in partnership with the Spaniards in other areas of the New World that diverted energies from the home region. Tlaxcalans aided the Spaniards in dominating conquered tribes moving North. The oldest church in the US, found in Santa Fe, New México was constructed by Tlaxcalan artisans.

(Morado, 2003).[23] As Marichal (1997) notes, the emerging industry in these areas gave impetus to a set of *de facto* and eventually *de jure* institutions and pro-industry regulations which may well have only been able to emerge in an environment where the regulatory structure had not been driven by extractive considerations. That said, the fact that a positive correlation emerges when we abstract from the border states causes us to think that the proximity to the US was the primary driver of the prosperity of the low density North.

4 Conclusion

This paper documents that, within countries, economic activity in the Western Hemisphere has tended to persist over the last half millennium and in spite of the major trauma that was the colonization. Despite the literal decimation of the previous populations, the imposition of new cultural and political forms and technologies, areas that were rich and populated before the arrival of Colombus tend to be rich today. In this sense, in the same way that Davis and Weinstein (2002) demonstrated the resiliency of Hiroshima to a 25% fall in population, we show that the distribution of Native Indian Populations strongly influences prosperity today, despite their 90% reduction in their population and the imposition of cultural, political institutions which have been argued to do the reverse.

We construct a data set on subnational population densities and incomes derived from poverty maps, and show that pre-colonial population densities are strongly positively correlated with present day population and with income per capita. This is clearly the case for low pre-colonial density countries like the US, but also for classic Latin conquest cases like Colombia, and, for the extreme high density cases like El Salvador, México, Nicaragua

[23] As Mora-Torres (2001) notes, these foundries emerged largely as a result of the McKinley tariffs of 1890, which taxed foreign imports at roughly 50 percent. This threatened both Mexican exports of ore to the US, as well as the smelters on the US side that processed them. The response was to move the smelters over the border to the railway center of Monterrey. The result of the accumulated US capital investment was "that the northern economy became an extension of the U.S. economy and that the North turned into the new center of Mexican capitalism" p. 9.

and Perú. We also generate proxies for suitability for agriculture and river density that contribute to a comprehensive set of locational fundamentals. These appear as significant determinants of the location of pre-colonial densities and of our present day measures, but they do not appear to drive the observed persistence. Including locational fundamentals lowers the contribution of initial density only slightly in some population specifications, and actually increases it in the preferred income specification suggesting that agglomeration effects are dominant on average. In fact, we find that whereas a favorable agricultural evironment was, expectedly, positively correlated with pre-colonial populations, the reverse is true for present income. That said, across all our case studies, the large changes in relative positions, such as in Popayán (Colombia), Lambayeque (Perú), Buenos Aires, (Argentina), or the North of México appear largely driven by shifts in locational fundamentals.

Further, the case studies suggest reasons for both fundamentals and pre-colonial densities to be important. Not only would colonizers also value the rivers, coasts, fertile land, natural resources, and climate that attracted the native populations, but they would need the native populations themselves as sources of human capital (architects, agronomists, and craftsmen), trading partners, sources of information, strategic bulwarks against enemy encroachment, and souls to save. Hence, scale economies and Marshallian externalities related to population were probably as relevant to determining where colonists located their settlements as locational fundamentals. In turn, the contact with new technologies may have, after the initial trauma, strengthened these agglomerations. Many of the regions of the very highest pre-colonial density remain among the most prosperous regions today: the persistent prosperity of California and New England, in the US, or Antioquia, or Bogotá in Colombia, despite massive structural transformations away from natural resource based production toward more sophisticated manufacturing and services does suggest that the forces arising from concentrations of knowledge, trade or labor are critical. At the subnational level, geographical and especially agglomeration factors appear to cause fortune to persist.

5 Annex I. The Institutional Channel: Slavery in Brazil, Colombia, and the US

Though the paper documents the relative importance of locational fundamentals, agglomeration externalities and perhaps technological transfer in determining present income, we also find evidence for the negative impact of extractive institutions, even if they did not dominate the others.[24] As a proxy for extractive institutions we are able to collect data on the incidence of slavery at the subnational level for Brazil and Colombia and the US where censuses are available. As a direct measure of extractive institutions, we exploit the data on slavery, measured as the percentage of enslaved and "free colored people," in the three countries for which they are national historic censuses. For Brazil, we used the 1872 Census, for Colombia the 1851 Census and for the United States, we used the 1860 Census as well as the data compiled in Nunn (2008). [25] To capture the broader influence of slavery, both in the year of the census and in previous years, we include both slaves and the general black population which would include now-freed slaves.

While data comparability and classification issues are non-trivial, the average share of the population enslaved in the mid 19th century was 28% in the American South, 13% in Brazil and 2.9% in Colombia. We use the more expansive measure that includes free Blacks which raises Brazil to first position, although the results do not change qualitatively when we use the more narrow measure.

$$Log(Y_{2005,ij}) = \alpha + \beta g(D_{precol,ij}) + \delta SLAVERY_{ij} + \delta_{int} SLAVERY_{ij} * g(D_{precol_ij}) + \gamma LF_{ij} + \mu_i + \epsilon_{ij} \quad (3)$$

where δ_{int} captures the interaction of pre-colonial density and slavery and μ_i are now three fixed effects for Brazil, Colombia and the US South with the US North as the omitted category. Columns 1-5 in Table 7 progressively introduce the elements of equation 3. Column 1 includes pre-colonial density along with dummies for Brazil, Colombia and the American South.[26] In the full sample in both log-level and log-log specifications (column 1), pre-colonial density is significant and positive, lending support from a smaller sample to the case for persistence. Column 2 repeats the same regression with the smaller sample dictated by the more restrictive slavery variable with a loss in significance of the persistence term. Columns 3 add the slavery term and, for both log and level specifications, it enters negatively and significantly. Column 4 adds slavery interacted with initial population density. It

[24]The negative impact of slavery cannot be taken as a foregone conclusion since disentangling the endowment and institutional effect is difficult. Acemoglu et al. (2012) find that in Colombian municipalities where slave labor was demanded poverty is higher and school enrollment, vaccination coverage and public good provision is lower, than where it was not. On the other hand, in São Paulo, Brazil Summerhill (2010) finds no relationship between slavery and present incomes while Rocha et al. (2012) find slavery is positively correlated with human capital.

[25]We thank Jaime Bonet and Adolfo Meisel Roca for providing their colonial data for Colombia and for pointing us towards Tovar-Pinzón et al. (1994)'s compendium of colonial statistics.

[26]The South is comprised of Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, Missouri, North Carolina, South Carolina, Tennessee, Texas, Virginia, West Virginia.

enters negatively in both specifications and of similar sign although only significantly in the levels specification. Further, the coefficient on pre-colonial density roughly doubles with the inclusion of the interaction of slavery and density in the levels specification and increases by 30 percent in the log specification suggesting that extractive institutions did have a negative agglomeration effect as postulated by Acemoglu et al. (2002). The same results are found in the MS specifications (column 6).

Adding locational fundamentals (column 5) changes the coefficient little but renders it insignificant in the OLS FE estimates. However, the MS estimator finds both the free standing and interactive terms significant in both level and log specifications. Though the sample is small, nonetheless, the results offer support for extractive institutions at least reducing, if not overturning the persistence induced by agglomeration externalities and fundamentals.

In sum, despite legitimate concerns about the exogeneity of slavery, using a direct proxy for exclusionary institutions is suggestive of the existence of an Acemoglu et al. (2002) effect: slavery works against persistence. However, at the subnational level, the net effect of factors associated with of indigenous population densities– extractive institutions, agglomeration externalities, or locational fundamentals–tends to leave a positive correlation with prosperity today.

References

Acemoglu, D. and Dell, M. (2010). Productivity Differences Between and Within Countries. *American Economic Journal: Macroeconomics*, 2(1):169–88.

Acemoglu, D., García-Jimeno, C., and Robinson, J. A. (2012). Finding Eldorado: Slavery and Long-run Development in Colombia. Working Paper 18177, National Bureau of Economic Research.

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

Acemoglu, D., Johnson, S., and Robinson, J. A. (2002). Reversal of Fortune: Geography and Institutions in The Making of The Modern World Income Distribution. *The Quarterly Journal of Economics*, 117(4):1231–1294.

Acemoglu, D. and Robinson, J. A. (2006). De Facto Political Power and Institutional Persistence. *The American Economic Review*, 96(2):pp. 325–330.

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

Andrews, G. R. (1980). *The Afro-Argentines of Buenos Aires 1800-1900.* University of Winsconsin Press. Madison.

Aroca, P., Bosch, M., and F., W. (2005). Spatial Dimensions of Trade Liberalization and Economic Convergence: Mexico 1985–2002. *World Bank Economic Review*, 19(3):345–378.

Ashraf, Q. and Galor, O. (2011a). Cultural Assimilation, Cultural Diffusion and the Origin of the Wealth of Nations. *mimeo.*

Ashraf, Q. and Galor, O. (2011b). Dynamics and Stagnation in the Malthusian Epoch: Theory and Evidence. *American Economic Review.*

Ashraf, Q. and Galor, O. (2012). The" Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development. *American Economic Review.*

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

Becker, G. S., Glaeser, E. L., and Murphy, K. M. (1999). Population and Economic Growth. *American Economic Review*, 89(2):145–149.

Bleakley, H. and Lin, J. (2012). Portage and Path Dependence. *Quarterly Journal of Economics.*

Bonet, J. and Roca, A. M. (2006). El Legado Colonial como Determinante del Ingreso per cápita Departamental en Colombia. Documentos de Trabajo sobre Economía Regional 002520, Banco de la República - Economía Regional.

Breen, T. H. (1984). Creative Adaptations: Peoples and Cultures. *Colonial British America.*

Bruhn, M. and Gallego, F. A. (2011). Good, Bad, and Ugly Colonial Activities : Studying Development Across the Americas. *Review of Economics and Statistics*.

Bushnell, D. (1993). *The Making of Modern Colombia: A Nation in Spite of Itself*. University of California Press.

Ciccone, A. and Hall, R. E. (1996). Productivity and the Density of Economic Activity. *American Economic Review*, 86(1):54–70.

Coatsworth, J. H. (2008). Inequality, Institutions and Economic Growth in Latin America. *Journal of Latin American Studies*, 40:545–569.

Comin, D., Easterly, W., and Gong, E. (2010). Was the Wealth of Nations Determined in 1000 BC? *American Economic Journal: Macroeconomics*, 2(3):65–97.

Davis, D. R. and Weinstein, D. (2002). Bones, Bombs, and Break Points: The Geography of Economic Activity. *American Economic Review*, 92(5):1269–1289.

De Vorsey, L. (1978). Amerindian Contributions to the Mapping of North America: A Preliminary View. *Imago Mundi*, 30:71–78.

De Vorsey Jr, L. (1986). The Colonial Georgia Backcountry. *Colonial Augusta:"Key of the Indian Countrey"(Macon, Ga., 1986)*, pages 3–26.

Degler, C. N. (1970). *Out of Our Past : The Forces that Shaped Modern America*. Harper & Row, New York.

Dell, M. (2011). The Persistent Effects of Peru's Mining Mita. *Econometrica*. Forthcoming.

Denevan, W. (1992a). *The Native Population of the Americas in 1492*. University of Wisconsin Press.

Denevan, W. (1992b). The Pristine Myth: the Landscape of the Americas in 1492. *Annals of the Association of American Geographers*, 82(3):369–385.

Duranton, G. (2005). Testing for Localization Using Micro-Geographic Data. *Review of Economic Studies*, 72(4):1077–1106.

Easterly, W. and Levine, R. (2003). Tropics, Germs, and Crops: How Endowments Influence Economic Development. *Journal of Monetary Economics*, 50(1):3–39.

Elbers, C., Lanjouw, J. O., and Lanjouw, P. F. (2003). Micro–Level Estimation of Poverty and Inequality. *Econometrica*, 71(1):355–364.

Ellison, G. and Glaeser, E. L. (1997). Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy*, 105(5):889–927.

Ellison, G. and Glaeser, E. L. (1999). The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration? *American Economic Review*, 89(2):311–316.

Fujita, M. and 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.

Gallup, J. L., Sachs, J. D., and Mellinger, A. D. (1998). Geography and Economic Development. Technical report.

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

Galor, O. and Weil, D. N. (1999). From malthusian stagnation to modern growth. *American Economic Review*, 89(2):150–154.

García-Jimeno, C. (2005). Colonial Institutions and Long-Run Economic Performance in Colombia: Is There Evidence of Persistence? *Documento CEDE*.

Gennaioli, N, L. P., Lopez-de Silanes, F., and Shleifer, A. (2011). Human Capital and Regional Development. *mimeo*.

Glaeser, E. L. (2005). Urban Colossus: Why is New York America's Largest City? NBER Working Papers 11398, National Bureau of Economic Research, Inc.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., and Shleifer, A. (1992). Growth in Cities. *Journal of Political Economy*, 100(6):1126–52.

Greenstone, M., Hornbeck, R., and Moretti, E. (2010). Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. *Journal of Political Economy*, 118(3):536–598.

Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy*, 99(3):483–99.

Krugman, P. (1992). *Geography and Trade*, volume 1 of *MIT Press Books*. The MIT Press.

Krugman, P. and Elizondo, R. L. (1996). Trade Policy and the Third World Metropolis. *Journal of Development Economics*, 49(1):137–150.

Krugman, P. R. (1993). On the Relationship between Trade Theory and Location Theory. *Review of International Economics*, 1(2):110–22.

Lucas, R. E. (2004). The Industrial Revolution: Past and Future. *Annual Report*, (May):5–20.

Marichal, C. (1997). Avances Recientes en la Historia de las Grandes Empresas y su Importancia para la Historia Económica de México. In Marichal, C. and Cerutti, M., editors, *Historia de las grandes empresas en México, 1850-1930*. Universidad Autónoma de Nuevo León - Fondo de Cultura Económica.

Marks, R. L. (1994). *Cortés: The Great Adventurer and the Fate of Aztec Mexico*. Alfred A. Knopf.

Maronna, R. and Yohai, V. (2000). Robust Regression with Both Continuous and Categorical Predictors. *Journal of Statistical Planning and Inference*, 89(1-2):197–214.

McEvedy, C., Jones, R., et al. (1978). *Atlas of world population history*. Penguin Books Ltd, Harmondsworth, Middlesex, England.

Michalopoulos, S. and Papaioannou, E. (2013). Pre-colonial ethnic institutions and contemporary african development. *Econometrica*, 81(1):113–152.

Miguel, E. and Roland, G. (2011). The Long Run Impact of Bombing Vietnam. *Journal of Development Economics*, 96(1):1–15.

Mitchener, K. J. and McLean, I. W. (2003). The Productivity of US States since 1880. *Journal of Economic Growth*, 8(1):73–114.

Monteón, M. (1982). *Chile in the Nitrate Era: The Evolution of Economic Dependence, 1880-1930*. University of Wisconsin Press Madison, Wisc.

Mora-Torres, J. (2001). *The Making of the Mexican Border*. University of Texas Press.

Morado, C. (2003). Empresas Mineras y Metalurgicas en Monterrey, México. 1890-1908. Parte ii. Tres Plantas Metalúrgicas. Technical report.

Naritomi, J., Soares, R. R., and Assunï¿½ï¿½o, J. J. (2012). Institutional Development and Colonial Heritage within Brazil. *Journal of Economic History*, 72(1):393–422.

Newson, L. (1982). The Depopulation of Nicaragua in the Sixteenth Century. *Journal of Latin American Studies*, 14(2):235–286.

Nunn, N. (2008). Slavery, Inequality, and Economic Development in the Americas: An Examination of the Engerman-Sokoloff Hypothesis. In Helpman, E., editor, *Institutions and Economic Performance*, pages 148–180. Harvard University Press.

Parkes, H. B. (1969). *A History of Mexico*. Houghton Mifflin.

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

Ramirez, S. E. (1986). *Patriarchs: Land Tenure and the Economics of Power in Colonial Perú*. University of New Mexico Press, Albuquerque, 1st edition.

Rappaport, J. and Sachs, J. D. (2003). The United States as a Coastal Nation. *Journal of Economic Growth*, 8(1):5–46.

Rocha, R., Ferraz, C., and Soares, R. (2012). Settlement Colonies Across Plantation Fields: Evidence on the Relationship Between Human Capital and Long Term Development. mimeo, PUC-Rio.

Safford, F. and Palacios, M. (1998). *Colombia: Fragmented Land, Divided Society*. 1st ed. edition.

Scobie, J. R. (1964). *Argentina : A City and a Nation*. Oxford University Press, New York.

Severnini, E. R. (2012). The power of hydroelectric dams: Agglomeration spillovers.

Simpson, L. (1962). *Many Mexicos*. Univ of California Press.

Steckel, R. H. and Prince, J. M. (2001). Tallest in the World: Native Americans of the Great Plains in the Nineteenth Century. *The American Economic Review*, 91(1):pp. 287–294.

Summerhill, W. (2010). Colonial Institutions, Slavery, Inequality, and Development: Evidence from São Paulo, Brazil. MPRA Paper 22162, University Library of Munich, Germany.

Taylor, A. (2001). *American Colonies: The Settling of North America the Penguin History of the United States, Volume 1*. Penguin History of the United States. Penguin Books, New York, N.Y.

Thomas, H. (1993). *Conquest: Montezuma, Cortés, and the Fall of Old México*. Touchstone.

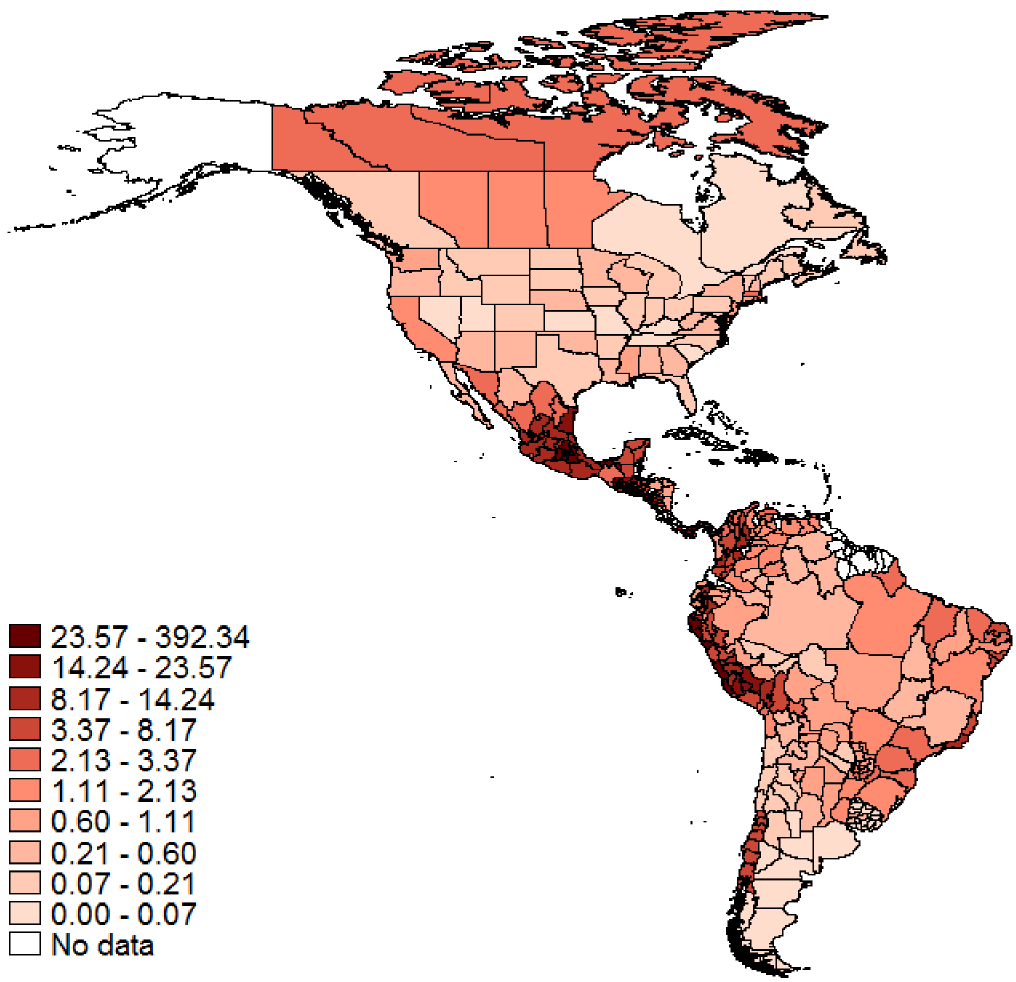
Tovar-Pinzón, H., Tovar-Mora, J. A., and Tovar-Mora, C. E. (1994). *Convocatoria al Poder del Número: Censos y Estadísticas de la Nueva Granada, 1750-1830*. Archivo General de la Nación (Colombia).

Ubelaker, D. (1988). North American Indian Population Size, AD 1500 to 1985. *American Journal of Physical Anthropology*, 77(3):289–294.

World Bank (2008). *World Development Report 2009: Reshaping Economic Geography*.

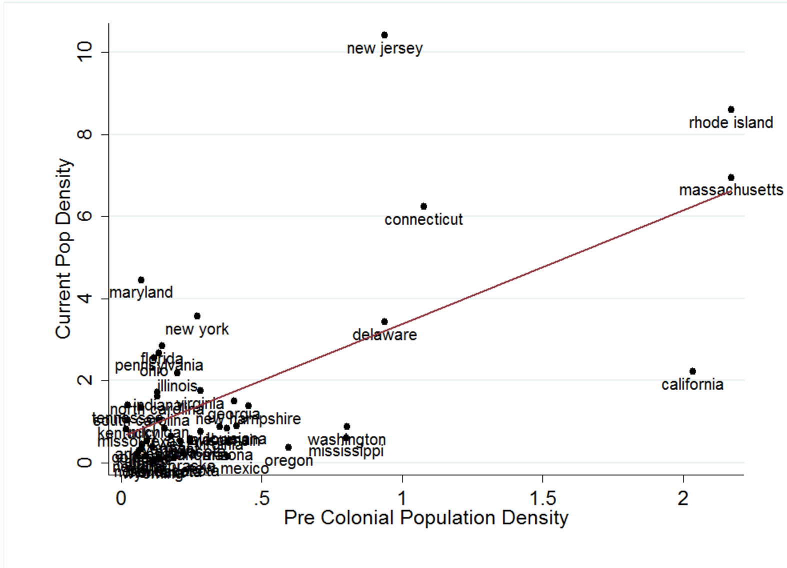
Tables and Figures

Figure 1: Pre-colonial Population Density



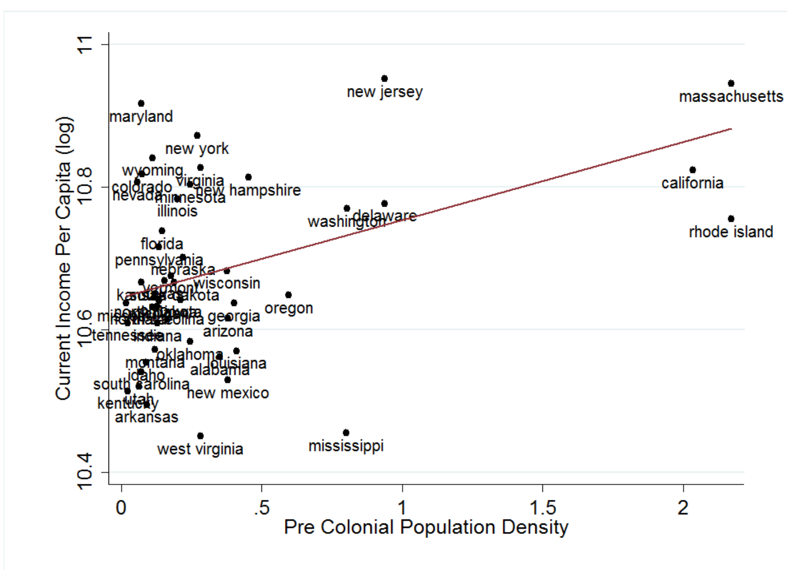
Note: Pre-colonial Population Density is the number of indigenous people per square kilometer before the arrival of Columbus, Income is per capita (PPP 2005 US dollars) in 2000. Data from national censuses, Denevan (1992), and Bruhn and Gallego (2010). More detailed data sources and descriptions in the text.

Figure 2: Population Density in 2000 against Pre-colonial Population Density (United States)



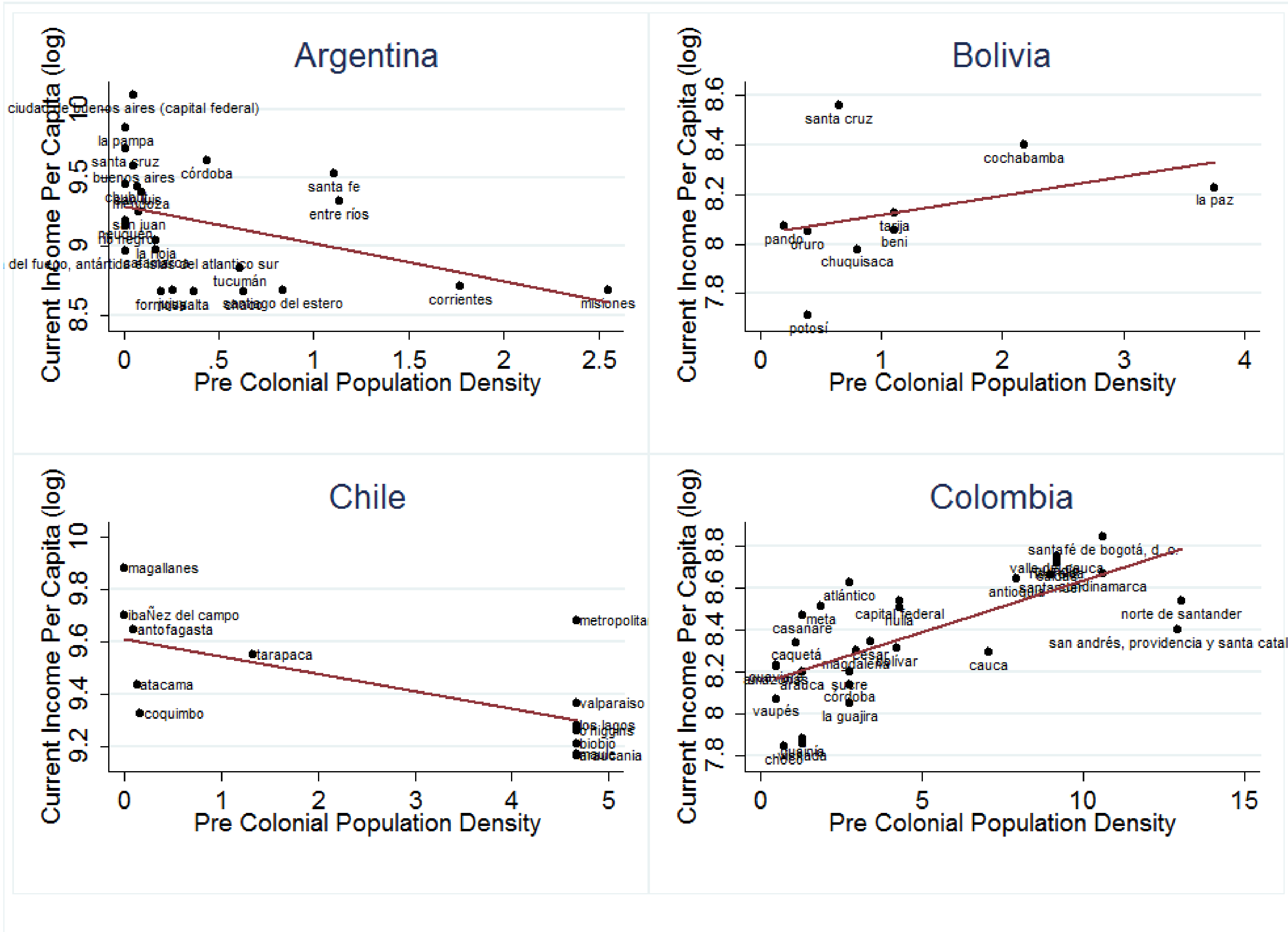
Note: Pre-colonial Population Density is the number of indigenous people per square kilometer before the arrival of Columbus, Current Population Density is the total population in 2000 divided by the area of the state or province in square kilometers. Data from national censuses, Denevan (1992), and Bruhn and Gallego (2010). More detailed data sources and descriptions in the text.

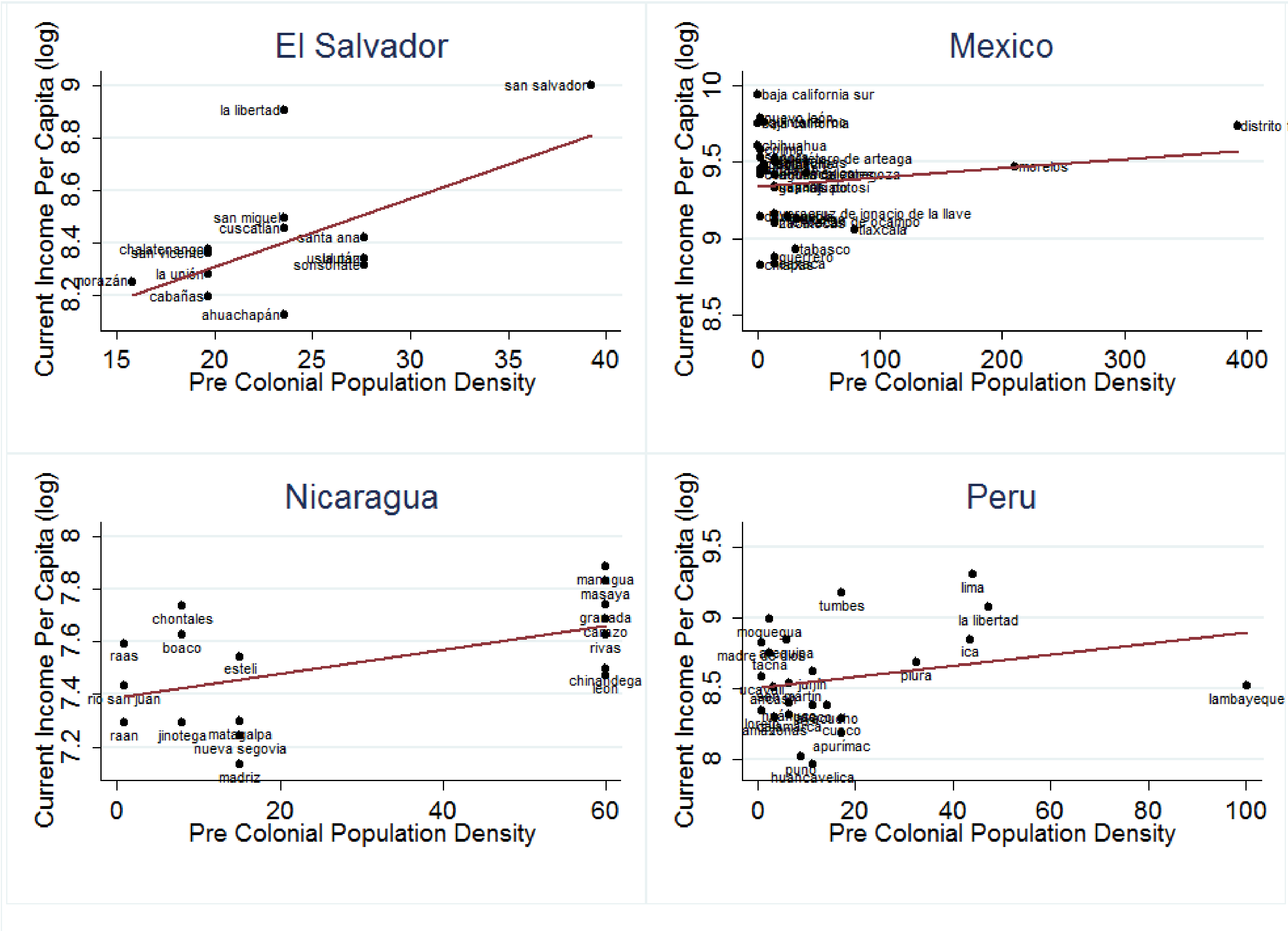
Figure 3: Log Income per Capita in 2005 against Pre-colonial Population Density (United States)



Note: Pre-colonial Population Density is the number of indigenous people per square kilometer before the arrival of Columbus, Income is per capita (PPP 2005 US dollars) in 2000. Data from national censuses, Denevan (1992), and Bruhn and Gallego (2010). More detailed data sources and descriptions in the text.

Figure 4: Log Income per Capita in 2005 against Pre-colonial Population Density (Latin America)





Note: Pre-colonial Population Density is the number of indigenous people per square kilometer before the arrival of Columbus, Income is per capita (PPP 2005 US dollars) in 2000. Data from national censuses, Denevan (1992), and Bruhn and Gallego (2010). More detailed data sources and descriptions in the text.

Table 1: Summary Statistics- Population Density and Income

	Obs	**Pre-colonial Population Density**				**Current Population Density**				**Income**			
		Mean	Coef. Var	Min	Max	Mean	Coef. Var	Min	Max	Mean	Coef. Var	Min	Max
Argentina	24	0.44	1.45	0.01	2.55	626.06	4.80	1.20	14727.03	10576.16	0.46	5834.35	24328.34
Bolivia	9	1.18	0.96	0.20	3.74	9.53	0.84	0.82	26.17	3494.36	0.25	2239.15	5219.44
Brazil	27	2.55	0.97	0.20	8.58	53.39	1.40	1.41	346.75	7590.93	0.46	3343.24	18287.33
Canada	13	1.22	1.06	0.02	3.00	6.34	1.19	0.01	24.40	34540.71	0.17	27479.80	48436.04
Chile	13	2.65	0.87	0.01	4.66	53.05	1.99	1.05	393.50	12852.48	0.24	9545.53	19533.39
Colombia	30	4.96	0.82	0.49	13.04	424.40	2.36	0.48	4310.09	4554.56	0.27	2546.91	6917.57
Ecuador	22	5.76	0.78	0.01	12.06	56.10	0.92	2.01	182.80	5764.57	0.30	3738.26	10463.96
El Salvador	14	24.19	0.24	15.80	39.25	326.73	1.30	95.58	1768.80	4669.67	0.29	3378.47	8094.27
Guatemala	8	22.95	0.35	5.64	29.08	248.97	1.57	10.23	1195.48	3699.73	0.56	2132.71	8526.96
Honduras	18	8.09	0.55	1.00	17.64	134.67	1.22	15.81	614.83	3171.35	0.30	1512.21	5170.91
Mexico	32	31.90	2.38	0.40	392.34	227.55	3.36	5.61	4352.62	12119.95	0.29	6780.40	20709.32
Nicaragua	17	29.82	0.89	1.00	60.00	103.28	1.20	8.58	473.80	1896.24	0.22	1250.37	2658.39
Panama	9	13.40	0.67	0.06	24.78	38.66	0.88	2.42	116.80	9046.41	0.31	4880.31	13950.97
Paraguay	18	1.27	0.56	0.20	3.29	58.62	2.28	0.10	579.36	4162.39	0.18	2923.94	5516.21
Peru	24	17.36	1.30	0.78	100.15	31.80	0.18	1.08	222.23	5623.75	0.35	2846.11	10980.10
US	48	0.39	1.34	0.02	2.17	169.50	0.99	5.16	1041.54	44193.13	0.14	34533.35	62765.91
Uruguay	19	0.11	2.05	0.00	0.85	33.44	1.80	2.25	263.51	8195.26	0.21	6024.20	13965.81
Venezuela	19	1.78	0.42	0.35	2.78	96.70	0.48	0.40	415.52	9788.84	0.13	7843.90	13191.90

Note: Pre-colonial Population Density is the number of indigenous people per square kilometer before the arrival of Columbus, Current Population Density is the total population in 2000 divided by the area of the state or province in square kilometers and Income is in per capita (PPP 2005 US dollars) in 2000. Data from national censuses, Denevan (1992), and Bruhn and Gallego (2010). More detailed data sources and descriptions in the text.

Table 2: Summary Statistics- Population Density and Income

	Mean	**Median**	**Sd**	**Min**	**Max**
Agriculture	0.56	0.58	0.28	0.00	1.00
Rivers	3.28	3.29	1.23	0.00	6.92
Landlocked	0.57	1.00	0.50	0.00	1.00
Temperature	19.97	20.40	5.83	2.38	29.00
Altitude	0.66	0.19	0.92	0.00	4.33
Rainfall	1.28	1.10	0.95	0.00	8.13

Note: Agriculture is an index of probability of cultivation given cultivable land, climate and soil composition, from Ramankutty, Foley and McSweeney (2002). Rivers captures the density of rivers as a share of land area derived from HydroSHEDS (USGS 2011). Landlocked is a dummy variable for whether the state has access to a coast or not; temperature is a yearly average in ^{o}C; altitude measures the elevation of the capital city of the state in kilometers; and Rainfall captures total yearly rainfall in meters, all are from Bruhn and Gallego (2010). More detailed data sources and descriptions in the text.

Table 3: Pre-colonial Population Density and Locational Fundamentals (pooled)

	(1) OLS	(2) OLS	(3) OLS FE	(4) OLS FE	(5) MS	(6) MS	(7) MS FE	(8) MS FE
Agriculture	0.1	-0.04	0.06	-0.02	0.005**	0.03**	0.006***	0.02**
	(0.08)	(0.11)	(0.06)	(0.11)	(0.00)	(0.01)	(0.00)	(0.01)
Agriculture^2		0.1		0.08		-0.02*		-0.01*
		(0.12)		(0.15)		(0.01)		(0.01)
Rivers	-0.03***	-0.03	-0.002	0.02	0.0004	-0.0005	-0.002***	-0.002
	(0.01)	(0.03)	(0.01)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)
Rivers^2		0.0008		-0.002		-0.00005		-0.00009
		(0.00)		(0.00)		(0.00)		(0.00)
Landlocked	0.01	-0.005	0.02	0.02	-0.007***	-0.005***	-0.001	-0.0007
	(0.03)	(0.02)	(0.03)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)
Temperature	0.004*	0.01	-0.003	0.01	0.0008***	-0.0009*	0.0002*	0.00009
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Temperature^2		-0.0002		-0.0004		0.00005***		0.000002
		(0.00)		(0.00)		(0.00)		(0.00)
Altitude	0.06	0.1	0.02	0.02	0.004***	0.005**	0.0006	0.0005
	(0.04)	(0.08)	(0.04)	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)
Altitude^2		-0.02		-0.003		-0.0006		0.00003
		(0.02)		(0.00)		(0.00)		(0.00)
Rainfall	-0.01	-0.02	-0.01	-0.03	-0.0005	-0.0002	0.00003	0.001
	(0.01)	(0.03)	(0.01)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)
Rainfall^2		0.002		0.003		-0.0002		-0.0004**
		(0.00)		(0.00)		(0.00)		(0.00)
Constant	0.02	-0.02	0.02	-0.10	-0.009***	0.002	0.006**	0.003
	(0.06)	(0.06)	(0.09)	(0.09)	(0.00)	(0.01)	(0.00)	(0.00)
N	330	330	330	330	330	330	330	330
R^2	0.061	0.058	0.109	0.099				

Note: Regression of sub national Pre-colonial Population Density on locational fundamentals. Estimation by OLS and robust MS regression with country fixed effects. Pre-colonial Population Density is the number of indigenous people per square kilometer before the arrival of Columbus, from Denevan (1992), and Bruhn and Gallego (2010). Agriculture is an index of probability of cultivation given cultivable land, climate and soil composition, from Ramankutty, Foley and McSweeney (2002). Rivers captures the density of rivers as a share of land area derived from HydroSHEDS (USGS 2011). Landlocked is a dummy variable for whether the state has access to a coast or not; temperature is a yearly average in ^{o}C; altitude measures the elevation of the capital city of the state in kilometers; and Rainfall captures total yearly rainfall in meters, all are from Bruhn and Gallego (2010). More detailed data sources and descriptions in the text. Robust SE for OLS and MS SE are in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Population Density in 2000 and Pre-colonial Population Density (country by country)

	N	β Pop. Density		Rank
		Log-Log	Level-Level	Correlation
Argentina	24	0.29**	-602.5	0.61***
		(0.12)	(637.67)	
Brazil	27	0.81***	7.34	0.63***
		(0.20)	(5.29)	
Bolivia	9	0.85***	5.16**	0.68**
		(0.28)	(2.02)	
Chile	13	0.61***	19.9*	0.84***
		(0.08)	(11.47)	
Canada	13	-0.86***	-3.70***	-0.69**
		(0.28)	(1.17)	
Colombia	30	1.64***	108.1**	0.70***
		(0.32)	(51.20)	
Ecuador	22	0.50***	3.84*	0.49**
		(0.10)	(2.24)	
El Salvador	14	2.79**	61.4***	0.79**
		(0.67)	(21.40)	
Guatemala	8	1.98***	20.9	0.83**
		(0.38)	(14.85)	
Honduras	18	0.79***	21.2**	0.47**
		(0.13)	(10.11)	
Mexico	32	0.65***	9.09***	0.68***
		(0.12)	(2.32)	
Mexico[1]	25	0.80***	9.18***	0.65***
		(0.17)	(2.34)	
Nicaragua	17	0.67***	2.99***	0.80***
		(0.08)	(1.10)	
Panama	9	0.034	-0.31	0.08
		(0.14)	(1.11)	
Paraguay	18	1.37***	-12.3	0.34
		(0.51)	(22.14)	
Peru	24	0.70***	1.11**	0.74 ***
		(0.11)	(0.52)	
Peru[2]	23	0.73***	2.00*	0.70***
		(0.12)	(1.05)	
US	48	0.44***	276.9***	0.37***
		(0.15)	(71.65)	
Uruguay	19	-0.16	-41.3	-0.25
		(0.13)	(35.27)	
Venezuela	19	0.70***	1.11**	0.76***
		(0.11)	(0.52)	

Note: Beta from OLS regression of Current Population Density on Pre-colonial Population Density in both Log-Log and Level-Level forms. Final column is Spearman rank correlation coefficient. Current Population Density is the log of the total population in 2000 divided by the area of the state or province in square kilometers, from national censuses, and Bruhn and Gallego (2010). Pre-colonial Population Density is the log of the number of indigenous people per square kilometer before the arrival of Columbus, from Denevan (1992), and Bruhn and Gallego (2010). More detailed data sources and descriptions in the text. 1. Mexico without border states. 2 Peru without Lambayeque. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Population Density in 2000, Pre-colonial Population Density, and Locational Fundamentals (pooled)

	(1) OLS	(2) Between	(3) Within FE	(4) Within FE	(5) Within FE	(6) MS FE	(7) MS FE
Pre-colonial Density	7.2***	1.9	8.6***	8.8***	8.8***	3.0***	0.6***
	(1.37)	(3.80)	(0.73)	(0.57)	(0.68)	(0.11)	(0.13)
Agriculture					-4.3		0.3
					(8.42)		(0.41)
Agriculture^2					8.4		0.1
					(11.90)		(0.48)
Rivers					0.2		0.10
					(2.89)		(0.11)
Rivers^2					-0.2		-0.02*
					(0.39)		(0.01)
Landlocked					-1.6		-0.1**
					(2.03)		(0.05)
Temperature					0.6		0.03
					(0.38)		(0.02)
Temperature^2					-0.02*		-0.0009
					(0.01)		(0.00)
Altitude					-0.8		-0.04
					(1.28)		(0.06)
Altitude^2					0.3		0.010
					(0.40)		(0.02)
Rainfall					2.0*		0.10*
					(1.12)		(0.05)
Rainfall^2					-0.3**		-0.04***
					(0.14)		(0.01)
Constant	1.1*	1.3**	0.9***	1.0***	-4.2	0.07***	-0.3***
	(0.53)	(0.54)	(0.07)	(0.05)	(7.31)	(0.01)	(0.11)
N	365	365	365	330	330	330	330
R^2	0.045	-0.045	0.057	0.060	0.068		

Note: Regression of Current Population Density against Pre-colonial Population Density. Estimation by OLS and robust MS regression with country fixed effects. Current Population Density is the total population in 2000 divided by the area of the state or province in square kilometers, from national censuses, and Bruhn and Gallego (2010). Pre-colonial Population Density is the number of indigenous people per square kilometer before the arrival of Columbus, from Denevan (1992), and Bruhn and Gallego (2010). Agriculture is an index of probability of cultivation given cultivable land, climate and soil composition, from Ramankutty, Foley and McSweeney (2002). Rivers captures the density of rivers as a share of land area derived from HydroSHEDS (USGS 2011). Landlocked is a dummy variable for whether the state has access to a coast or not; temperature is a yearly average in ^{o}C; altitude measures the elevation of the capital city of the state in kilometers; and Rainfall captures total yearly rainfall in meters, all are from Bruhn and Gallego (2010). More detailed data sources and descriptions in the text. Robust SE for OLS and MS SE are in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Log Income per Capita in 2005, Pre-colonial Population Density, and Locational Fundamentals (pooled)

	(1) OLS	(2) Between	(3) Within FE	(4) Within FE	(5) Within FE	(6) MS FE	(7) MS FE
Pre-colonial Density	-0.4	-2.8	0.1**	0.09**	0.09***	0.06***	0.1***
	(0.58)	(1.70)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)
Agriculture					-0.4		0.3
					(0.23)		(0.30)
Agriculture^2					0.2		-0.3
					(0.24)		(0.31)
Rivers					-0.04		0.0003
					(0.10)		(0.04)
Rivers^2					-0.0004		-0.004
					(0.01)		(0.01)
Landlocked					0.01		0.01
					(0.05)		(0.03)
Temperature					0.02		0.006
					(0.02)		(0.01)
Temperature^2					-0.0008		-0.0003
					(0.00)		(0.00)
Altitude					-0.02		-0.1*
					(0.09)		(0.07)
Altitude^2					-0.03		-0.02
					(0.02)		(0.02)
Rainfall					-0.05		-0.2***
					(0.04)		(0.04)
Rainfall^2					-0.004		0.01***
					(0.01)		(0.00)
Constant	9.1***	9.1***	9.0***	9.0***	9.4***	9.5***	9.1***
	(0.28)	(0.24)	(0.00)	(0.00)	(0.21)	(0.17)	(0.14)
N	365	365	365	330	330	330	330
R^2	0.010	0.093	0.004	0.003	0.128		

Note: Regression of the Log of Income per capita in 2000 (PPP 2005 US dollars) against Pre-colonial Population Density. Estimation by OLS and robust MS regression with country fixed effects. Income per capita (in logs) is taken from national censuses. Pre-colonial Population Density is the number of indigenous people per square kilometer before the arrival of Columbus, from Denevan (1992), and Bruhn and Gallego (2010). Agriculture is an index of probability of cultivation given cultivable land, climate and soil composition, from Ramankutty, Foley and McSweeney (2002). Rivers captures the density of rivers as a share of land area derived from HydroSHEDS (USGS 2011). Landlocked is a dummy variable for whether the state has access to a coast or not; temperature is a yearly average in ^{o}C; altitude measures the elevation of the capital city of the state in kilometers; and Rainfall captures total yearly rainfall in meters, all are from Bruhn and Gallego (2010). More detailed data sources and descriptions in the text. Robust SE for OLS and MS SE are in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Log Income per Capita in 2005, Pre-colonial Population Density, and Slavery (Brazil, Colombia and United States)

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS	(6) MS	(7) MS
Pre-colonial Density	2.9**	1.9	2.6**	5.5***	2.5*	6.5***	8.6***
	(1.16)	(1.33)	(1.27)	(1.45)	(1.46)	(1.56)	(1.65)
Brazil	-1.9***	-2.0***	-1.6***	-1.6***	-1.5***	-2.0***	-2.1***
	(0.09)	(0.11)	(0.21)	(0.20)	(0.23)	(0.07)	(0.05)
Colombia	-2.5***	-2.4***	-2.4***	-2.6***	-2.3***	-2.6***	-2.9***
	(0.07)	(0.09)	(0.08)	(0.08)	(0.23)	(0.08)	(0.16)
South	-0.09**	-0.1***	0.2	0.09	0.09	-0.01	0.1***
	(0.04)	(0.04)	(0.13)	(0.13)	(0.14)	(0.07)	(0.03)
Slavery			-0.009**	-0.006	-0.006*	-0.002	-0.006***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Slavery*Pop				-0.1**	-0.09	-0.1***	-0.1***
				(0.05)	(0.05)	(0.04)	(0.02)
Agriculture					-0.08		0.5***
					(0.89)		(0.18)
Agriculture^2					-0.1		-0.4**
					(0.75)		(0.18)
Rivers					-0.002		0.09*
					(0.17)		(0.05)
Rivers^2					-0.009		-0.02***
					(0.02)		(0.01)
Landlocked					0.02		0.01
					(0.10)		(0.03)
Temperature					0.04		-0.01
					(0.03)		(0.01)
Temperature^2					-0.001		0.0006
					(0.00)		(0.00)
Altitude					0.3		0.4***
					(0.20)		(0.03)
Altitude^2					-0.1		-0.2***
					(0.08)		(0.01)
Rainfall					-0.004		0.1***
					(0.10)		(0.02)
Rainfall^2					-0.007		-0.02***
					(0.01)		(0.00)
Constant	10.7***	10.7***	10.7***	10.7***	10.7***	10.7***	10.4***
	(0.02)	(0.03)	(0.03)	(0.02)	(0.32)	(0.04)	(0.09)
N	105	78	78	78	78	78	78
R^2	0.937	0.940	0.947	0.953	0.953		

Note: Dependent variable is the Log Income per capita in 2000 (PPP 2005 US dollars). Pre-colonial Population Density is the number of indigenous people per square kilometer before the arrival of Columbus. Estimation by OLS and robust MS regression with country fixed effects. Income per capita (in logs) is taken from national censuses. Pre-colonial Population Density is the number of indigenous people per square kilometer before the arrival of Columbus, from Denevan (1992), and Bruhn and Gallego (2010). Dummies for Brazil, Colombia, and the US South (according to the US Census). Slavery is measured as a fraction of the population and is taken from Bergad (2008) and Nunn (2008). Interaction of slavery with Pre-colonial population density. Agriculture is an index of probability of cultivation given cultivable land, climate and soil composition, from Ramankutty, Foley and McSweeney (2002). Rivers captures the density of rivers as a share of land area derived from HydroSHEDS (USGS 2011). Landlocked is a dummy variable for whether the state has access to a coast or not; temperature is a yearly average in ^{o}C; altitude measures the elevation of the capital city of the state in kilometers; and Rainfall captures total yearly rainfall in meters, all are from Bruhn and Gallego (2010). More detailed data sources and descriptions in the text. Robust SE for OLS and MS SE are in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.