Pre-Colonial Institutions, Ethnic Homelands, and Economic Development in Latin America

Aldo Elizalde
Queen's University Belfast*

Sotiris Kampanelis
Cardiff University†

December 18, 2023

Abstract

This paper revisits the effects of pre-colonial institutions on present-day economic development for Latin America at the ethnic homeland level. The main hypothesis is that homelands inhabited by more advanced ethnic groups -as measured by their levels of institutional complexity- relate to better economic development today. To trace these long-term effects, we have compiled a newly digitised and georeferenced dataset of ethnic homelands. After controlling for country-specific characteristics and applying a large battery of geographical and climatic controls, we find that the effects of pre-colonial institutions are associated with a higher light density at night–as a proxy for economic development- in homelands where more advanced ethnic groups lived.

Keywords: Pre-colonial institutions, ethnic homelands, development, Latin America
JEL Codes: C21, N91, J62

1. Introduction

In recent years, there has been a growing effort to empirically uncover the enduring influence of pre-colonial institutions on contemporary economic development (Gennaioli and Rainer, 2007; Michalopoulos and Papaioannou, 2013; Angeles and Elizalde, 2017). Indeed, such efforts have been directed on the Sub-Saharan African case, a continent where colonialism arrived late (most of the African interior was

*Department of Economics, Queen's University Belfast; a.elizalde@qub.ac.uk
†Corresponding author: Cardiff Business School, Cardiff University; kampaneliss@cardiff.ac.uk

largely unknown to Europeans until the 1880s), did not last very long (African decolonisation took off during the late 1950s), and where, with the exception of South Africa and its neighbours, European settlement was of very limited importance. Under these conditions, it is perhaps not surprising that pre-colonial institutions transcended the colonial period and affect African societies to this day. While the significance of these institutions in the African context is therefore well-documented (Gennaioli and Rainer, 2007 and Michalopoulos and Papaioannou, 2013), research has increasingly recognised their vital role in shaping Latin American outcomes as well (Angeles and Elizalde, 2017; Elizalde, 2020; Elizalde et al., 2023). Notably, Angeles and Elizalde (2017) explore this relationship by analysing 324 subnational states from Latin America, revealing a robust and statistically significant association between pre-colonial institutions and current indicators of education, health, and economic development.

This paper aims to revisit the above relationship in Latin America using ethnic homelands as the core unit of observation. Indeed, while Michalopoulos and Papaioannou (2013)'s influential work has explored this relationship in Africa using ethnic homelands as the main unit of analysis, such evidence is surprisingly absent for Latin America. To address this gap in the literature, we construct a digitised and georeferenced map of ethnic homelands based on the ethnographic work of George P. Murdock in the Americas (Murdock, 1951; Murdock, 1972).

Our georeferenced map encompasses 257 ethnic homelands across Latin America. Consistent with Michalopoulos and Papaioannou (2013), we assign to each homeland a level of political complexity of the ethnic groups that inhabited them before colonisation (ranging from 0 for bands and tribes to 4 for complex states). This information is taken from George Peter Murdock's Ethnographic Atlas and via his *Outline of South American Cultures*'s book (Murdock, 1951; Murdock, 1967). By utilising satellite light density at night as a proxy for contemporary economic development, we find a robust and positive correlation between pre-colonial institutions and economic development in Latin American ethnic homelands. Our results are robust to a comprehensive set of geographic and climatic controls, as well as country-specific characteristics, including colonial and post-colonial institutions, facilitated by the use of country-fixed effects.

Our research builds upon previous economic studies investigating the connection between pre-colonial institutions and contemporary development. While prior work has focused on African nations (Gennaioli and Rainer, 2007), African ethnic homelands (Michalopoulos and Papaioannou, 2013), and in Latin American subnational states (Angeles and Elizalde, 2017), our contribution extends this understanding to Latin America by digitising and georeferencing a novel map of historical ethnic homelands. This methodology allows us to explore the impact

of pre-colonial institutions on present-day development at the ethnic homeland level. In line with previous evidence, our results demonstrate a strong and positive association between these institutions and modern development within Latin American ethnic homelands.

More broadly, our work also advances the economics literature searching for the ultimate drivers of the wealth and poverty of nations. Traditionally, studies within this literature have given considerable attention to the role of institutions (Hall and Jones, 1999; Acemoglu et al., 2001; Acemoglu et al., 2005). Yet most of this literature has focused on the role of colonial institutions —the institutional package that European colonial powers put in place throughout the world between the 16th and 20th centuries. While the importance of colonial institutions in long-run development has therefore been well-recognised, recent research has progressively uncovered a large role for pre-colonial institutions too as noted above (Gennaioli and Rainer, 2007; Michalopoulos and Papaioannou, 2013; Angeles and Elizalde, 2017). This paper contributes to these efforts by revisiting the role of pre-colonial institutions on present-day development utilising Latin American ethnic homelands.

We also contribute to the growing empirical efforts in exploring the role of historical factors affecting modern-day economic development in Latin America. Bruhn and Gallego (2012) examine the long-run effects of colonial economic activities in the American continent and found a negative correlation between present-day development and "bad" colonial economic activities –which were mostly undertaken in Latin American countries such as mining, plantation, and sugar. Maloney and Valencia Caicedo (2016) explore the lasting economic patterns in the New World, linking pre-colonial density to modern agglomerations and income persistence. Barsanetti (2021) investigates the long-term influence of the pre-colonial Peaubiru path on modern cities in Brazil, uncovering that regions in proximity to this path exhibit higher population density and urbanisation today. Our work analyses the long-run effects of pre-colonial institutional factors in Latin America using historical ethnic homelands and find that regions with a legacy of more advanced pre-colonial institutions relate to better modern economic development.

The remainder of the paper is as follows. Section 2 provides a brief historical context. Section 3 describes the data and sources. Section 4 presents the main empirical strategy and results. Section 5 concludes.

2. Historical context

Certainly, various reasons would make challenging in using ethnic homelands in Latin America to explore the long-run effects of pre-

colonial institutions on present-day development outcomes. No single ethnic group in Latin America escaped from the various havocs introduced through colonisation. Diseases, destruction and exploitation transformed the life and space of the large variety of ethnic groups that lived prior to colonisation (Adams and Macleod, 2000; Mann, 2005). The map of the Latin America was then drawn based on the delineations marked by Europeans or Meztizos (descendants of Indigenous and European ancestors), thus overlooking the existence of numerous ethnic homelands. However, colonisation expanded under different paths depending on major geographical determinants, as well as on key ethnic-specific characteristics that the Europeans encountered at their arrival in the Americas (Engerman and Sokoloff, 1997; Acemoglu et al., 2001; Acemoglu et al., 2002).

In the advent of colonisation, historical evidence suggests that several ethnic groups lived within their own ethnic homeland all along Latin America (Tanck de Estrada, 2005; Mann, 2005). These ethnic groups had a rather different way of organisation. As such, a few of them had established large centralised political systems in Central Mexico and some parts in South America. For instance, in South America, the Incas implemented the largest and more sophisticated political and administrative systems in the Americas (e.g. corvée labour or mita). In Central Mexico, the Aztecs established institutions rivalling those in the Inca Empire. Some other groups located, for instance, in what is today Colombia and Ecuador had also similar paramount chiefdom characteristics (e.g. Chibchas, Quito or Catios) for which systems large enough to control various kingdoms beyond the fringe of their local communities were one of their singular ethnic features.

On the other hand, many smaller groups such as the ones that lived in the Amazonas and the Pampas regions had less centralised and more political autonomous ethnic features. These groups were organised into small bands, and their ways of living did not require to run systems capable of administering complex political activities other than the necessary tasks that allowed them to subsist independently.

At the time of contact, Europeans somehow understood these differences of ethnic power structures, as well as the strategic advantages that came with them. Besides the expected struggles, colonisers found it easier to subjugate and settle in areas where larger groups inhabited. The already established complex ethnic systems of more advanced groups allowed them to control the native population and extract resources at large through the help of local leaders (Lang, 1974; Adams and Macleod, 2000). *Caciques* or *kurakas* provided, for at least a century or so, essential administrative and authority tasks over many small communities by generating the main tributes and supply of labour for the expansion of the colonial enterprise. A prominent envoy of the Spanish crown wrote (Saignes, 1999, p. 65):

"*The day we decide to count and tax all Indians individually, so that if they have paid they are no longer under the domination and authority of the caciques and under their orders, being free to go wherever they want, we will have taken away the restraint that holds them together in an orderly manner, for this is the way in which they survive and have survived before the Christians obtained these realms: if one could put this fact to the test for only one year, one would clearly see their destruction*".

Recent studies have therefore found a large variation of the economic activity undertaken in the colonial period within present-day Latin America's nations (Dell, 2010; Bruhn and Gallego, 2012; Maloney and Valencia Caicedo, 2016). As colonisation progressed, the relative stability that these areas offered well into the second century of colonisation may have enabled Europeans to develop some solid economic activities, as well as the main political and financial centres[1]. Indeed, although the configuration of most - if not all- the ethnic homelands vanished as the colonial power entered in the continent, it seems that one important determinant for colonisers to settle and operate economically and politically was the degree of political complexity of ethnic groups.

3. Data and Sources

3.1. Ethnic homelands and key variable of interest

To create our main unit of analysis, we have digitised and georeferenced a comprehensive map detailing the homelands of ethnic groups in mainland Latin America. This effort is founded on two pivotal works authored by George Peter Murdock, encompassing a series of maps delineating the geographical boundaries of ethnic homelands across the American continent. Our first source is the *Outline of South American Cultures* by Murdock (1951). Within this work, Murdock drew 27 distinct maps showing the homelands of all the ethnic groups from Central and South America. We managed to digitise 211 ethnic homelands, excluding mainly those located in the Caribbean.

The second source at our disposal is the *Ethnographic Bibliography of North America*. Within this compendium, Murdock provides the borders of ethnic homelands across North America, encompassing regions from Canada, Mexico, and the United States. With this source, we have undertaken the digitisation of another 46 distinctive ethnic homelands

[1]It is important to note that while the literature points out that the areas with higher development gave way to the establishment of "extractive" economic activities -resulting in negative consequences for economic development over the long-run, recent evidence addressing the "reversal of fortune" hypothesis at the subnational level is somewhat less clear. Please see Bruhn and Gallego (2012) and Maloney and Valencia Caicedo (2016) for two important contrasting views on such hypothesis.

located in Mexico. Therefore, our complete map comprises 257 ethnic homelands across mainland Latin America. Figure 1a shows this map. The black colour denotes the boundaries of ethnic homelands, while the grey one corresponds to the contemporary administrative divisions of Latin American countries.

Regrettably, one geographic segment remained unmapped due to lack of ethnographic data, which is indicated in red on the maps. This particular area is situated within the Yucatan Peninsula and extends into the northern reaches of Guatemala. While our baseline analysis has omitted this region, we assign ethnographic information in our empirical robustness checks. This is important to consider given that this geographical area has predominantly been inhabited by Maya-affiliated groups, and the absence of such information may potential affect our estimations.

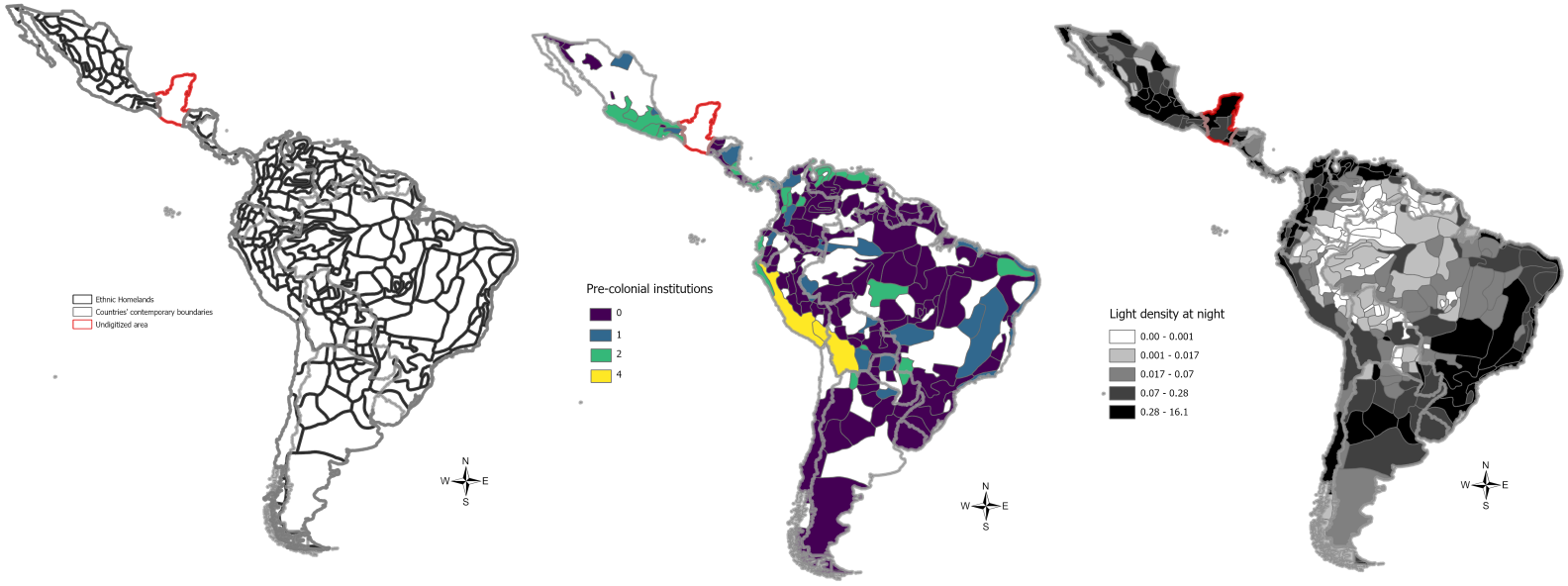
Next, we take our map and superimpose it onto a digital map showing the current administrative divisions in Latin American countries. This process helps us identify and assign the partitioned ethnic homelands to their respective countries. As a result, 371 distinct ethnic homelands were created, which we use as our main sample for the empirical analysis.[2]

To assign political complexity values to the 371 units in our study, we drew on ethnographic information from two of Murdock's key sources: the 'Ethnographic Atlas' (Murdock, 1967) and the 'Outline of South American Cultures' (Murdock, 1951). Murdock's Ethnographic Atlas provides data on various characteristics of 1,267 ethnic groups from around the world, including their level of political complexity.

For our analysis, we used the categorical variable called '*Jurisdictional Hierarchy beyond the Local Community Level*,' which ranges from 0 (representing stateless societies) to 4 (representing complex states). Initially, we found data for 79 ethnic groups in Murdock's Atlas that could be directly linked to our map. To fill in missing data, we referred to Murdock (1951)'s work, allowing us to assign political complexity values to an additional 80 ethnic groups on our map. Unfortunately, due to data limitations, we were unable to include 94 ethnic groups.[3]

[2]This increase in the number of ethnic homelands is due to the fact that some of them are spread across two or more countries. For example, the Charrua group in South America is divided among three countries: Argentina, Brazil, and Uruguay.

[3]This leaves us with a sample of 247 units of the 371 created after the intersection with the present-day countries' boundaries.



(a) Ethnic homelands (b) Pre-colonial institutions index (c) Light Density at Night

Figure 1. Ethnic homelands in Latin America

Notes: Figure 1a shows the digitisation of the ethnic homelands in Latin America based on the work by Murdock (1951) and Murdock (1972). The black colour denotes ethnic homelands and the grey colour represents contemporary administrative divisions of Latin American countries. In all three figures (1a-1b), the red colour indicates the undigitised area due to lack of ethnographic data. Figure 1b shows the variation of the index of pre-colonial institutions at the ethnic homeland level. The index was constructed using the categorical variable *Jurisdictional Hierarchy beyond the Local Community Level* drawn from Murdock's Ethnographic Atlas (Murdock, 1967), which ranges from 0 (denoting stateless societies) to 4 (complex states). Political complexity values were assigned to each ethnic group using information from Murdock (1967) and Murdock (1951). The white areas indicate ethnic groups with missing ethnographic data on political complexity. Figure 1c shows the variation of satellite light density at night at the ethnic homeland level. Lighter colour denotes lower density while darker the opposite. Light density data was drawn from the Visible Infrared Imaging Radiometer Suite (VIIRS) sensor (Elvidge et al., 2017).

In particular, our methodology for assigning political complexity values using these two sources is as follows. We assign a political complexity value to an ethnic group based on Murdock (1951)'s work only if the Ethnographic Atlas lacks this information. In cases where political complexity values are documented in both sources, the Ethnographic Atlas and the 'Outline of South American Cultures,' we prioritise the information from the Atlas. This approach is supported by the fact that the correlation of the value of political complexity collected in these two sources is 89%. However, as a robustness check, we assigned a value of zero from Murdock's index –indicating stateless characteristics– to the ethnic groups with missing values. This step serves to test for potential selection issues in our estimates.

Figure 1b shows the variation of the measure of pre-colonial institutions of ethnic groups in Latin America at the ethnic homeland level. The yellow colour denotes complex states societies; the blue colour represents paramount chiefdoms; the green colour indicates petty chiefdoms; and the purple colour signifies stateless societies. The white areas indicate the ethnic groups with missing ethnographic data on political complexity. Yet, as mentioned above, in our robustness check, we include these areas in the analysis by assigning a value of zero from Murdock's institutional complexity data. As it can be seen, the variation is prominent. Areas like the Gran Chaco and Amazonas in South America have the lowest value of pre-colonial institutions. Meanwhile, areas such as Central Mexico, the Western and some parts in the northeastern of South America concentrate ethnic groups with paramount chiefdom and state charactesitics.

Indeed, contrary to Angeles and Elizalde (2017), our approach links pre-colonial institutions to geographical regions, not individuals. Angeles and Elizalde (2017) computes an index based on the proportion of Indigenous groups in each subnational state using Murdock's Jurisdictional Hierarchy data. While valid, this method may encounter challenges due to the changing nature of Indigenous populations over time, influenced by diseases, destruction, exploitation, and forced displacements during colonial and post-colonial periods. Instead, our approach assumes that as new settlers migrated into a given geographic region, they were influenced by the existing institutional framework. Thus, in this paper, we argue that regions where more advanced ethnic groups lived, as determined by their pre-colonial institutional complexity, tend to exhibit higher levels of economic development today, irrespective of the current or historical presence of Indigenous populations.

3.2. Outcome and controls

We utilise nighttime satellite data from 2015 as our main dependent variable. Specifically, we employ the Visible Infrared Imaging Radiome-

ter Suite (VIIRS) sensor, which gives us more detailed light resolution than the widely used Defence Meteorological Satellite Programme data (Elvidge et al., 2017). Following a methodology akin to Michalopoulos and Papaioannou (2013), in instances where an ethnic homeland is partitioned in more than two countries, we attribute the luminosity to the country where the ethnic group falls. Figure 1c illustrates the variation of our dependent variable at the ethnic homeland level. To control for potential confounders, we construct a large battery of geographical and climatic variables. Namely, we control for absolute latitude, elevation, ruggedness (standard deviation of elevation), rainfall, temperature, agriculture suitability, and nearest distance to coastlines. The description and sources of all variables employed in the analysis can be found in Table B.1 in the Appendix.

4. Empirical strategy and results

To examine the effects of pre-colonial institutions on present-day economic development at the ethnic homeland level in Latin America, we use the following equation:

$$Luminosity_{e,c} = \beta PCI_{e,c} + \alpha X'_{e,c} + Z_c + \varepsilon_{e,c}, \tag{1}$$

where e indexes ethnic homelands and c indexes countries. $Luminosity_{e,c}$ is a continuous variable measuring economic development as proxy by light density at night. This variable is measured in logarithm. $PCI_{e,c}$ is a categorical variable ranging from 0 to 4 measuring the level of pre-colonial institutions. The coefficient of interest is β, which captures the relationship between pre-colonial institutions and economic development. A positive coefficient will indicate that homelands inhabited by ethnic groups with more complex political institutions relate to higher economic development in Latin America today.

$X'_{e,c}$ is a vector of geographical and climatic variables. To capture within-country variations, we introduce country-fixed effects denoted as Zc, allowing us to control for various country-specific factors, including colonial and post-colonial institutions. In all specifications, we report double-clustering at both the country and ethnic-family levels, following the methodology of Michalopoulos and Papaioannou (2013). While this double-clustering approach already addresses spatial correlation, we also estimate in our baseline regressions standard errors with a Conley-type clustering structure as proposed by Colella et al. (2019).

The results from equation (1) are presented in Table 1. The first column shows coefficients from a model that includes only the pre-colonial institutions variable, $PCI_{e,c}$, and country-fixed effects (Zc). In the second column, we augment the model with all geographical and climatic

controls, $X'e,c$, providing a fully specified model as per equation (1). The estimates for $PCIe,c$ are positive and statistically significant, indicating that homelands inhabited by ethnic groups with more complex political institutions before colonisation show higher luminosity levels in Latin America. The effect of the coefficient in column 2 is large. Since this coefficient is 0.639 (and considering that the standard deviation of luminosity is 1.09 in the sample), this means that by increasing the pre-colonial ethnic groups from tribesmen to multi-city chiefdoms (an increase of 2 units) this would be associated with a one standard deviation change in luminosity.

Table 1. Baseline results: Pre-colonial institutions and economic development in Latin American ethnic homelands

	Dependent Variable: Light density at night		
	(1)	(2)	(3)
Pre-colonial institutions	0.871**	0.639**	
Double-clustered s.e.	(0.327)	(0.258)	
Arbitrary clustered s.e.	[0.199]	[0.193]	
Pre-colonial centralisation			1.170**
Double-clustered s.e.			(0.433)
Arbitrary clustered s.e.			[0.334]
Country FE	✓	✓	✓
Geo & Climate controls		✓	✓
N	247	247	247
R^2	0.247	0.456	0.454

Notes: The unit of observation is ethnic homeland. In columns (1)-(3), the dependent variable is the logarithm of light density at night for the year 2015, ln(0.001+$Luminosity_{e,c}$). This variable uses data from the Visible Infrared Imaging Radiometer Suite (VIIRS) sensor. In columns (1) and (2), Pre-colonial institutions is a categorical variable ranging from 0 (stateless societies) to 4 (complex states) using data from Murdock's Ethnographic Atlas (Murdock, 1967) and the 'Outline of South American Cultures' by Murdock (1951). In column (3), pre-colonial centralisation is an indicator that takes the value of 1 if the Pre-colonial institutions categorical variable takes the values from 1 to 4, and zero if this variable is equal to zero. It measures the degree of political centralisation of ethnic groups in Latin America following the work of Gennaioli and Rainer (2007). All columns control for local country fixed effects. Geo & Climate controls in columns (2)-(3) refer to the set of geographical and climatic controls, which include: absolute latitude, agricultural suitability, elevation, ruggedness (standard deviation of elevation), rainfall, temperature, and distance to the sea. The descriptions of the geographical and climatic variables can be found in the Appendix. Robust standard errors clustered at both the country and ethnic-family level. Below the double clustered errors, standard errors that follows a Conley-type clustering structure to account for spatial correlations are reported (Colella et al., 2019). The constant term was omitted for space. *, ** and *** mean that the coefficient is statistically significant at 10%, 5% and 1% respectively.

Finally, in the third column, we measure the degree of pre-colonial centralisation of ethnic groups in Latin America, following Gennaioli and Rainer (2007)'s work. In particular, our categorical variable on pre-colonial institutions is transformed into an indicator by assigning a value of zero to ethnic groups lacking any political organisation beyond the local community (essentially, those without a formal state). Conversely, a value of one is assigned to groups reflecting some level of organised political structure, spanning from small or large chiefdoms to fully developed states. This transformation therefore allows us to reduce measurement errors in Murdock's categorical index as pointed out by Michalopoulos and Papaioannou (2013). We label this indicator

as pre-colonial centralisation. The coefficient in column (3) shows too a strong and positive association with luminosity.

As an alternative outcome, we utilise population density for the year 2015. Specifically, we use the Gridded Population of the World dataset drawn from the Socioeconomic Data and Applications Center (SEDAC). This data offers estimates of human population density distribution at 30 arc-second level for the whole world, allowing us to construct a variable of population density at our main unit of analysis: ethnic homeland. Table A.1 in the Appendix shows the results. The significance of our main coefficient, β, carries through all along the specifications. If anything, the estimates are more statistically robust than our main outcome, light density at night. This suggests potential path dependence effects in population dynamics arising from the legacy of the pre-existing institutional framework that settlers encountered across the Latin American land.

4.1. Grid cell level analysis

In this subsection, we delve into a detailed analysis by examining the relationship between pre-colonial institutions and modern economic development using 10 km x 10 km grid cells across Latin America. This gives us thousands of observations grouped within each ethnic homeland and country. Following the approach of Michalopoulos and Papaioannou (2013), this analysis has a dual purpose: first, it helps overcome challenges related to constructing the luminosity variable at the ethnic homeland level, where outliers from exceptionally bright pixels may have influenced the results. Second, this approach lets us uncover more precise insights into how pre-colonial institutions impact local development in Latin America.

To explore the effects of pre-colonial institutions on today's economic development at the 10 km x 10 km grid cell level, we use the following equation:

$$Luminosity_{g,e,c} = \beta PCI_{g,e,c} + \alpha X'_{g,e,c} + Z_c + \varepsilon_{g,e,c}, \tag{2}$$

This equation is similar to our main equation (1), with the key difference being that we are measuring all variables at the grid cell level. Therefore, in equation (2), g indexes for grid cell, e indexes ethnic homelands, and c indexes countries. The dependent variable $Luminosity_{g,e,c}$ is an indicator, taking the value of 1 if grid cell g has positive nightlight and 0 otherwise. The construction of the remaining variables mirrors the methodology detailed in equation (1). Likewise, we cluster our standard errors at both the country and ethnic-family levels.

The results from equation (2) are presented in Table 2. All models in Table 2 follow the same structure as those from Table 1. In Panel A, we

include all grid cells in our sample, and in Panel B, we exclude those with zero population. Notably, the results in Table 2 provide strong empirical support for the main hypothesis of the paper. Specifically, the coefficients associated with the two variables measuring pre-colonial institutions demonstrate a positive and statistically significant impact on present-day economic development.

Table 2. Grid cell analysis

Panel A:	Dependent Variable: Binary variable for nightlights		
	(1)	(2)	(3)
Pre-colonial institutions	0.100***	0.088***	
Double-clustered s.e.	(0.025)	(0.028)	
Pre-colonial centralisation			0.154**
Double-clustered s.e.			(0.065)
Country FE	✓	✓	✓
Geo & Climate controls		✓	✓
N	83692	83352	83352
R^2	0.111	0.205	0.200
Panel B:	Dependent Variable: Binary variable for nightlights		
	(1)	(2)	(3)
Pre-colonial institutions	0.099***	0.087***	
Double-clustered s.e.	(0.025)	(0.029)	
Pre-colonial centralisation			0.151**
Double-clustered s.e.			(0.066)
Country FE	✓	✓	✓
Geo & Climate controls		✓	✓
N	82221	81892	81892
R^2	0.114	0.204	0.199

Notes: The unit of observation is a grid cell (10km to 10km). Panel A includes all grid cells and Panel B excludes those with zero population. In columns (1)-(3), the dependent variable is an indicator that takes the value of 1 if grid cell *i* has light density at night for the year 2015, and 0 otherwise. This variable uses data from the Visible Infrared Imaging Radiometer Suite (VIIRS) sensor. In columns (1) and (2), Pre-colonial institutions is a categorical variable ranging from 0 (stateless societies) to 4 (complex states) using data from Murdock's Ethnographic Atlas (Murdock, 1967) and the 'Outline of South American Cultures' by Murdock (1951). In column (3), pre-colonial centralisation is an indicator that takes the value of 1 if the Pre-colonial institutions categorical variable takes the values from 1 to 4, and zero if this variable is equal to zero. It measures the degree of political centralisation of ethnic groups in Latin America following the work of Gennaioli and Rainer (2007). All columns control for local country fixed effects. Geo & Climate controls in columns (2)-(3) refer to the set of geographical and climatic controls, which include: absolute latitude, agricultural suitability, elevation, ruggedness (standard deviation of elevation), rainfall, temperature, and distance to the sea. The descriptions of the geographical and climatic variables can be found in the Appendix. Robust standard errors clustered at both the country and ethnic-family level. The constant term was omitted for space. *, ** and *** mean that the coefficient is statistically significant at 10%, 5% and 1% respectively.

4.2. Robustness Checks

We proceed with a set of robustness checks that aim to fortify the credibility of our main findings. First, we address the potential selection bias arising from the absence of political complexity values for all societies. To mitigate this concern, we assign a value of zero (indicating stateless societies) to societies lacking information on political complexity. Following the methodology proposed by Angeles and

Elizalde (2017), we assume that the absence of this information is attributed to the fact that anthropologists generally overlook smaller groups. Column (1) in Table 3 shows the results. Panel A presents the coefficients using the categorical variable on pre-colonial institutions, while in Panel B we use the indicator on political centralisation. The coefficients remain positive and statistically significant, suggesting that our results are not being affected by the absence of observations in the sample.

Table 3. Robustness checks

Panel A:	Dependent Variable: Light density at night		
	(1) all obs.	(2) Maya==2	(3) Maya==1
Pre-colonial institutions	0.581***	0.643**	0.648**
Double-clustered s.e.	(0.177)	(0.227)	(0.254)
Country FE	✓	✓	✓
Geo & Climate controls	✓	✓	✓
N	371	252	252
R^2	0.464	0.463	0.462
Panel B:	Dependent Variable: Light density at night		
	(1) all obs.	(2) Maya==2	(3) Maya==1
Pre-colonial centralisation	1.069***	1.186***	1.186***
Double-clustered s.e.	(0.326)	(0.406)	(0.406)
Country FE	✓	✓	✓
Geo & Climate controls	✓	✓	✓
N	371	252	252
R^2	0.463	0.460	0.460

Notes: The unit of observation is ethnic homeland. Panel A reports estimates using the categorical variable for Pre-colonial institutions, which ranges from 0 (stateless societies) to 4 (complex states). Panel B presents estimates using the indicator for Pre-colonial centralisation. This indicator takes the value of 1 if the Pre-colonial institutions categorical variable takes the values from 1 to 4, and zero if this variable is equal to zero. It measures the degree of political centralisation of ethnic groups in Latin America following the work of Gennaioli and Rainer (2007). Both variables were constructed using using data from Murdock's Ethnographic Atlas (Murdock, 1967) and the 'Outline of South American Cultures' by Murdock (1951). In columns (1)-(3) for both Panels A & B, the dependent variable is the logarithm of light density at night for the year 2015, $\ln(0.001+Luminosity_{e,c})$. This variable uses data from the Visible Infrared Imaging Radiometer Suite (VIIRS) sensor. In column (1), all observations in the sample were included by assigning a value of zero to the ethnic groups with missing ethnographic data on political complexity. In column (2), region linked to the Maya group (red area highlighted on Figure 1) was assigned a value a 2 (paramount chiefdom) from Murdock's categorical variable '*Jurisdictional Hierarchy beyond the Local Community Level*, whereas in column (3) a value of 1 (petty chiefdom) is assigned to this group. All columns control for local country fixed effects and for geographical and climatic controls, which include: absolute latitude, agricultural suitability, elevation, ruggedness (standard deviation of elevation), rainfall, temperature, and distance to the sea. The descriptions of the geographical and climatic variables can be found in the Appendix. Robust standard errors clustered at both the country and ethnic-family level. The constant term was omitted for space. *, ** and *** mean that the coefficient is statistically significant at 10%, 5% and 1% respectively.

An additional potential concern regarding selection bias arises from the omission of ethnic homelands in the Yucatan Peninsula in Mexico and Guatemala due to the absence of ethnographic data (highlighted in red on Figure 1). Given the predominantly Maya-affiliated groups in this region—among the largest and most renowned in the Americas— its exclusion could impact our estimates. Addressing this issue in columns (2) and (3) Table 3, we assign a political complexity value of 2

and 1 from Murdock's Jurisdictional Hierarchy index to this entire area, respectively. The coefficients in these columns continue demonstrating a relevant impact on economic development, mitigating the aforementioned concern.

Another source of concern relates to omitted variable problems. We address this issue in Appendix Table A.2 via Oster (2019)'s method. This approach results in estimates with a 30% higher R^2 than our baseline results and assumes proportional selection of observed to unobserved controls, producing a new β coefficient and a parameter δ. To eliminate bias risk, the range between our baseline and the new β should confidently exclude zero, and δ should be higher than 1. The results presented in Table A.2 show that the new β coefficient closely aligning with our baseline estimate and δ highly exceeds 1. This strongly suggests that our model is unlikely to be impacted by omitted variable bias.

5. Conclusion

In this paper, we have revisited the role of pre-colonial institutions on present-day development for Latin America. To do so, we have digitised and georeferenced a novel map of ethnic homelands. This effort has allowed us to examine the long-term effects of pre-colonial institutions at the ethnic homeland level. In line with previous evidence, our results demonstrate that areas inhabited by more advanced ethnic groups -as measured by their levels of institutional complexity- relate to better economic development today -as reflected by satellite light density at night.

Indeed, in the Latin American context, historical trajectories diverged from those in Africa —a region where economists have somewhat paid more attention in understanding the role of pre-colonial institutions. Colonialism was not only much longer lasting than in other regions of the world, about three centuries for most Latin American nations, but it was also accompanied by a massive transformation of the land and ethnic structure of the population. While correct, what remained after colonisation was not a mirror image of European society but a new reality where pre-colonial culture and institutions survived. The evidence presented in this paper is an effort to progress this line of scholarly enquiry through the lens of Latin America's long history of pre-existing societies and territorial domains.

References

Acemoglu, Daron, Simon Johnson, and James A Robinson (2001). The colonial origins of comparative development: An empirical investigation. *American economic review* 91(5), pp. 1369–1401.

— (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *The Quarterly journal of economics* 117(4), pp. 1231–1294.

— (2005). Institutions as a fundamental cause of long-run growth. *Handbook of economic growth* 1, pp. 385–472.

Adams, Richard EW and Murdo J Macleod (2000). History of the native peoples of the Americas. Cambridge University Press.

Angeles, Luis and Aldo Elizalde (2017). Pre-colonial institutions and socioeconomic development: The case of Latin America. *Journal of Development Economics* 124, pp. 22–40.

Barsanetti, Bruno (2021). Cities on pre-Columbian paths. *Journal of Urban Economics* 122, p. 103317.

Bruhn, Miriam and Francisco A Gallego (2012). Good, bad, and ugly colonial activities: do they matter for economic development? *Review of Economics and Statistics* 94(2), pp. 433–461.

Colella, Fabrizio, Rafael Lalive, Seyhun Orcan Sakalli, and Mathias Thoenig (2019). Inference with arbitrary clustering.

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

Elizalde, Aldo (2020). On the economic effects of Indigenous institutions: Evidence from Mexico. *Journal of Development Economics* 147, p. 102530. ISSN: 0304-3878.

Elizalde, Aldo, Eduardo Hidalgo, and Nayeli Salgado (2023). Public good or public bad? Indigenous institutions and the demand for public goods. *techreport*. CUCEH Working Paper Series.

Elvidge, Christopher D, Kimberly Baugh, Mikhail Zhizhin, Feng Chi Hsu, and Tilottama Ghosh (2017). VIIRS night-time lights. *International Journal of Remote Sensing* 38(21), pp. 5860–5879.

Engerman, Stanley L and Kenneth L Sokoloff (1997). Factor endowments, institutions, and differential paths of growth among new world economies. *How Latin America Fell Behind*, pp. 260–304.

Fischer, Guenther, Freddy Nachtergaele, Sylvia Prieler, HT Van Velthuizen, Luc Verelst, and David Wiberg (2008). Global agro-ecological zones assessment for agriculture (GAEZ 2008). *IIASA, Laxenburg, Austria and FAO, Rome, Italy* 10.

Gennaioli, Nicola and Ilia Rainer (2007). The modern impact of pre-colonial centralization in Africa. *Journal of Economic Growth* 12(3), pp. 185–234.

Hall, Robert E and Charles I Jones (1999). Why do some countries produce so much more output per worker than others? *The quarterly journal of economics* 114(1), pp. 83–116.

Lang, James Joseph (1974). Conquest and Commerce: Spain and England in the Americas. University of Michigan.

Maloney, William F and Felipe Valencia Caicedo (2016). The persistence of (subnational) fortune. *The Economic Journal* 126(598), pp. 2363–2401.

Mann, Charles C (2005). 1491 The Americas before Columbus. Granta Books, London.

Michalopoulos, Stelios and Elias Papaioannou (2013). Pre-colonial ethnic institutions and contemporary African development. *Econometrica* 81(1), pp. 113–152.

Murdock, George Peter (1951). Outline of South American Cultures. Human Relations Area Files.

— (1967). Ethnographic atlas: a summary. *Ethnology* 6(2), pp. 109–236.

— (1972). Ethnographic Bibliography of North America. Human Relations Area Files.

Oster, Emily (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics* 37(2), pp. 187–204.

Saignes, Thierry (1999). The Colonial Condition in the Quechua-Aymara Heartland (1570–1780). na.

Tanck de Estrada, Dorothy (2005). Atlas ilustrado de los pueblos de indios. Nueva España 1800. El Colegio de Mexico, El Colegio Mexiquense, CNDPI, Fomento Cultural Banamex.

Appendix (not for publication)

A. Tables and Figures

Table A.1. Baseline results: Population

	Dependent Variable: Population density		
	(1)	(2)	(3)
Pre-colonial institutions	0.654**	0.439**	
Double-clustered s.e.	(0.229)	(0.155)	
Arbitrary clustered s.e.	[0.136]	[0.124]	
Pre-colonial centralisation			0.840***
Double-clustered s.e.			(0.258)
Arbitrary clustered s.e.			[0.236]
Country FE	✓	✓	✓
Geo & Climate controls		✓	✓
N	244	244	244
R^2	0.269	0.525	0.525

Notes: The unit of observation is ethnic homeland. In columns (1)-(3), the dependent variable is the logarithm of population density for the year 2015, ln($PopDen_{e,c}$). This variable uses data from the Gridded Population of the World Database. In columns (1) and (2), Pre-colonial institutions is a categorical variable ranging from 0 (stateless societies) to 4 (complex states) using data from Murdock's Ethnographic Atlas (Murdock, 1967). In column (3), pre-colonial centralisation is an indicator that takes the value of 1 if the Pre-colonial institutions categorical variable takes the values from 1 to 4, and zero if this variable is equal to zero. It measures the degree of political centralisation of ethnic groups in Latin America following the work of Gennaioli and Rainer (2007). All columns control for local country fixed effects. Geo & Climate controls in columns (2)-(3) refer to the set of geographical and climatic controls, which include: absolute latitude, agricultural suitability, elevation, ruggedness (standard deviation of elevation), rainfall, temperature, and distance to the sea. The descriptions of the geographical and climatic variables are presented in the Appendix. Robust standard errors clustered at both the country and ethnic-family level. Below the double clustered errors, standard errors that follows a Conley-type clustering structure to account for spatial correlations are reported (Colella et al., 2019). The constant term was omitted for space. *, ** and *** mean that the coefficient is statistically significant at 10%, 5% and 1% respectively.

Table A.2. Robustness: Omitted Variable Bias

	Dependent variable: Light density at night			
	(1) Baseline specification coefficient b	(2) Identified Set (b,b'(Rmax,Delta=1),b)	(3) Exclude Zero	(4) Absolute Delta (δ)
Pre-colonial institutions	0.639^{**}	$[0.553, 0.639]$	✓	$4.3 > 1$
Country FE	✓	✓		
Geo & Climate controls	✓	✓		
N	247	247		

Notes: The unit of observation is ethnic homeland. The dependent variable is the logarithm of light density at night for the year 2015, ln(0.001+$Luminosity_{e,c}$). This variable uses data from the Visible Infrared Imaging Radiometer Suite (VIIRS) sensor. Pre-colonial institutions is a categorical variable ranging from 0 (stateless societies) to 4 (complex states) using data from Murdock's Ethnographic Atlas (Murdock, 1967) and the 'Outline of South American Cultures' by Murdock (1951). In the first column, we present the conservative coefficient obtained in column (2) Table 1. Moving to column (2), we illustrate a range of coefficients —-from the initial conservative estimate (b) to the updated b coefficient ($\dot{b}$(Rmax,Delta=1)) derived using the method outlined by Oster (2019). In the third column, a checkmark (✓) indicates that($\dot{b}$(Rmax,Delta=1),b) reliably excludes zero, confirming the absence of omitted variable bias. In column (4), the comparison 4.1 ¿ 1 tells us that the parameter δ exceeds one, suggesting that the model in equation 1 is likely unaffected by omitted variable bias. Regression controls for local country fixed effects and for geographical and climatic controls, which include: absolute latitude, agricultural suitability, elevation, ruggedness (standard deviation of elevation), rainfall, temperature, and distance to the sea. The descriptions of the geographical and climatic variables can be found in the Appendix. Robust standard errors clustered at both the country and ethnic-family level. *, ** and *** mean that the coefficient is statistically significant at 10%, 5% and 1% respectively.

B. Variable definitions

Table B.1. Variable definitions and sources

Variable	Description	Source
Main Outcome Variable		
Luminosity	Log average satellite density light at night for the year 2015.	VIIRS's lights in 2015 from Elvidge et al. (2017) and NOOA/National Centers for Environmental information: here
Alternative Outcome Variable		
Population density	Log average population density for the year 2015	Gridded Population of the World (GPW), v4 drawn from Socioeconomic Data and Applications Center (SEDAC): here
Main Control Variables		
Latitude	Absolute latitude	Author's elaboration using Arcgis Pro.
Agriculture Suitability	Log average values of critical soil characteristics crucial for effective crop growth.	Authors' elaboration using data from Fischer et al. (2008): here
Elevation	Log average altitude in meters above or below the sea level.	Author's elaboration using data from WorldClim data.
Ruggedness	Log standard deviation of the altitude.	Author's elaboration using data from WorldClim data.
Precipitation	Log annual average precipitation	Author's elaboration using data from WorldClim data.
Temperature	Log annual average of temperature	Author's elaboration using data from WorldClim data.
Distance to the Sea	The log geodesic distance (in kilometres) from the centroid of each ethnic homeland to the nearest coastline.	Authors' elaboration using ArcGIS Pro.