

# Explaining Spatial Variations in Productivity

Evidence from Latin America and the Caribbean

*Luis E. Quintero*

*Mark Roberts*



**WORLD BANK GROUP**

Social, Urban, Rural and Resilience Global Practice

&

Latin America and Caribbean Region

August 2018

## Abstract

There is a large and extensive literature examining the strength of agglomeration economies and, more generally, the determinants of spatial variations in productivity for developed countries. However, the corresponding literature for developing countries is comparatively scant. This paper contributes to filling this knowledge gap by providing estimates for city productivity premiums and different sources of agglomeration effects for 16 countries in the Latin America and Caribbean region. While two of the countries in our sample—Brazil and Colombia—have been considered by the literature, the remaining 14 countries have not been previously analyzed. The paper presents estimates for the region as well as comparable estimates for each country

using a harmonized data set with characteristics of individual workers and features of the cities in which the workers live. In addition to examining the strength of agglomeration economies, the roles of human capital externalities and market access in explaining subnational productivity variations are assessed. The paper finds that citywide human capital externalities appear much stronger than agglomeration economies in explaining productivity variation in all the considered countries. There is considerable heterogeneity in the estimated strength of human capital externalities across countries, which could be a reflection of country differences in educational quality.

---

This paper is a product of the Social, Urban, Rural and Resilience Global Practice and the Office of the Chief Economist, Latin America and Caribbean Region. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://www.worldbank.org/research>. The authors may be contacted at [mroberts1@worldbank.org](mailto:mroberts1@worldbank.org).

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.*

# Explaining Spatial Variations in Productivity: Evidence from Latin America and the Caribbean.

Luis E. Quintero\* and Mark Roberts†

**JEL Classification:** R12, R3,O47, O54, O18. **Keywords:** Agglomeration economies, human capital externalities, market access, city productivity, developing economies, Latin America. subsidies, diversification.

---

\*Carey Business School, Johns Hopkins University. leq@jhu.edu.

†Social, Urban, Rural and Resilience (SURR) Global Practice, The World Bank. mroberts1@worldbank.org.

Funding for this work from the World Bank is gratefully acknowledged. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

We are grateful to Gilles Duranton for valuable discussion at different stages of the paper. We also thank Diego Puga, Matthew Turner, Edward Glaeser, Joshua Gottlieb, Victor Couture, María Marta Ferreyra, Daniel Lederman, and Ming Zhang for useful comments, as well as participants of the 2017 NBER Summer School on Urban Economics, the Authors' Workshop for the World Bank's Flagship Report, "Raising the Bar for Productive Cities in Latin America and the Caribbean" (for which this paper was prepared as input), the 2017 NARSC Conference, the DC Urban Economics Day, and the 18<sup>th</sup> European Meetings of the Urban Economics Association. We thank Joao Jatene and Jane Park for exceptional research assistance, and Julia Branson, Andrew Campbell-Sutton, Graeme M. Hornby, Duncan D. Hornby and Chris Hill at the GeoData Institute in the University of Southampton for excellent data work.

# 1 Introduction

There exist wide productivity differences between cities within countries (Combes et al., 2012). How different productivity is in different cities may depend on many factors. On the one hand, a city may attract or repel an unusually talented workforce and group of firms, the individual members of which would be highly productive independently of their location. The sorting of high skill workers into some cities would make these places more productive. If larger or denser cities are more attractive to these workers, sorting would imply higher productivity will be observed in larger scale cities.<sup>1</sup> On the other hand, a city may exhibit a higher level of productivity because of specific attributes associated with its environment and the externalities generated by these attributes on productivity. If larger or denser cities are associated with more positive externalities, a larger scale would imply higher productivity.<sup>2</sup>

The policy implications of these two explanations for variations in city productivity are rather different. The second explanation implies that cities can be *special places* with the potential to help firms and workers be more productive if they have certain attributes. In the context of this second explanation, urban economics has identified three main theories of urban success related to city attributes.<sup>3</sup>

The first theory focuses on *agglomeration economies*, a set of positive externalities that increase productivity and that are generated in places with larger populations and/or densities. These externalities arise through a variety of potential mechanisms, which Duranton and Puga (2004) identify as *sharing, matching and learning*. Thick labor markets that characterize larger or denser cities can help generate better matches between firms and workers. Moreover, these cities can also host a large and diversified array of specialized suppliers of goods and services, which provide cheaper and better intermediate inputs to firms. Finally, the close geographic proximity of both people and firms implied by higher density can give rise to the, often unintended, productivity-enhancing spillover of ideas, as workers from different firms learn from each other, both through observation and interaction.

The second theory refers to positive human capital externalities. As with agglomeration economies, people learn from observing and interacting with other people in a city as they go about their daily working lives. The crucial difference here, however, is that learning is more likely to occur from higher skilled than from lower skilled workers. This leads to the prediction that a worker's individual productivity will be increasing in the average level of human capital of the city in which she lives (Rauch, 1993; Moretti, 2004b; Davis and Dingel, 2017).<sup>4</sup>

---

<sup>1</sup>Davis and Dingel (2017) develop theory to show that higher skills and high skill intensive sectors are expected to be disproportionately located in large scale cities and provide evidence of their predictions for US cities. Davis et al. (2018) shows the same for Brazil, China, and India.

<sup>2</sup>We use the term city to refer to subnational locations that conceptually share the same workforce. This includes locations that might have a rural component. The specific definition of locations will change across the countries we analyze. We attempt to make them as comparable as possible. See more details in section 3.

<sup>3</sup>As is evident from the discussion below, these three theories are inter-related and, to a certain extent, overlapping. See also (Roberts, 2018) for a more detailed discussion on this point.

<sup>4</sup>See Duranton (2015) for a review of the issue for developing countries.

Finally, the third theory is that cities can generate higher levels of productivity because they also tend to benefit from higher levels of access to both large consumer markets and markets for intermediate inputs. Superior market access stems from a city’s connectivity to the markets of surrounding areas.<sup>5</sup> Higher levels of market access lower export costs (Krugman, 1980), lower input costs for local firms (Krugman and Venables, 1995; Bartelme, 2018) and increase effective demand for local firms’ products, which would stimulate increases in productivity, especially in a context of economies of scale (Krugman, 1991).<sup>6</sup>

There exists an extensive empirical literature analyzing the above three theories of urban success (see Combes and Gobillon (2015) for a detailed review of this literature). However, as has been noted by, for example, Overman and Venables (2005), Desmet and Rossi-Hansberg (2013), and Duranton (2015), this literature is largely confined to developed countries, and rarely does it consider all three theories together. As such, there exists an important empirical blind spot with regards to the strength of the basic forces which help to govern the productivity of cities in developing countries. This is at a time when there are good reasons to suspect that the strength of agglomeration economies may differ significantly for developing countries on account of, *inter alia*, differences in their economic structures, levels of institutional development, and infrastructure stocks.

In response to the above knowledge gap, there has been a recent effort in the literature to generate estimates of the strength of agglomeration economies for several developing countries. Duranton (2016) presents such estimates for Colombia, while Chauvin et al. (2017) do likewise for Brazil, China and India. In both cases, in order to allow for the comparability of their results, the authors make use of empirical specifications that have previously been applied to developed countries in the literature. Chauvin et al. (2017) empirically examine the roles of both agglomeration economies and human capital externalities in driving urban productivity differences for Brazil, China and India. Duranton (2016) also examines both of these theories for Colombia in addition to empirically investigating the role of market access. Authors who have studied the empirical relationship between sub-national levels of productivity and levels of market access, without necessarily accounting for agglomeration economies and human capital externalities in a developing country context, include Au and Henderson (2006), Roberts et al. (2012) and Hering and Poncet (2010) for China, Amiti and Cameron (2007) for Indonesia, and Fally et al. (2010) for Brazil.

This paper contributes to the above effort to fill the knowledge gap on the determinants of urban success for developing economies by providing estimates of the roles of population density, human capital and market access in driving sub-national productivity differences within 16 Latin American and Caribbean (LAC) countries. By controlling for key individual worker characteristics

---

<sup>5</sup>The measure of market access that we use in this paper excludes a city’s own market. This helps to distinguish market access from the measure of agglomeration that we use, while also helping to mitigate endogeneity concerns associated with reverse causation.

<sup>6</sup>The effect of market access on local productivity is not clear theoretically, as higher access to final products can also decrease investment in local industries and prevent their development. Also, access to cheaper consumption goods would increase real wages in cities with higher market access, which would imply lower nominal wages in spatial equilibrium (Handbury and Weinstein, 2014; Duranton, 2016). However, Combes and Gobillon (2015) review analyses of market access on local productivity and conclude that “*the positive effect of the economic size of distant locations and the spatial decay of this effect are rarely rejected empirically*”.

in our estimation, we aim to mitigate bias due to sorting effects. While, as indicated above, two of our countries - Brazil and Colombia - have already been considered by [Chauvin et al. \(2017\)](#) and [Fally et al. \(2010\)](#), and [Duranton \(2016\)](#) respectively - the remaining 14 countries have not, to our knowledge, been previously analyzed in the literature.

To ensure comparability of our estimates across countries, we use a harmonized dataset of nominal wages and characteristics of individual workers, and of the characteristics of the locations in which the workers live.<sup>7</sup> This data set is constructed from successive rounds of country specific household surveys extracted from the Socio-Economic Database for Latin America and the Caribbean (SEDLAC), and from a Geo-Spatial Database for Latin America and the Caribbean that was constructed for the World Bank's Regional Flagship report Raising the Bar for Productive Cities in Latin America and the Caribbean ([Ferreyra and Roberts, 2018](#)). To further ensure the comparability of results both across the countries we study and with those available elsewhere in the literature for other countries, we apply a similar empirical specification to previous studies in the literature.

We find sizeable variations in sub-national levels of nominal wages - and, hence, productivity - within countries. Higher levels of productivity are typically observed in and around larger and more densely populated cities. Much of the variation in productivity across sub-national areas can be explained by observable compositional differences in the workforce, the skill-selective sorting of workers discussed previously. More productive areas tend to be populated by better educated workers.

Yet, there remains an important component of sub-national productivity whose variation cannot be explained by compositional differences in the workforce. This is consistent with the existence of positive externalities and spillovers between workers and/or firms and, therefore, the three theories of urban success, which we test empirically. Out of the three theories of urban success, evidence is found to support that the most important channel in the countries we analyze is human capital externalities, both in the pooled analysis and in the individual estimations by country.<sup>8</sup> Market access measures are found to be a statistically significant predictor of variations in sub-national productivity in the pooled analysis. However, this is only robust in 4 countries in the individual analysis.<sup>9</sup> Finally, agglomeration economies appear to be weak or non-existent across LAC countries in the pooled analysis. Five countries show positive and significant elasticities with respect to density, but all with very small magnitude. The effect is not significant in the pooled analysis. One potential explanation for the weakness of the agglomeration economies in our sample is sub-optimal infrastructure that increases congestion under high densities, which in turn overwhelms

---

<sup>7</sup>Ideally, we would like to estimate agglomeration effects on the basis of where the workers are employed. However, our data only identify where they live.

<sup>8</sup>Our results on the importance of human capital could also, at least partly, be attributable to human capital complementarities in production. Empirically, it is difficult to distinguish between human capital externalities and complementarities. See [Moretti \(2004b\)](#) for a discussion on the interpretation of empirical estimates under externalities and complementarities and [Ferreyra \(2018\)](#) for analysis of the roles of human capital externalities and complementarities in driving high returns to human capital in LAC.

<sup>9</sup>Our market access measure does not include access to international markets.

agglomeration economies.

The structure of the remainder of the paper is as follows. In the next section, we present the empirical strategy used to estimate a city’s productivity premium controlling for the composition of its workforce, as well as the effect of the determinants of urban success on this premium. Following this, we discuss the construction of the dataset. We then present results of the analysis for all countries pooled, before proceeding to examine the heterogeneity of results found across both different countries and different sub-groups of workers (i.e. between young and old workers, male and female workers, public and private sector workers, and formal and informal workers). Finally, we offer some concluding remarks on our findings.

## 2 Estimation

To test different mechanisms of urban success, we first focus on estimating city productivity premiums controlling for sorting of workers with different education levels and other observable characteristics, and then attempt to estimate the effects on these premiums of variables that are representative of the different theories discussed above. To estimate the city productivity premium we focus on a regression of the natural logarithm of a worker’s nominal wage on a set of location dummies and a set of observable worker characteristics.<sup>10</sup> Following [Combes and Gobillon \(2015\)](#), our estimating equation is derived from a setting where a representative firm located in city  $l$  at time  $t$ , produces goods  $Y_{l,t}$ , using labor  $L_{l,t}$  and capital  $K_{l,t}$ , at factor prices  $\omega_{l,t}$  and  $r_{l,t}$  respectively, and sells them at price  $p_{l,t}$ . The firm’s profits are given by

$$\begin{aligned}\pi_{l,t} &= p_{l,t}Y_{l,t} - \omega_{l,t}L_{l,t} - r_{l,t}K_{l,t} \\ \pi_{l,t} &= p_{l,t} \left( \frac{A_{l,t}}{D} (s_{l,t}^\beta L_{l,t})^\alpha K_{l,t}^{1-\alpha} \right) - \omega_{l,t}L_{l,t} - r_{l,t}K_{l,t}\end{aligned}\quad (1)$$

where the term in brackets is output assuming a Cobb-Douglas production function.  $\alpha$  is the share of (augmented) labor and  $D = \alpha^\alpha(1-\alpha)^{1-\alpha}$  is a constant.  $A_{l,t}$  captures the firm productivity (TFP).  $s_{l,t}$  are labor skills, which augment the impact of labor on production by the factor  $s_{l,t}^\beta$ , where  $\beta$  is the wage elasticity of labor skills. Profit maximization determines wages in a competitive equilibrium:

$$\omega_{l,t} = \left( \frac{p_{l,t}A_{l,t}}{r_{l,t}^{1-\alpha}} \right)^{\frac{1}{\alpha}} s_{l,t}^\beta \quad (2)$$

Nominal wages are a combination of worker skills and firm productivity, product and capital

---

<sup>10</sup>We assume a competitive labor market. Otherwise, the bargaining power of workers could change with the size of a city and this could be perceived as an additional source of agglomeration effects. In the case of a monopsonic market structure, wages would be proportional but not equal to productivity. To avoid confounding agglomeration economies and market structure, some studies use TFP as a productivity measure ([Henderson, 2003](#); [Cingano and Schivardi, 2004](#); [Combes et al., 2010](#); [Cainelli et al., 2015](#)). We cannot follow such an approach due to the absence of such data with harmonized methodology for a large number of countries in our study.

prices. Taking logarithms we get

$$\ln \omega_{l,t} = B_{l,t} + \beta \ln s_{l,t} \quad (3)$$

where  $B_{l,t}$  represents the term accompanying the skills productivity augmenting component. We do not have data on local prices or TFP. However, we can measure these compounded effects ( $B_{l,t}$ ) with fixed effects for locations and years, since these vary by location and time. This effect is the city productivity premium.<sup>11</sup> In the data, skills and wages vary by worker, so we relax the implied assumption of a representative worker, and estimate the equation above at the worker level  $i$ .

$$\ln \omega_{i,l(i),t} = B_l + \beta \ln s_{i,l(i),t} + \epsilon_{i,l(i),t} \quad (4)$$

where  $l(i)$  is the location (city) of worker  $i$  at time  $t$ , and  $\epsilon_{i,l(i),t}$  is a stochastic error.  $B_l$  is now indexed by location because our data prevent us from estimating a separate fixed effect for each city in each time-period.<sup>12</sup> Rather, we estimate a single city fixed effect while controlling for survey-year fixed effects.

This specification raises an important estimation concern. We do not have a perfect measure of skills. We can separate skills into an observed component comprised of demographic characteristics as well as education history of the worker, and an unobserved component. Ideally, we would identify the unobserved effects, assuming they are time-invariant, by tracking workers in a panel (Combes et al., 2010; D’Costa and Overman, 2014). However, we do not have access to a panel and therefore are subject to an endogeneity (omitted variable) bias. In estimation, the observable worker characteristics that we control for are years of schooling, age and its square, gender, and marital status. Similar to Duranton (2016) and Chauvin et al. (2017), we see the potential biases in the estimated coefficients arising from omitted heterogeneity as a price worth paying for the knowledge obtained about developing countries, given the alternative of not undertaking the empirical work. We also note that, in recent work for Spain using rich administrative data, De La Roca and Puga (2017) have found that, once learning effects in cities are controlled for, workers in larger cities do not exhibit higher initial unobserved ability than workers in smaller cities, which eases to some extent the concern for bias arising from not controlling for these unobserved skills. Under these circumstances, a higher productivity premium in larger scale cities can be interpreted as capturing both the static and dynamic agglomeration benefits associated with such cities.

After estimating 4, we estimate a second stage using the estimated city fixed effects or productivity premiums  $\hat{B}_l$  as the dependent variable.

---

<sup>11</sup>We are assuming a representative firm in this derivation and do not separate firm specific effects and location effects, but instead average over all firms by imposing that this variable does not change for each location and time pair. This is a standard approach. To incorporate firm heterogeneity, we also estimate versions of the model as robustness tests where we incorporate firm and job characteristics to disentangle the citywide effects and the sorting of different types of firms into different locations. The results are similar and are available from the authors by request.

<sup>12</sup>This is because the exact survey years differ between countries and we do not observe every city in our sample in every survey.

$$\hat{B}_l = \theta_1 \ln \text{density}_l + \theta_2 \ln \text{human capital}_l + \theta_3 \ln \text{market access}_l + \text{controls}_l + \mu_l \quad (5)$$

This regression relates the estimated city productivity effect net of skill sorting to variables associated with the three urban success theories discussed above. Population density captures the theory of agglomeration economies, a set of positive externalities that increase productivity in larger or denser places. Human capital captures human capital externalities - i.e. spillovers of knowledge that increase with a city's stock of human capital.<sup>13</sup> We initially assume the impact of human capital externalities to affect all workers equally.<sup>14</sup> To measure human capital, we use average years of education completed in the working-age population and the share of the working-age population that has completed higher education (each in different specifications). The effect of the density and human capital variables would be reflected in the productivity component,  $A_{l,t}$  of the city effect  $B_{l,t}$  in equation 2.

Market access captures the third theory that a city can generate higher productivities because they benefit from access to larger markets for both final goods (higher average perceived  $p_{l,t}$  because of lower trade costs) and intermediate inputs (lower  $r_{l,t}$ ) (Combes, 2011). The effect of market access then would impact the other variables in the city effect  $B_{l,t}$ .

Additional controls include a location's average temperature, terrain ruggedness, and total precipitation, as well as country and survey-year fixed effects.

Further estimation concerns appear in this second stage. First,  $\hat{B}_{l,t}$  is itself an estimated variable and therefore has some estimation error. In the second stage, we are not taking that estimation error into account. This can be overcome by either using the standard errors from the first stage as weights in a generalized least squares estimation in the second stage, or by performing all the estimation in a single stage, nesting equation 4 into 5 as follows:

$$\begin{aligned} \ln \omega_{i,l(i),t} &= \kappa \ln s_{i,l(i),t} + \tau_1 \ln \text{density}_{l(i)} + \tau_2 \ln \text{human capital}_{l(i)} \\ &\quad + \tau_3 \ln \text{market access}_{l(i)} + \text{controls}_{l(i)} + \rho_{i,l(i),t} \end{aligned} \quad (6)$$

We present our main results using the two stage approach as it allows us to describe and map the spatial distribution of  $\hat{B}_{l,t}$ , which is interesting in its own right.<sup>15</sup> We estimate 6 as a robustness test.<sup>16</sup>

Second, we are not comparing the wages of workers randomly assigned to cities. Instead, we are comparing workers that have endogenously chosen, up to affording mobility costs, their location. This can imply that there are some city characteristics that are correlated with wages and that are

<sup>13</sup> This might also be capturing complementarities in production between skilled and unskilled workers.

<sup>14</sup> Combes et al. (2008) analyze differential effects for high and low skilled workers, which also sheds light on possible substitutability between these two types of workers. We estimate regressions with subsamples that partially capture some of these effects in section 4.2.2.

<sup>15</sup> Our adoption of a 2-stage approach to estimation follows Combes et al. (2008) and De La Roca and Puga (2017).

<sup>16</sup> In estimating equation 6, we allow the effects of individual worker characteristics to differ across countries so as to mimic our 2-stage regression approach as closely as possible.

absent from our equation, introducing an omitted variable problem. In an effort to solve this, we include control variables, as well as a set of job and firm characteristics in robustness tests.<sup>17</sup>

### 3 Data

Our first stage (equation 4) analysis draws on successive rounds of household survey micro-data for 16 LAC countries that, apart from Brazil, are taken from the Socio-Economic Database for Latin America and the Caribbean (SEDLAC). This database has been jointly constructed by the Center for Distributive, Labor and Social Studies (CEDLAS) at the Universidad Nacional de La Plata and the World Bank’s Poverty Group for the Latin American and the Caribbean region.<sup>18</sup> Previously, data availability has been an important obstacle to carrying out large scale work on agglomeration economies for developing countries. In particular, surveys are not uniform across LAC countries. Comparability is, therefore, an issue of great concern. One of the great advantages of SEDLAC is that it provides harmonized survey micro-data. Strong comparability of the data is ensured across time and countries by using similar definitions of variables in each country/year, and by applying consistent methods of processing the data (CEDLAS and The World Bank, 2014). SEDLAC provides as close to perfect comparability as can be hoped for given the continuing prevalence of differences in coverage and questionnaires of household surveys across LAC countries.

The version of SEDLAC we use covers different survey years for different countries - for example, from 1974 to 2014 for Argentina, from 1987 to 2013 for Chile, and from 2001 to 2014 for Colombia. To ensure consistency across the LAC countries on which our analysis focuses, we make use only of SEDLAC data from 2000 onwards. To allay potential concerns over a lack of representativeness of the survey data at the level at which we analyze it, we pool successive cross-sections of data from this period. SEDLAC covers 24 countries. However, eight countries had to be dropped either because: (i) changes in administrative units and their coding over time prevent SEDLAC from providing reliable geographic identifiers; or (ii) technical difficulties prevented the loading of the micro-data from SEDLAC. The 16 countries analyzed are: Argentina, Brazil, Bolivia, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Peru, and Uruguay.

Our analysis focuses on a broad sample of workers covering both formal and informal sectors irrespective of job characteristics. However, we choose to restrict our sample only to wage workers, excluding self-employed workers whose reported income levels may not be comparable across countries (Duranton, 2016). Likewise, we exclude workers who report zero-income (mostly family helpers in the sectors of agriculture and retail trade). Our final sample comprises 4,083,256 employed wage workers ages 14 to 65. A worker’s wage is taken to be the nominal hourly wage earned

---

<sup>17</sup>A possible third concern is that of reverse causality - i.e. that density or human capital in a city can be determined by its productivity premium. Using lagged right hand side variables in equation 6 could partially address this issue, unless there is strong path dependence, as seems likely, in which case both the premiums and the right hand side variables could be determined by a city’s historical productivity premium. Additionally, market access measures used exclude a city’s own market to mitigate reverse causation concerns.

<sup>18</sup> <https://data.worldbank.org/data-catalog/sedlac>

in the primary occupation.<sup>19</sup> We also perform estimation considering an alternative narrower sample of workers which is restricted to *prime age* males - i.e. males aged 20-55 - working in the private sector (see table B1 in the appendix).<sup>20</sup> In general, we find very similar results for both our broad and narrow samples.

For Brazil, the coding of administrative units in SEDLAC did not allow us to identify specific locations and map them to variables from other sources used in the second stage. Thus, we take micro-data on workers from the population census sample for 2000 that is available in IPUMS International.<sup>21</sup> We perform our own harmonization of the IPUMS International data for Brazil with the SEDLAC data for the other 15 countries in our sample.

For our second stage (equation 5) analysis of the determinants of the city premium, we further carefully match the harmonized survey data from SEDLAC and IPUMS with data from a LAC Geo-spatial Database that was constructed for the World Bank by the University of Southampton’s GeoData Center (Branson et al., 2017).<sup>22</sup> This database was constructed with the specific purpose of aligning with the identifiers for sub-national areas we have in SEDLAC. This dataset includes information on a location’s population density (population per km<sup>2</sup> of admin unit), aggregate stock of human capital, market access, terrain ruggedness (elevation variation calculated as in Nunn and Puga (2012)), mean air temperature, and total precipitation.<sup>23</sup>

Gridded Population of the World (GPW) v4 global 1 km population count data for 2015 was used to determine the population of locations, which correspond to administrative units, specified in table 1. UN adjusted measurements were used, which adjust raster cell values so that when summed to the national level they are consistent with UN population estimates. Robustness tests were carried out with other population data sources, such as WorldPop and Landscan-2012.<sup>24</sup> Similar results were obtained.

Average years of schooling by administrative unit were estimated using the household surveys

<sup>19</sup>Wages are measured at 2005 Purchasing Power Parity (PPP) exchange rates. In SEDLAC, rural wages are also increased by 15 percent to capture differences in rural-urban prices (CEDLAS and The World Bank, 2014, p 23). In our analysis, we undo this by multiplying the mean hourly wage of a rural worker by the factor 0.8695.

<sup>20</sup>Bacolod et al. (2009), Glaeser and Resseger (2010), Duranton (2016), and Chauvin et al. (2017) focus only on prime-age males.

<sup>21</sup>The second stage results for the pooled sample excluding Brazil are very similar to those including Brazil.

<sup>22</sup>This database was constructed as input into the World Bank’s LAC Flagship Report “Raising the Bar for Productive Cities in Latin America and the Caribbean” (Ferreira and Roberts, 2018).

<sup>23</sup> The market access variable attempts to capture the spillovers of agglomeration of other locations in the focal city. The construction of the measure is similar to a market potential measure, but it leaves the focus economy size put and focuses on the effect of proximity to other markets. This helps to distinguish market access from our measure of agglomeration, while also mitigating reverse causality concerns. Following Henderson and Wang (2007), we define market access for city  $l$  as

$$MA_l \equiv \sum_{k=1|k \neq l}^N \frac{n_k}{d_{lk}^2} \quad (7)$$

where  $n_k$  is the size of city  $k$  measured by population,  $d_{lk}$  is the estimated travel time by automobile between the focal city  $l$  and each of the *other* cities, indexed by  $k$ , and  $N$  is the total number of cities in the country. Travel times incorporate average road congestion, current speed limits, and road type details between city centroids (Branson et al., 2017).

<sup>24</sup>GPW v4 can be found at <http://sedac.ciesin.columbia.edu/data/collection/gpw-v4>, World Pop at <http://www.worldpop.org.uk/>, and Landscan-2012 at <http://web.ornl.gov/sci/landscan/>.

Table 1: Summary of data

Country	Admin. Unit (Division level)	Survey years	1st-stage observations	2nd-stage # locations	% national population
Guatemala	Departamento (1)	2006, 2011, 2014	58,030	22	100
Chile	Comuna (3)	2000, 2003, 2006, 2009, 2011, 2013	405,221	328	99.4
Honduras	Municipio (2)	2004 - 2011, 2013	43,261	275	97.8
Peru	Distrito (3)	2000 - 2014	459,915	1,428	96.3
Costa Rica	Distrito (3)	2002 - 2006, 2008 - 2010	117,517	401	94
El Salvador	Municipio (2)	2014	27,117	220	93.2
Nicaragua	Municipio (2)	2001, 2005	24,730	116	90.2
Ecuador	Parroquia (3)	2005 - 2012, 2014	237,801	637	88.4
Uruguay	Municipio (2)	2005	15,915	13	87.6
Bolivia	Provincia (2)	2006, 2012	14,874	73	83.9
Dominican Republic	Municipio (2)	2000 -2014	131,608	207	80.8
Brazil	Municipio (2)	2000	1,809,596	1,488	80.7
Panama	Provincia (1)	2003 - 2008, 2010 - 2013	186,956	9	80
Argentina	Aglomerado (1)	2000 - 2011, 2013	225,261	13	75.9
Mexico	Municipio (2)	2000, 2002, 2008, 2010	94,105	430	64.6
Colombia	Municipio (2)	2008, 2009, 2010	231,349	212	60.9
Total			4,083,256	5,872	77.9

from SEDLAC (and IPUMS for Brazil). Workers under 14 years of age are excluded from the calculation, to prevent confounding between a large share of young population and low educational averages. Robustness tests were carried out using instead percentage of workers with tertiary education, following [Behrens et al. \(2014\)](#) and [Chauvin et al. \(2017\)](#).

For most countries, the locations that our data set captures cover more than 80% of the total national population, as calculated with GPW 2015 and shown in table 1. Table 2 shows summary statistics of the data used in the 1st-stage estimation calculated using survey weights. As expected when working with developing countries, the percentage of workers with complete higher education is rather small (an overall average of 11.3 percent in our sample vs 35 percent as the average of OECD (Organisation for Economic Co-operation and Development) countries and 43 percent for the US, as reported by the OECD for 2017 ([OECD, 2017](#))). Similarly, the working population is younger (35.8 in our sample vs a median of 41 years in the US for 2014 according to the Bureau of Labor Statistics, [BLS \(2017\)](#)).

Table 3 provides summary statistics of the specific variables used in the second stage.<sup>25</sup>

<sup>25</sup>Population density for Argentina seems significantly lower than for the rest of the sample. This is explained partially by the fact that we are using an administrative level (aglomerados) for Argentina, which is generally slightly larger than the administrative units used in other countries. This choice is due to data constraints.

Table 2: Summary Statistics 1st stage

Country	# of observations	% Males	% Married	Mean age	Mean years of schooling	% Workers higher education
Argentina	225,261	59.5	60.6	38.5	10.7	17.4
Bolivia	14,874	60.2	66.6	38.2	9.5	17.1
Brazil	1,809,596	57.5	56.6	33.0	7.8	8.3
Chile	405,221	65.1	61.9	39.7	11.2	13.0
Colombia	231,349	56.2	57.7	37.7	8.8	15.6
Costa Rica	117,517	66.9	59.8	36.2	8.9	10.9
Dominican Republic	131,608	66.0	57.7	36.8	8.8	11.1
Ecuador	237,801	64.1	54.2	38.8	9.4	12.6
Guatemala	58,030	66.8	63.7	34.9	6.0	2.70
Honduras	43,261	63.1	58.9	35.9	6.6	5.3
Mexico	94,105	62.4	61.9	36.4	9.4	13.6
Nicaragua	24,730	63.0	60.3	35.5	6.6	8.1
Panama	186,956	64.7	61.8	38.0	10.7	11.2
Peru	459,915	60.7	60.7	37.8	9.7	17.6
El Salvador	27,117	59.3	58.3	37	8.3	5.3
Uruguay	15,915	54.2	61.0	39.9	10.6	13.6
All	4,083,256	60.2	58.4	35.8	8.60	11.3

Table 3: Summary Statistics Administrative units for 2nd stage<sup>a</sup>

Variables	Statistic	All	ARG	BOL	BRA	CHL	COL	CRI	DOM	ECU
Population density (people/km <sup>2</sup> )	Mean	473	19	32	341	434	401	1,357	469	181
	Median	55.7	12	12	61	28	69	171	115	59
	Max	25,821	68	356	13,393	8,652	9,039	16,355	11,188	4,801
	Min	0	4	0.6	0.2	0.1	3	3	6	1
	S.D.	1,737	20	63	1,139	1,375	1,092	2,564	1,292	426
Average years of schooling (years)	Mean	6.8	10.6	6.8	5.4	9.5	7.2	7.7	6.9	6.9
	Median	6.7	10.6	7.0	5.6	9.4	6.9	7.5	6.9	6.8
	Max	14.2	11.1	11.6	9.8	13.5	10.9	14.2	11.3	11.8
	Min	1.4	9.9	1.5	1.8	5.4	3.6	3.1	3.0	2.8
	S.D.	2.0	0.4	2.3	1.5	1.1	1.4	1.8	1.4	1.6
Share of workers w/ higher education (%)	Mean	5.3	11.5	6.8	4.4	7.1	5.8	6.8	4.4	2.9
	Median	3.5	11.3	4.8	3.6	5.6	4.8	4.0	3.9	1.7
	Max	55.8	14.3	24.1	24.3	37.5	23.2	55.8	18.7	22.4
	Min	0.0	9.2	0.0	0.3	0.0	0.0	0.0	0.0	0.0
	S.D.	5.8	1.7	6.4	3.2	5.0	4.3	7.6	3.3	3.6
Market access index (ln)	Mean	14.2	18.4	13.2	14.5	14.1	15.6	15.5	14.3	13.9
	Median	13.6	19.6	12.7	14	13.7	14.7	15.4	14.2	13.6
	Max	36.2	25.5	19.5	36.2	22.9	26.5	22	22.3	25
	Min	8.1	11.3	9.1	9.1	8.1	11.5	11.1	11.8	10.9
	S.D.	2.5	5.7	2.7	2.6	2.9	2.9	2.5	1.8	1.7
		GTM	HND	MEX	NIC	PAN	PER	SLV	URY	
Population density (people/km <sup>2</sup> )	Mean	282	89	1,024.40	128	53	478	500	193	
	Median	167	66	128	65	50	24	213	8	
	Max	1,531	835	19,743	1,584	164	25,821	6,903	2,289	
	Min	20	5	0.3	7	5	0	30	5	
	S.D.	319	102	2,700	202	46	2,346	996	630	
Average years of schooling (years)	Mean	5.2	5.2	7.8	5.1	8.9	7.4	7	9.2	
	Median	5.1	5.2	7.9	5.2	8.9	7.2	6.7	9.0	
	Max	7.9	9.1	13.4	9.1	10.8	13.2	12.7	10.4	
	Min	3.8	1.4	2.4	1.5	6.7	1.6	2.8	8.7	
	S.D.	0.9	1.2	1.9	1.8	1.2	2.0	1.6	0.5	
Share of workers w/ tertiary education (%)	Mean	1.4	1.1	6.9	3.3	6.6	7.3	2.4	6.2	
	Median	1.0	0.6	5.3	1.9	6.8	4.3	1.4	5.8	
	Max	6.5	8.2	44.9	19.3	9.8	51.4	25.7	10.9	
	Min	0.5	0.0	0.0	0.0	3.3	0.0	0.0	1.6	
	S.D.	1.3	1.5	6.2	3.8	2.2	8.1	3.3	2.2	
Market access index (ln)	Mean	13.7	13.4	15.3	13	12.9	13.3	13.3	12.6	
	Median	13.1	13.2	14.7	12.6	11.6	12.7	13.1	12.1	
	Max	18.8	19.7	24.6	18.1	22.4	28.5	18.7	21.5	
	Min	10.4	11.3	9.8	9.8	9.9	8.9	9.8	9.2	
	S.D.	1.9	1.4	2.9	1.7	4.0	2.3	1.8	3.3	

<sup>a</sup>ARG = Argentina, BOL = Bolivia, BRA = Brazil, CHL = Chile, COL = Colombia, CRI = Costa Rica, DOM = Dominican Republic, ECU = Ecuador, GTM = Guatemala, HND = Honduras, MEX = Mexico, NIC = Nicaragua, PAN = Panama, PER = Peru, SLV = El Salvador, URY = Uruguay.

## 4 Empirical results

### 4.1 1st stage results - city productivity premiums

We observe large differences in average nominal wages across sub-national areas (figures 1a and 2a), which typically correspond to level 2 administrative units or municipios. Moreover, it is notable that the highest levels of nominal wages, and, hence, productivity, tend to be observed in sub-national areas that correspond to major cities such as Bogota, Buenos Aires, Lima, Mexico City, Panama City, Santa Cruz, Santiago, and Sao Paulo. After controlling for worker skills  $s_{i,t}$  as shown in equation 4, we get the estimated premium  $\hat{B}_{l,t}$  by location (figures 1b and 2b). As discussed previously, the observable worker characteristics that we control for are years of schooling, age and its square, gender, and marital status. After these controls are introduced to account for sorting, we see a significantly lower variation in the premium across cities than for the simple average nominal wages. Figure 1 and table 4 show the dampening of the variation by reporting productivity premiums before and after controlling for sorting on observable worker characteristics. This indicates that compositional differences in the workforce associated with the sorting of workers into different places is a major factor driving productivity differences. The overall standard deviation is reduced by 80% for all countries pooled. Tables A1 and A2 show 1st-stage results by country. Additionally, we have performed estimation including job characteristics (available upon request from the authors) as a robustness check.<sup>26</sup> Results are robust.

Sorting does not, however, tell the full story of productivity differences between places. Even after controlling for observable characteristics, some important variation in productivity premiums remains across cities within countries.<sup>27</sup> This can be seen more clearly in figure 3, which shows box plots of estimated city premiums for the sample of 16 countries. The variation in premiums is particularly pronounced in Peru, Ecuador, Costa Rica and Honduras, whereas it is much more muted in, for example, Guatemala and Uruguay.

We explain this remaining variation in premiums net of skill sorting with variables that represent the three urban success theories discussed above.

### 4.2 2nd stage results - the roles of the three theories of urban success

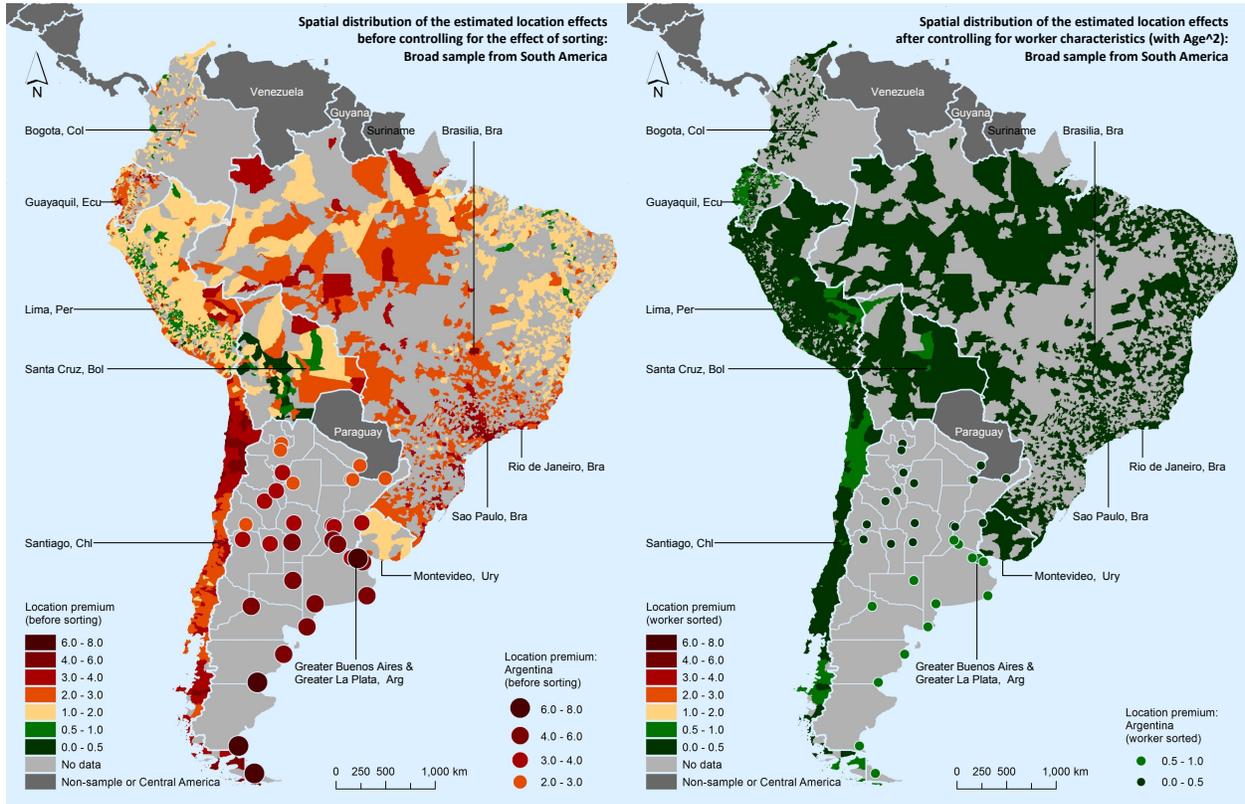
Consistent with the three theories of urban success (i.e. the theories of agglomeration economies, human capital externalities, and market access), we see that estimated city premiums are positively and significantly correlated with population density, average years of schooling among the working-age population, and market access (figure 4) across our sample of 16 LAC countries. With respect to population density, which provides our measure of agglomeration, the estimated elasticity of the city premium is 5.3 percent. This is higher than corresponding estimates that have been reported in the literature for developed countries, but lower than the estimates that have been reported

---

<sup>26</sup>Job characteristics are 1 digit ISIC sectoral dummies, informal sector dummies, and controls for size of firm.

<sup>27</sup>We still see a range of approximately 2 dollars per hour between the 5<sup>th</sup> and 95<sup>th</sup> percentiles of the distribution of the location premiums across cities for all countries pooled (see table 4).

Figure 1: City premiums in South America<sup>a</sup>

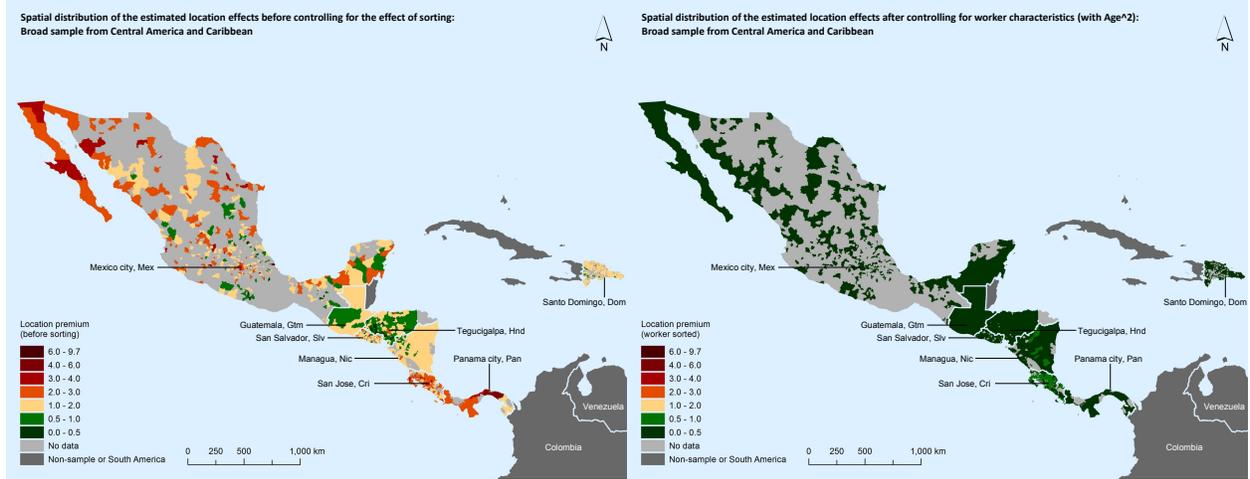


(a) City premiums before controlling for sorting.

(b) City premiums after controlling for sorting.

<sup>a</sup> We employ a separate point layer for Argentina to retain all locations for which SEDLAC allows us to estimate city premiums. Unlike other countries in our sample, these correspond to major cities/urban agglomerations, e.g. City of Buenos Aires and Greater La Plata, for which we lack a GIS shapefile of administrative boundaries. City premiums in the maps are displayed as  $\exp(\hat{B})$  and expressed in 2005 PPP exchange rates, where  $\hat{B}$  is the estimated city fixed effect in equation 4. In (a), the city premium (before sorting) is the premium without controlling for observable worker characteristics, but controlling for survey-year fixed effects. In (b), the city premium is the premium after controlling for observable worker characteristics (age, age<sup>2</sup>, years of schooling, gender, marital status) and survey-year fixed effects. Same considerations apply for figure 2.

Figure 2: City premiums in Central America<sup>a</sup>



(a) City premiums before controlling for sorting.

(b) City premiums after controlling for sorting.

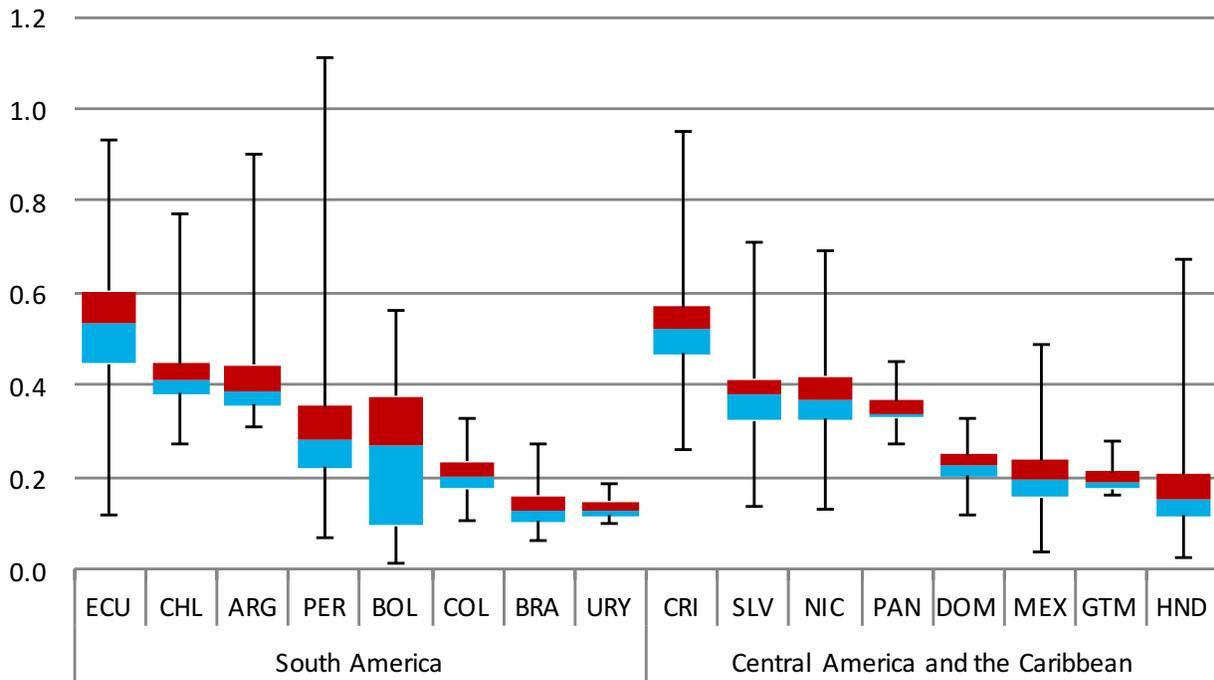
<sup>a</sup> Similar considerations apply as in figure 1.

Table 4: Standard deviation of productivity premiums before and after controlling for sorting<sup>a</sup>

country	ARG	BOL	BRA	CHL	COL	CRI	DOM	ECU	
$\sigma_{\hat{\omega}}$	1.370	0.886	0.937	0.864	0.453	0.987	0.296	0.635	
$\sigma_{\hat{B}_l}$	0.152	0.152	0.037	0.071	0.040	0.097	0.035	0.118	
	GTM	HND	MEX	NIC	PAN	PER	SLV	URY	Total
$\sigma_{\hat{\omega}}$	0.270	0.446	0.914	0.351	0.712	0.729	0.490	0.338	0.950
$\sigma_{\hat{B}_l}$	0.030	0.072	0.064	0.083	0.056	0.109	0.083	0.023	0.161

<sup>a</sup> $\sigma_{\hat{\omega}}$  is the standard deviation of the average wage net of survey-year fixed effects (estimated through the location fixed effect in a regression similar to equation 4 but without any worker observable characteristics - i.e. without the component  $\beta \ln s_{i,l(i),t}$ ).  $\sigma_{\hat{B}_l}$  is the standard deviation of the estimate location premium  $\hat{B}_l$  from estimating equation 4. Variances are taken after exponentiating the estimated parameters so that they are expressed in the hourly wages units. ARG = Argentina, BOL = Bolivia, BRA = Brazil, CHL = Chile, COL = Colombia, CRI = Costa Rica, DOM = Dominican Republic, ECU = Ecuador, GTM = Guatemala, HND = Honduras, MEX = Mexico, NIC = Nicaragua, PAN = Panama, PER = Peru, SLV = El Salvador, URY = Uruguay.

Figure 3: Variation of city premiums after controlling for sorting <sup>a</sup>.



<sup>a</sup> This chart is organized in descending order of the median level of estimated city premium within each sub-region. Estimated city premium measures sub-national variations in underlying productivity after controlling for observable worker characteristics within the broad sample (all wage/salary employees aged from 14 to 65 years old). The upper and lower caps respectively indicate the maximum and the minimum estimated city premiums for each country. The bottom of the blue box, the border of two colors, and the top of the red box respectively depict the first quartile, the median, and the third quartile of the estimated city premiums in each country. ARG = Argentina, BOL = Bolivia, BRA = Brazil, CHL = Chile, COL = Colombia, CRI = Costa Rica, DOM = Dominican Republic, ECU = Ecuador, GTM = Guatemala, HND = Honduras, MEX = Mexico, NIC = Nicaragua, PAN = Panama, PER = Peru, SLV = El Salvador, URY = Uruguay.

for China and India. Using comparable regression specifications, [Chauvin et al. \(2017\)](#) report an elasticity of nominal wages with respect to population density of 4.6 percent for US Metropolitan Statistical Areas (MSAs), 19.2 percent for a sample of Chinese provincial and prefectural level cities, and 7.6 percent for the urban parts of Indian districts. Meanwhile, the estimated elasticities of the city premium with respect to average years of schooling and market access for our sample of 16 LAC countries are 62 percent and 4 percent respectively.

While the above correlations are suggestive of the ability of the three theories of urban success to explain variations in sub-national underlying productivity, it is also the case that population density, average years of schooling and market access are all positively correlated with one another. Hence, more densely populated areas tend to exhibit both higher average years of schooling and market access.<sup>28</sup> To disentangle the relative importance of the three theories of urban success, we estimate equation 5. Table 5 shows the results. Columns 1-3 report the results of regressions of the estimated city premium on population density, average years of schooling, and our measure of market access. In these regressions, we also control for mean air temperature, terrain ruggedness and total precipitation, physical geographical conditions that could be correlated with both city premiums and the main explanatory variables.

In column 1, we see that the estimated elasticity of the city premium with respect to population density declines to 4.9 percent once we control for geographic conditions. Although this estimate remains statistically significant, it is less than the estimated elasticity of 5.3 percent that we reported above when not controlling for geography. Once we also introduce average years of schooling, in column 2, however, both the estimated size and statistical significance of the elasticity of the city premium with respect to population density fall dramatically. Including market access in column 3 then leads population density to lose its statistical significance completely.<sup>29</sup>

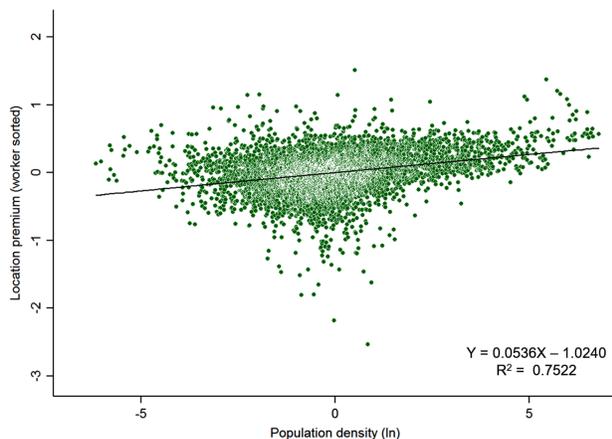
Although both average years of schooling and market access are significant, the former has a much larger effect on underlying productivity. While an increase in average years of schooling from the 25<sup>th</sup> to the 75<sup>th</sup> percentile in our sample implies an estimated productivity increase of 23.4 percent, moving from the 25<sup>th</sup> to the 75<sup>th</sup> percentile for market access implies a productivity increase of only 4.1 percent. The point estimate for the elasticity of underlying productivity with respect to average years of schooling is 0.57. This implies, at the overall average level of education of 8.6 years, that an increase in one year of education is associated with an increase in wages of

---

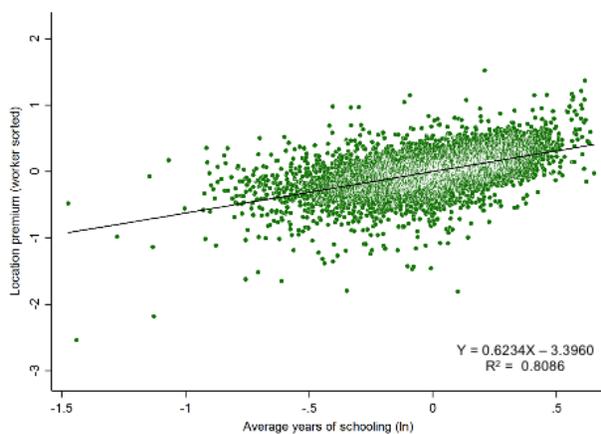
<sup>28</sup>The Pearson correlation coefficient for population density and average years of schooling is 0.30, while that for population density and market access is 0.67. For average years of schooling and market access, the correlation coefficient is 0.32. All estimated correlation coefficients are significant at the 5 percent level.

<sup>29</sup>Robustness tests were implemented to test a possible non-linear relationship of the variables. Results are available from the authors. We find a positive quadratic relationship (a U shape), which contrasts with the inverted U shape more commonly found in the literature ([Desmet and Rossi-Hansberg, 2014](#)). Instead of the optimal density of a city ([Au and Henderson, 2006](#)) after which cities become too congested, this would suggest a minimum density threshold in cities after which agglomeration economies start emerging (or overcoming congestion costs). This could be affected by relatively low levels of infrastructure and institutional quality compared to those in developed countries. However, these results could be a consequence of pooling different years and countries. Future work testing this relationship independently by country could shed light on the agglomeration vs congestion costs relationships in developing countries, which the literature has not fully addressed simultaneously ([Desmet and Rossi-Hansberg, 2013](#); [Desmet and Henderson, 2015](#); [Duranton, 2016](#); [Akbar and Duranton, 2017](#); [Hanlon and Tian, 2015](#)).

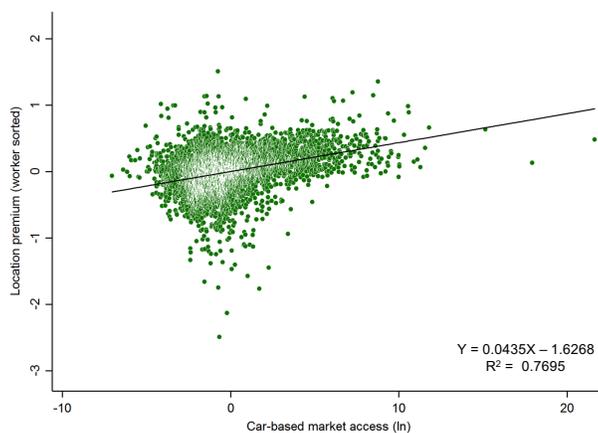
Figure 4: Correlation between city premium and: (a) population density; (b) average years of schooling; and (c) market access <sup>a</sup>



(a) Population Density.



(b) Human Capital.



(c) Market Access

<sup>a</sup> Scatterplots show the correlation between the estimated city premiums and the natural logs of population density, average years of schooling and market access controlling for country fixed effects. Hence, sub-national administrative areas are the units of observation and the correlations are estimated based on the within-country variation in the data.

Table 5: Results of regressions on the determinants of location premiums across sub-national areas<sup>a</sup>

Dependent variable:	1	2	3	4	5
Location premium (ln)					
Population density (ln)	0.049***	0.013*	0.005	0.023*	0.002
Average years of schooling (ln)		0.576***	0.574***		
% of WAP with higher education				0.021***	0.020***
Market access (ln)			0.015***		0.027***
Mean air temperature (ln)	0.03	0.044	0.051	0.036	0.045
Terrain ruggedness (ln)	-0.031**	-0.024***	-0.017	-0.026*	-0.024
Total precipitation (ln)	-0.028	-0.008	-0.01	-0.003	-0.001
Constant	-0.99***	-2.37***	-2.70***	-1.28***	-1.82***
n	5,750	5,750	5,050	5,750	5,050
$R^2$	0.757	0.814	0.831	0.785	0.804
Adj. $R^2$	0.756	0.813	0.830	0.785	0.803

<sup>a</sup> \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ . In all columns, country effects have been controlled for and standard errors have been clustered by country. In all columns, the dependent variable is the estimated location premium (measured in natural logs) from a series of country-level first-stage regressions after controlling for observable worker characteristics within the broad sample (all wage/salary employees aged from 14 to 65 years old) and survey-year fixed effects. Worker characteristics included age, age-squared, marital status, gender, and years of schooling. WAP stands for the working-age population. The source of the population data is the Gridded Population of the World (GPW), v4.

around 6 percent <sup>30</sup>

From the above, it seems that the theories of human capital externalities and, to a lesser extent, market access provide more important explanations of urban success in LAC than does the theory of agglomeration economies. Traditional agglomeration economies that are associated with population density, but with neither human capital nor market access, appear to be virtually non-existent. This could be linked to the high levels of population density that are characteristic of cities in LAC. One plausible hypothesis is then that the high levels of population density are leading to excessive congestion, which is in turn negating the positive externalities that are normally associated with urban density.<sup>31</sup>

<sup>30</sup>These are comparable to other measures of human capital externalities in the literature. For example, [Rauch \(1993\)](#) find that one year of average education increases wages by between 3 and 5 percent. [Acemoglu and Angrist \(2000\)](#) find a comparable growth of 7 percent. However, IV estimates are much smaller and not significant. The latter results exclude lower education levels. The aforementioned estimates come from approaches that generally do not control for market access and agglomeration effects like ours, but that incorporate methods that control for the endogeneity of average education.

<sup>31</sup>We must acknowledge some caveats about drawing this conclusion. It may be that, relative to our measures of human capital and market access, population density is a relatively poor measure of agglomeration. Indeed, boundaries of sub-national administrative units often conform only poorly to the *true* boundaries of cities (i.e. to true labor or housing market boundaries). To the extent that agglomeration can be poorly measured, this could

In addition to average years of schooling, we also report results in table 5 using instead the share of the working-age population who have completed higher education as a measurement of human capital (columns 4 - 5). This has been a preferred measure in the literature on human capital externalities, where it is argued it better captures the evidence that raising the top of the human capital distribution will generate learning spillovers but raising the bottom will not (see, for example, Glaeser (1999), Behrens et al. (2014) and Chauvin et al. (2017)). This amounts to arguing that workers only experience significant learning from the highly-educated. The point estimate for the corresponding elasticity with respect to percentage of workers with higher education is .02.<sup>32</sup> As can be seen, this alternative measure of human capital is also a highly statistically significant predictor of variations in the city premium. Comparing the results in columns 4 - 5 with those in columns 2 - 3, however, we can also see that our regressions fit better when using average years of schooling as the measure of human capital, and that the estimated effect is orders of magnitude higher. Given initial lower levels of education among workers than in developed countries, additional years of education, even if at pre-tertiary levels, seem to have strong externalities in these developing countries. In other words, raising the bottom seems also very important here.

As discussed before, as an alternative to the two-stage approach adopted so far, many studies instead adopt a single-stage approach in order to have more precise confidence intervals on the parameters estimated. The concern is that the second stage regressions use estimated productivity premiums but do not incorporate their estimation error (see, inter alia, Duranton (2016); Chauvin et al. (2017)). As table 6 shows, when we adopt this approach, we obtain higher and significant estimated coefficients on population density than those reported in table 5 using the two-stage approach. Nevertheless, the overall qualitative picture remains the same, especially when it comes to comparing the difference in elasticities between measures of population density and average years of schooling. As before, the estimated elasticity of a worker’s nominal wage with respect to the population density of the location in which she lives drops rather drastically as we introduce, first, a location’s overall average years of schooling, and then its level of market access (see columns 2 and 3 in table 6 respectively). At the same time, its level of statistical significance declines. Whether the variable becomes statistically insignificant at all conventional levels (i.e. at all levels up to the 10 percent level) depends crucially on how we cluster the standard errors. When the standard errors are clustered at the area-level (reported in table 6), population density remains significant at the 5 percent level. It may be argued, however, that this is too restrictive because it rules-out correlation between errors for workers who live in, for example, neighboring areas. When we instead cluster standard errors at the national level, population density is insignificant even at the 10 percent level.<sup>33</sup>

We also report single-stage results with the alternative measure of human capital in columns 4

---

provide another explanation of why we find no effect of population density having controlled for average years of schooling and market access.

<sup>32</sup>Moretti (2004a) find an elasticity of approximately 0.011 between wages and share of college graduates in the US. Furthermore Moretti (2004c) finds that this result is consistent with the elasticity estimated using TFP.

<sup>33</sup>The correct level of clustering for the standard errors likely lies between the area and national levels. Results based on clustering at the national level are available upon request from the authors.

and 5 based on the clustering of standard errors at the area level. In this case, population density continues to remain significantly positive at the 1 percent level even after controlling for both the share of working-age population who possess higher education and market access.

Table 6: Results of regressions on the determinants of underlying productivity variations based on the single-stage approach<sup>a</sup>

Nominal hourly wage (ln)	1	2	3	4	5
Population density (ln)	0.057***	0.024***	0.012**	0.035***	0.017***
Average years of schooling (ln)		0.636***	0.605***		
% of WAP with higher education				0.015***	0.014***
Market access (ln)			0.013***		0.016***
Mean air temperature (ln)	0.013	0.018	0.02	0.018	0.022
Terrain ruggedness (ln)	0	0.002	0.006	-0.005	0
Total precipitation (ln)	-0.044***	-0.027***	-0.025***	-0.035***	-0.033***
Constant	-2.02***	-3.41***	-3.49***	-2.09***	-2.27***
n	4,000,142	4,000,142	3,766,690	4,000,142	3,766,690
$R^2$	0.337	0.346	0.349	0.343	0.346
Adj. $R^2$	0.337	0.346	0.349	0.343	0.346

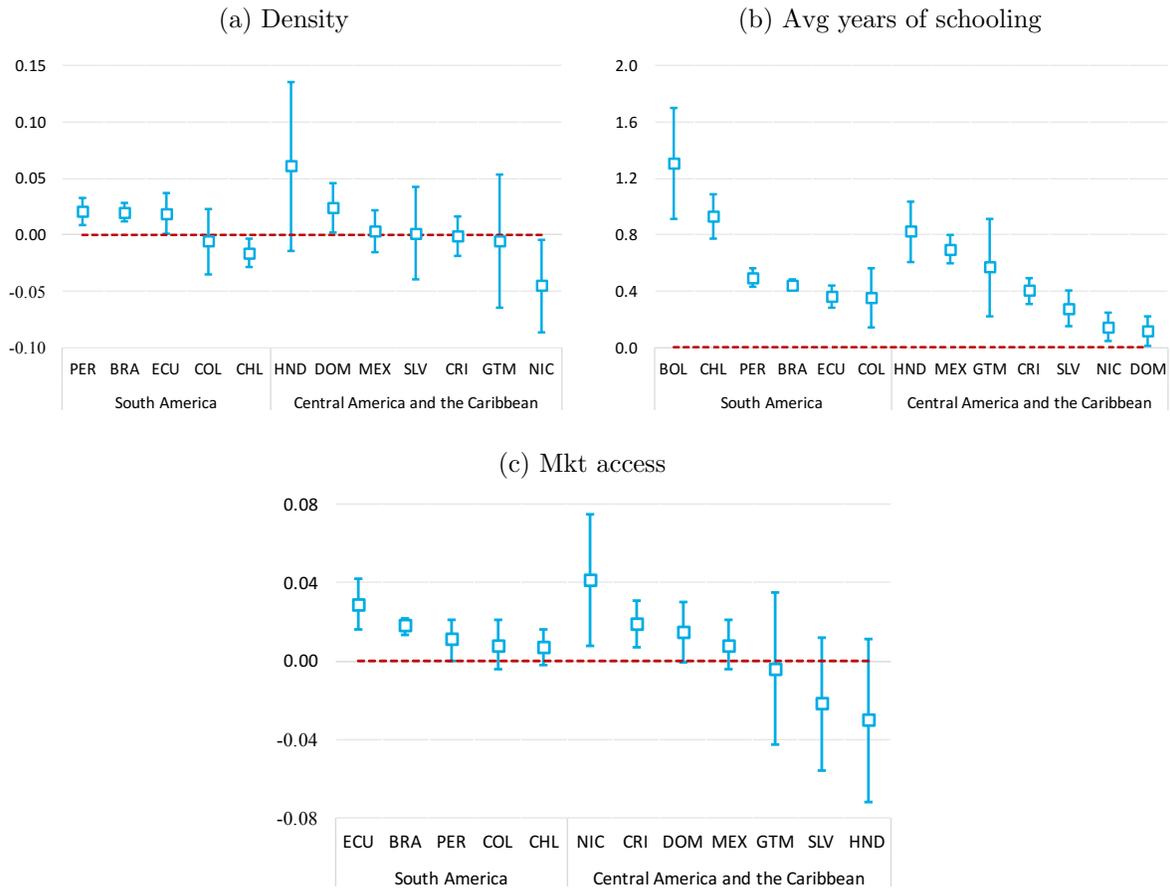
<sup>a</sup> \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ . In all columns, country effects have been controlled for and standard errors have been clustered by location. In all columns, the dependent variable is the worker's nominal wage, as shown in equation 6. Controls include observable worker characteristics within the broad sample (all wage/salary employees aged from 14 to 65 years old) and survey-year fixed effects. Worker characteristics include age, age-squared, marital status, gender, and years of schooling. We also include interactions between the worker characteristics and country fixed effects. The source of the population data is the Gridded Population of the World (GPW), v4.

#### 4.2.1 Country heterogeneity

The above results on the relative importance of agglomeration economies, human capital externalities, and market access are based on estimates for all 16 LAC countries pooled. However, there also exists considerable heterogeneity in estimated effects across the individual countries. Figure 5 shows the summarized coefficients for the 2nd stage regression when equation 5 is estimated separately for each country.<sup>34</sup> The coefficient reported is obtained from estimating equation 5 and the 95% confidence intervals are obtained from the coefficient distribution. There is a large difference in the magnitude of the estimated elasticities with respect to density, average years of schooling and market access (notice the figures use different scales for the vertical axes).

<sup>34</sup>Figure 5 does not show results for all 16 countries in our overall sample because of extremely wide confidence intervals or, as in the case of Panama, a lack of sufficient observations. None of the omitted coefficients is statistically significant. The points estimates are shown for all countries in table A3.

Figure 5: Heterogeneity in coefficients from 2<sup>nd</sup> stage.<sup>a</sup>



<sup>a</sup> Figures show the estimated elasticities between subnational underlying productivity and population density, average years of schooling and market access respectively, derived from individual country-level regressions that follow equation 5. The squares represent the point estimates, while the upper and lower caps indicate the upper and lower bounds of the 95 percent confidence intervals. Results are sorted by region and then by descending order of the estimated elasticities. ARG = Argentina, BOL = Bolivia, BRA = Brazil, CHL = Chile, COL = Colombia, CRI = Costa Rica, DOM = Dominican Republic, ECU = Ecuador, GTM = Guatemala, HND = Honduras, MEX = Mexico, NIC = Nicaragua, PER = Peru, SLV = El Salvador, URY = Uruguay. This figure omits countries whose coefficients have extremely wide confidence intervals due to small numbers of subnational areas or, as with Panama, a lack of sufficient observations (only 9 locations) to permit estimation. None of the omitted coefficients are statistically significant. The points estimates are shown for all countries in table A3.

Although, using the pooled data we find no evidence of significant agglomeration economies, the estimated elasticity of the city premium with respect to population density is both positive and statistically significant for Peru, Brazil, Ecuador, and Dominican Republic, as exemplified by the estimated 95 percent confidence intervals, which are shown by the blue lines in panel (a) of figure 5. In contrast, we estimate a significant negative elasticity for Chile and Nicaragua.

Meanwhile, for market access, the positive effect estimated in the pooled data is mainly driven by four countries: Ecuador, Brazil, Nicaragua, and Costa Rica. For the remaining 10 countries for which we report effects in figure 5, the impact of market access on city premium is not significantly different from zero.

Finally, regarding human capital, the effect of average years of schooling on underlying productivity is positive and statistically significant for all countries. Even here, the estimated strength of human capital externalities varies dramatically across countries. Hence, we estimate extremely strong human capital externalities in Bolivia, but comparatively weak externalities in El Salvador, Nicaragua and the Dominican Republic. The effect of the percentage of workers with completed higher education is also positive in most countries, but it is not significant for some of the countries in Central America (Dominican Republic, Honduras and Nicaragua) and in Bolivia.<sup>35</sup> These elasticities are significantly larger than those corresponding to the variables representing the other two success theories.

We hypothesize that the heterogeneity in human capital coefficients may be related to educational quality. We estimate the linear relationship between the estimated human capital coefficients and PISA grades and find a significant positive relationship that would be consistent with this hypothesis (figure 6).<sup>36</sup> Unfortunately, not all the countries in our sample take the PISA exam, so we cannot test this relationship for the full sample. Additionally, we report the relationship between the human capital externalities effect and educational expenditure per capita<sup>37</sup>, which supports the idea that this quality, and hence the effect of human capital externalities, can be increased with public expenditure on education. The extent to which the estimated variations in the strength of human capital externalities are attributable to educational quality differences is an important area of future research.

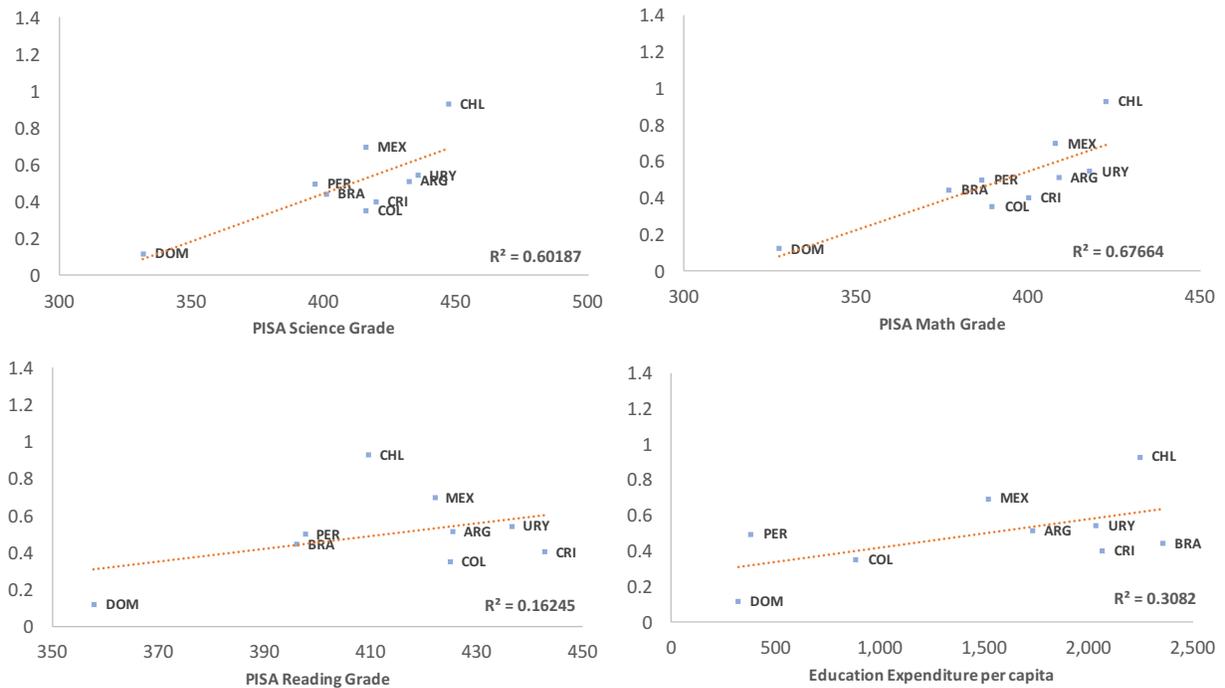
---

<sup>35</sup>All results based on the percentage of workers with completed higher education are available upon request from the authors.

<sup>36</sup>PISA (Programme for International Student Assessment) is a triennial international survey which aims to evaluate education systems worldwide by testing the skills and knowledge of 15 and 16-year-old students. In the 2015 version, 28 million students in 72 countries took this exam. Only students at school are tested, not home-schoolers. A higher grade represents better performance. This exam is administered by the OECD. To fulfill OECD requirements, each country must draw a sample of at least 5,000 students (although most countries use a much larger sample). The dotted line is a fitted line.

<sup>37</sup>This measure is obtained by dividing the current \$US expenditure on education by the population of inhabitants younger than 14 years old. Education expenditure refers to the current operating expenditures in education, including wages and salaries and excluding capital investments in buildings and equipment and comes from the United Nations Statistics Division's Statistical Yearbook, and the UNESCO Institute for Statistics online database. Population is based on the United Nations Population Division's World Population Prospects.

Figure 6: Relationship between estimated coefficients for human capital (average years of schooling) and educational quality.<sup>a</sup>



<sup>a</sup> The vertical axis show the estimated elasticities between subnational underlying productivity and city human capital derived from individual country-level regressions that follow equation 5. The horizontal axes show the 2015 science, math and reading grade in the Programme for International Student Assessment (PISA) exams. The dotted line is a fitted line. The panel in the low right corner shows educational expenditure per capita, obtained by dividing the current \$US expenditure on education by the population of inhabitants younger than 14 years old using data from the United Nations Statistics Division's Statistical Yearbook, UNESCO's Institute for Statistics, and United Nations Population Division's World Population Prospects.

## 4.2.2 Demographic and sector heterogeneity

In addition to heterogeneity across countries, it is also fruitful to ask whether, following the example of [Duranton \(2016\)](#) for Colombia, the elasticities of the city premium with respect to population density, average human capital and market access are heterogeneous across different sub-groups of workers. To explore this question, we perform a series of (1st and 2nd stage) regressions using sub-groups of workers for the following four dimensions that have been drawn from our broad sample (which consists of all employed wage workers aged 14 to 65):

1. Young (14-35 years old) versus Old (36-65 years old)
2. Male versus Female
3. Private versus Public Sector
4. Formal versus Informal<sup>38</sup>

For age groups, we use 35 years as the dividing line between Young and Old because this is roughly the mean age of workers across all 16 countries. Meanwhile, given that our sample already excludes all self-employed workers and workers who report zero income, we define informal workers as those who work for firms that have 5 or fewer employees. Our regressions for Private -v- Public Sector and Formal -v- Informal workers exclude Brazil. This is because, for Brazil, no public-sector workers were left in our original broad sample after data-cleaning, while IPUMS does not provide data that allow us to distinguish between formal and informal workers in a manner akin to that for other countries, for which the data instead come from SEDLAC.

Table 7 summarizes the results of our regressions (full results are available upon request from the authors). The most striking differences in estimated effects come when comparing Private -v- Public Sector workers and Formal -v- Informal workers. In both cases, the estimated elasticities of the city premium with respect to average years of schooling and market access are higher for private sector and formal workers than they are for public sector and informal workers respectively. For human capital externalities, the difference in the impact is the largest when comparing Private and Public Sector subsamples. We expect wages to be more closely connected with productivity in less regulated Private labor markets than in the Public Sector (wages are a better measure of productivity in private than in the public sector)<sup>39</sup> which is consistent with the stronger association observed between human capital externalities and productivity premiums. Although they are

---

<sup>38</sup>SEDLAC provides two different indicators for whether a worker is to be considered informal ([CEDLAS and The World Bank, 2014](#)). The first, which is based on a *productive* definition of informality, identifies a worker as informal if *(s)he belongs to any of the following categories: (i) unskilled self-employed, (ii) salaried worker in a small private firm, (iii) zero-income workers*. Meanwhile, the second, which is instead based on a *legalistic or social protection* notion of informality, identifies a salaried worker as informal if *s(he) does not have the right to a pension linked to employment when retired*. We choose to rely on the indicator based on the productive definition of informality because this suffers from fewer missing observations. Because our sample already excludes both self-employed and zero-income workers, this amounts to equating informal employment with employment by very small private firms (5 employees or fewer).

<sup>39</sup>[Lucifora and Meurs \(2006\)](#) [Christofides and Michael \(2013\)](#) study wage gaps between private and public sectors in Europe, while [Panizza \(2001\)](#) studies them in Latin America. They all find a positive gap in favor of the public

Table 7: Heterogeneous effects of determinants of location premium productivity across sub-groups of workers<sup>a</sup>

	Young	Old	Male	Female	Private	Public	Formal	Informal
Population density (ln)	-0.001	0.004	0.006	-0.001	0.002	0.012	0.004	-0.002
Avg. years schooling (ln)	0.466***	0.580***	0.541***	0.495***	0.548***	0.172*	0.559***	0.401***
Market access (ln)	0.014***	0.01	0.011*	0.008	0.018***	-0.001	0.019***	0.004
Constant	-2.554***	-2.392***	-2.356***	-2.281***	-2.586***	-0.804***	-2.349***	-1.565***
n	3,756	3,757	3,758	3,754	3,758	3,440	3,758	3,732
Adj. $R^2$	0.79	0.744	0.689	0.675	0.717	0.68	0.668	0.687

<sup>a</sup> \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ . In all columns, country effects have been controlled for and standard errors have been clustered by country. In all columns, the dependent variable is the estimated location premium (measured in natural logs) from a series of country-level first-stage regressions after controlling for observable worker characteristics within the broad sample (all wage/salary employees aged from 14 to 65 years old) and survey-year fixed effects. Worker characteristics include age, age-squared, marital status, gender, and years of schooling. The source of the population data is the Gridded Population of the World (GPW), v4.

smaller than for Private -v- Public Sector and Formal -v- Informal workers, differences can also be observed in the estimated elasticity of the city premium with respect to average years of schooling for Young -v- Old and Male -v- Female workers. Old workers appear to experience stronger human capital externalities than Young workers, while Male workers benefit from stronger human capital externalities than Female workers.

In the case of market access, the estimated elasticities are close to zero for public sector and informal workers, and statistically insignificant. Market access benefits more the productivity of workers in Formal and Private firms. This would be consistent with these types of firms selling to other regions in the country more and benefitting more from cheaper inputs brought from elsewhere. Market access also has an insignificant impact for both Old and Female workers. However, the estimated elasticities in both cases are much closer to those estimated for Young and Male workers respectively, which are statistically significant at the 10 percent level or better.

Finally, population density exerts a negligible, and statistically insignificant, impact on city premiums for all sub-groups of workers.

## 5 Conclusions

This paper contributes to filling a knowledge gap on agglomeration effects and the determinants of sub-national spatial variations in productivity for developing countries by providing estimates of city

sector that cannot be fully explained by worker characteristics. [Brassiolo et al. \(2016\)](#) find that the gap remains even when other unobservable characteristic, like motivation for public service and effort, are taken into account, for the same countries that we study here (except for Dominican Republic, Guatemala, and Nicaragua). Furthermore, they show that the public sector in Latin America exhibits more non-wage benefits and stability, more stable wages as careers progress, fewer performance based contracts, and more non-market regulations that govern their evolution, which further suggest the structurally different relationship between productivity and wages in the private and public sectors.

wage premiums net of skill sorting for 16 countries in the Latin America and the Caribbean region. While two of the countries in our sample - Brazil and Colombia - have already been considered by [Chauvin et al. \(2017\)](#) and [Duranton \(2016\)](#) respectively - the remaining 14 countries have not, to our knowledge, been previously analyzed in the literature. We generate these estimates using a harmonized data set which contains information on both the nominal wages and characteristics of individual workers, and of the characteristics of the locations in which the workers live. The scale of our analysis with harmonized data is also a novel contribution. Furthermore, we test how the city premiums are related to variables that represent different theories of urban success: agglomeration economies, human capital externalities, and market access.

Variation in productivity premiums is largely explained by sorting on observable worker characteristics, including observed variation in skill levels. Once this sorting is taken into account, much of the variation in nominal wages across locations disappears. Nevertheless, some variation does remain and this variation is correlated with city characteristics. When the data from all 16 countries are pooled, a location's underlying productivity level, as captured by its city premium, shows a strong positive correlation with average human capital, suggesting the presence of human capital externalities. It also exhibits a positive correlation with market access. Agglomeration economies (measured through population density) disappear when human capital externalities and market access are introduced. There is some heterogeneity across countries and groups of populations that is also analyzed, but the strength of human capital externalities compared to the other factors is robust.

In summary, urban success in LAC is crucially dependent on both the existence of a strong overall stock of human capital and, at least for certain countries, good access to large consumer and supplier markets through transportation networks. In contrast, agglomeration economies, in their more traditional sense, seem to be relatively weak across much of the region, which may be related to a lack of infrastructure investment and poor city management, relative to the high levels of urban population density that prevail in the region.

Heterogeneity across countries is investigated. It remains true, across most countries, that agglomeration economies, when measured linearly, lose in a horse race with market access and human capital externalities. In particular, the association of productivity and human capital at the municipality level is heterogeneous, which could be a reflection of country differences in educational quality.

This paper responds to an interest in studying agglomeration effects in developing countries empirically. More than just extending the state of knowledge on the extensive margin, this paper studies relationships in developing countries that could be significantly different from the ones found in the literature for developed areas, where, for example, congestion costs could play a weaker role due to better infrastructure and institutions.

## References

- Acemoglu, D. and Angrist, J. (2000). [How large are human-capital externalities? Evidence from compulsory schooling laws.](#) *NBER macroeconomics annual*, 15:9–59.
- Akbar, P. and Duranton, G. (2017). [Measuring the Cost of Congestion in Highly Congested City: Bogotá.](#) CAF (Latin American Development Bank).
- Amiti, M. and Cameron, L. (2007). [Economic geography and wages.](#) *The Review of Economics and Statistics*, 89(1):15–29.
- Au, C.-C. and Henderson, J. V. (2006). [Are Chinese cities too small?](#) *The Review of Economic Studies*, 73(3):549–576.
- Bacolod, M., Blum, B. S., and Strange, W. C. (2009). [Skills in the city.](#) *Journal of Urban Economics*, 65(2):136–153.
- Bartelme, D. (2018). [Trade costs and economic geography: evidence from the US.](#) Working paper.
- Behrens, K., Duranton, G., and Robert-Nicoud, F. (2014). [Productive cities: Sorting, selection, and agglomeration.](#) *Journal of Political Economy*, 122(3):507–553.
- BLS (2017). Age of the labor force. [https://www.bls.gov/emp/ep\\_table\\_306.htm](https://www.bls.gov/emp/ep_table_306.htm).
- Branson, J., Campbell-Sutton, A., A., H., D., H., and C., H. (2017). [A Geospatial Database for Latin America and the Caribbean.](#) Technical report, University of Southampton.
- Brassiolo, P., Sanguinetti, P., Álvarez, F., Quintero, L., Berniell, L., de la Mata, D., Maris, L., and Ortega, D., editors (2016). [RED 2015. A More Effective State. Capacities for designing, implementing and evaluating public policies,](#) chapter 2. CAF.
- Cainelli, G., Fracasso, A., and Vittucci Marzetti, G. (2015). [Spatial agglomeration and productivity in Italy: A panel smooth transition regression approach.](#) *Papers in Regional Science*, 94(S1).
- CEDLAS and The World Bank (2014). [A Guide to SEDLAC - Socio-Economic Database for Latin America and the Caribbean.](#) Technical report, The World Bank and CEDLAS.
- Chauvin, J. P., Glaeser, E., Ma, Y., and Tobio, K. (2017). [What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States.](#) *Journal of Urban Economics*, 98:17–49.
- Christofides, L. N. and Michael, M. (2013). [Exploring the public-private sector wage gap in European countries.](#) *IZA Journal of European Labor Studies*, 2(1):15.
- Cingano, F. and Schivardi, F. (2004). [Identifying the sources of local productivity growth.](#) *Journal of the European Economic association*, 2(4):720–742.

- Combes, P.-P. (2011). [The empirics of economic geography: how to draw policy implications?](#) *Review of World Economics*, 147(3):567–592.
- Combes, P.-P., Duranton, G., and Gobillon, L. (2008). [Spatial wage disparities: Sorting matters!](#) *Journal of Urban Economics*, 63(2):723–742.
- Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. (2012). [The productivity advantages of large cities: Distinguishing agglomeration from firm selection.](#) *Econometrica*, 80(6):2543–2594.
- Combes, P.-P., Duranton, G., Gobillon, L., and Roux, S. (2010). [Estimating agglomeration economies with history, geology, and worker effects.](#) In *Agglomeration economics*, pages 15–66. University of Chicago Press.
- Combes, P.-P. and Gobillon, L. (2015). [The Empirics of Agglomeration Economies](#), handbook of regional and urban economics edition.
- Davis, D. R. and Dingel, J. I. (2017). [The comparative advantage of cities.](#) Technical report, National Bureau of Economic Research.
- Davis, D. R., Dingel, J. I., and Miscio, A. (2018). [Cities, Skills, and Sectors in Developing Economies.](#)
- D’Costa, S. and Overman, H. G. (2014). [The urban wage growth premium: Sorting or learning?](#) *Regional Science and Urban Economics*, 48:168–179.
- De La Roca, J. and Puga, D. (2017). [Learning by working in big cities.](#) *The Review of Economic Studies*, 84(1):106–142.
- Desmet, K. and Henderson, J. V. (2015). [The geography of development within countries.](#) *Handbook of Regional and Urban Economics*, 5:1457–1517.
- Desmet, K. and Rossi-Hansberg, E. (2013). [Urban accounting and welfare.](#) *The American Economic Review*, 103(6):2296–2327.
- Desmet, K. and Rossi-Hansberg, E. (2014). [Analyzing urban systems: have megacities become too large?](#) Technical report, The World Bank.
- Duranton, G. (2015). [Growing through cities in developing countries.](#) *The World Bank Research Observer*, 30(1):39–73.
- Duranton, G. (2016). [Agglomeration effects in Colombia.](#) *Journal of Regional Science*, 56(2):210–238.
- Duranton, G. and Puga, D. (2004). [Micro-foundations of urban agglomeration economies.](#) *Handbook of regional and urban economics*, 4(4):2063–2117.

- Fally, T., Paillacar, R., and Terra, C. (2010). [Economic geography and wages in Brazil: Evidence from micro-data](#). *Journal of Development Economics*, 91(1):155–168.
- Ferreira, M. (2018). Human capital in cities. In Ferreira, M. and Roberts, M., editors, [Raising the bar for productive cities in Latin America and the Caribbean](#). Washington, DC: World Bank.
- Ferreira, M. and Roberts, M., editors (2018). [Raising the bar for productive cities in Latin America and the Caribbean](#). Washington, DC: World Bank.
- Glaeser, E. L. (1999). [Learning in cities](#). *Journal of urban Economics*, 46(2):254–277.
- Glaeser, E. L. and Resseger, M. G. (2010). [The complementarity between cities and skills](#). *Journal of Regional Science*, 50(1):221–244.
- Handbury, J. and Weinstein, D. E. (2014). [Goods prices and availability in cities](#). *The Review of Economic Studies*, 82(1):258–296.
- Hanlon, W. W. and Tian, Y. (2015). [Killer cities: Past and present](#). *The American Economic Review*, 105(5):570–575.
- Henderson, J. V. (2003). [Marshall’s scale economies](#). *Journal of urban economics*, 53(1):1–28.
- Henderson, J. V. and Wang, H. G. (2007). [Urbanization and city growth: The role of institutions](#). *Regional Science and Urban Economics*, 37(3):283–313.
- Hering, L. and Poncet, S. (2010). [Market access and individual wages: Evidence from China](#). *The Review of Economics and Statistics*, 92(1):145–159.
- Krugman, P. (1980). [Scale economies, product differentiation, and the pattern of trade](#). *The American Economic Review*, 70(5):950–959.
- Krugman, P. (1991). [Increasing returns and economic geography](#). *Journal of political economy*, 99(3):483–499.
- Krugman, P. and Venables, A. J. (1995). [Globalization and the Inequality of Nations](#). *The quarterly journal of economics*, 110(4):857–880.
- Lucifora, C. and Meurs, D. (2006). [The public sector pay gap in France, Great Britain and Italy](#). *Review of Income and wealth*, 52(1):43–59.
- Moretti, E. (2004a). [Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data](#). *Journal of econometrics*, 121(1-2):175–212.
- Moretti, E. (2004b). [Human capital externalities in cities](#). *Handbook of regional and urban economics*, 4:2243–2291.

- Moretti, E. (2004c). [Workers' education, spillovers, and productivity: evidence from plant-level production functions](#). *American Economic Review*, 94(3):656–690.
- Nunn, N. and Puga, D. (2012). [Ruggedness: The blessing of bad geography in Africa](#). *Review of Economics and Statistics*, 94(1):20–36.
- OECD (2017). Population with tertiary education. <https://data.oecd.org/eduatt/population-with-tertiary-education.htm>.
- Overman, H. G. and Venables, A. J. (2005). [Cities in the developing world](#). CEP Discussion Papers with number dp0695.
- Panizza, U. (2001). [Public sector wages and bureaucratic quality: Evidence from Latin America](#). *Economía*, 2(1):97–151.
- Rauch, J. E. (1993). [Productivity gains from geographic concentration of human capital: evidence from the cities](#). *Journal of urban economics*, 34(3):380–400.
- Roberts, M. (2018). The empirical determinants of city productivity. In Ferreyra, M. and Roberts, M., editors, [Raising the bar for productive cities in Latin America and the Caribbean](#). Washington, DC: World Bank.
- Roberts, M., Deichmann, U., Fingleton, B., and Shi, T. (2012). [Evaluating China's road to prosperity: A new economic geography approach](#). *Regional Science and Urban Economics*, 42(4):580–594.

# A Results by country

Table A1: 1st stage results <sup>a</sup>

	Argentina				Bolivia			
	broad		narrow		broad		narrow	
marriage	0.118***	0.0793***	0.112***	0.0855***	0.148***	0.105***	0.230***	0.219***
education	0.0855***	0.0834***	0.0763***	0.0755***	0.0905***	0.0886***	0.0814***	0.0811***
male	0.111***	0.119***			0.154***	0.163***		
Age	0.0108***	0.0442***	0.0107***	0.0471***	0.00541***	0.0354***	-0.000706	0.0149
Age <sup>2</sup>		-0.000414***		-0.000485***		-0.000378***		-0.000209
Observations	245,948	245,948	97,011	97,011	14,874	14,874	6,516	6,516
R-squared	0.760	0.762	0.769	0.770	0.461	0.463	0.467	0.468
	<b>Honduras</b>				<b>Mexico</b>			
marriage	0.128***	0.0712***	0.106***	0.0886***	0.158***	0.0925***	0.202***	0.176***
education	0.106***	0.104***	0.0907***	0.0905***	0.0948***	0.0912***	0.0848***	0.0839***
male	-0.0140	0.00811			0.151***	0.172***		
Age	0.0105***	0.0502***	0.00812***	0.0354***	0.0119***	0.0592***	0.0109***	0.0483***
Age <sup>2</sup>		-0.000515***		-0.000372***		-0.000613***		-0.000506***
Observations	38,269	38,269	17,685	17,685	94,105	94,105	41,579	41,579
R-squared	0.351	0.355	0.332	0.333	0.666	0.671	0.701	0.702
	<b>Chile</b>				<b>Colombia</b>			
marriage	0.138***	0.124***	0.161***	0.146***	0.105***	0.0737***	0.145***	0.126***
education	0.103***	0.102***	0.101***	0.101***	0.103***	0.101***	0.0827***	0.0823***
male	0.178***	0.182***			0.219***	0.228***		
Age	0.00973***	0.0255***	0.0108***	0.0386***	0.0122***	0.0463***	0.0102***	0.0492***
Age <sup>2</sup>		-0.000194***		-0.000369***		-0.000437***		-0.000525***
Observations	405,058	405,058	178,321	178,321	231,349	231,349	104,122	104,122
R-squared	0.726	0.726	0.744	0.745	0.552	0.556	0.586	0.588
	<b>Nicaragua</b>				<b>Panama</b>			
marriage	0.180***	0.146***	0.176***	0.168***	0.122***	0.101***	0.143***	0.131***
education	0.0651***	0.0635***	0.0704***	0.0702***	0.104***	0.102***	0.0851***	0.0849***
male	0.0376***	0.0519***			0.215***	0.221***		
Age	0.00861***	0.0353***	0.0123***	0.0247***	0.0106***	0.0322***	0.00792***	0.0317***
Age <sup>2</sup>		-0.000349***		-0.000169		-0.000273***		-0.000321***
Observations	24,730	24,730	7,486	7,486	205,122	205,122	95,382	95,382
R-squared	0.295	0.298	0.304	0.304	0.728	0.729	0.715	0.715

<sup>a</sup> \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

Table A2: 1st stage results (continued) <sup>a</sup>

	Costa Rica				Dominican Republic			
	broad		narrow		broad		narrow	
marriage	0.153***	0.125***	0.164***	0.145***	0.145***	0.106***	0.158***	0.138***
education	0.0918***	0.0902***	0.0766***	0.0761***	0.0673***	0.0657***	0.0530***	0.0528***
male	0.141***	0.150***			0.247***	0.260***		
Age	0.00734***	0.0307***	0.00563***	0.0327***	0.0125***	0.0486***	0.0126***	0.0509***
Age <sup>2</sup>		-0.000303***		-0.000366***		-0.000462***		-0.000519***
Observations	129,202	129,202	64,076	64,076	131,608	131,608	65,046	65,046
R-squared	0.772	0.773	0.769	0.770	0.527	0.531	0.557	0.559
	<b>Peru</b>				<b>El Salvador</b>			
marriage	0.108***	0.0658***	0.144***	0.125***	0.0777***	0.0424***	0.122***	0.0961***
education	0.0720***	0.0692***	0.0613***	0.0606***	0.0735***	0.0718***	0.0648***	0.0645***
male	0.250***	0.261***			-0.0467***	-0.0292***		
Age	0.00736***	0.0360***	0.00833***	0.0347***	0.00790***	0.0441***	0.00392***	0.0551***
Age <sup>2</sup>		-0.000363***		-0.000354***		-0.000459***		-0.000695***
Observations	459,915	459,915	195,175	195,175	27,117	27,117	11,978	11,978
R-squared	0.443	0.446	0.497	0.498	0.490	0.494	0.448	0.452
	<b>Ecuador</b>				<b>Guatemala</b>			
marriage	0.102***	0.0814***	0.118***	0.106***	0.116***	0.0400***	0.0777***	0.0558***
education	0.0719***	0.0707***	0.0556***	0.0553***	0.102***	0.0991***	0.0934***	0.0933***
male	0.197***	0.204***			0.0388***	0.0642***		
Age	0.00819***	0.0277***	0.00787***	0.0294***	0.00842***	0.0530***	0.00764***	0.0351***
Age <sup>2</sup>		-0.000244***		-0.000290***		-0.000581***		-0.000373***
Observations	237,801	237,801	109,634	109,634	58,030	58,030	28,424	28,424
R-squared	0.634	0.635	0.656	0.657	0.346	0.354	0.337	0.338
	<b>Uruguay</b>				<b>Brazil</b>			
marriage	0.181***	0.144***	0.220***	0.193***	0.231***	0.151***		
education	0.112***	0.109***	0.118***	0.117***	0.118***	0.114***		
male	0.189***	0.198***			-0.403***	-0.419***		
Age	0.0178***	0.0531***	0.0207***	0.0587***	0.0225***	0.0860***		
Age <sup>2</sup>		-0.000434***		-0.000501***		-0.000868***		
Observations	15,915	15,915	6,025	6,025	1,809,596	1,809,596		
R-squared	0.682	0.685	0.674	0.675	0.829	0.835		

<sup>a</sup> \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Table A3: 2nd stage results <sup>a</sup>

	ARG	BOL	BRA	CHL	COL	CRI	DOM	
Population density (ln)	0.123	0.057	0.020***	-0.016**	-0.006	-0.001	0.024**	
Avg. yrs of schooling (ln)	0.512	1.308***	0.442***	0.928***	0.352***	0.401***	0.119**	
Market access (ln)	0.002	0.006	0.018***	0.007	0.008	0.019***	0.015*	
Observations	12	62	1,292	281	192	370	180	
R-squared	0.583	0.578	0.664	0.526	0.168	0.467	0.315	
	ECU	GTM	HON	MEX	NIC	PER	SLV	URY
Population density (ln)	0.019**	-0.006	0.061	0.003	-0.045**	0.021***	0.001	0.109**
Avg. yrs of schooling (ln)	0.361***	0.568***	0.824***	0.695***	0.144***	0.497***	0.277***	0.542
Market access (ln)	0.020***	-0.004	-0.030	0.008	0.041**	0.011*	-0.022	-0.017
Observations	543	22	254	400	112	1,096	213	13
R-squared	0.336	0.754	0.335	0.542	0.425	0.383	0.130	0.922

<sup>a</sup> \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . In all columns, the dependent variable is the estimated location premium (measured in natural logs) from a series of country-level first-stage regressions after controlling for observable worker characteristics within the broad sample (all wage/salary employees aged from 14 to 65 years old) and survey-year fixed effects. Worker characteristics includes age, age-squared, marital status, gender, and years of schooling. WAP stands for the working-age population. The source of the population data is the Gridded Population of the World (GPW), v4. All regressions include controls for mean air temperature, terrain ruggedness and total precipitation. ARG = Argentina, BOL = Bolivia, BRA = Brazil, CHL = Chile, COL = Colombia, CRI = Costa Rica, DOM = Dominican Republic, ECU = Ecuador, GTM = Guatemala, HND = Honduras, MEX = Mexico, NIC = Nicaragua, PAN = Panama, PER = Peru, SLV = El Salvador, URY = Uruguay.

## B Other robustness tests for the 2nd stage results

Table B1: 2nd stage regression - location premiums from narrow sample<sup>a</sup>

Dependent variable: Location premium	1	2	3	4
Population density (ln)	0.057*** [6.108]	0.018* [2.025]	0.005 [0.532]	0.013 [1.362]
Average years of schooling (ln)		0.628*** [9.259]	0.620*** [8.496]	0.622*** [8.846]
Market access (ln)			0.020*** [5.314]	0.019*** [5.769]
Road density (ln)				-0.036*** [-3.784]
Mean air temperature (ln)	0.042 [0.781]	0.058 [1.158]	0.068 [1.503]	0.067 [1.660]
Terrain ruggedness (ln)	-0.036** [-2.624]	-0.028*** [-3.562]	-0.022* [-2.029]	-0.021* [-1.912]
Total precipitation (ln)	-0.037 [-0.763]	-0.015 [-0.537]	-0.018 [-0.764]	-0.022 [-0.928]
Constant	-0.917*** [-5.518]	-2.425*** [-11.656]	-2.822*** [-13.936]	-2.904*** [-14.286]
n	5,748	5,748	5,049	4,858
R-squared	0.564	0.651	0.681	0.687
Adjusted R-squared	0.562	0.65	0.679	0.685

<sup>a</sup> \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Robust t-statistics clustered at the country level in brackets. Country effects have been controlled in all columns. In all columns, the dependent variable is the estimated location premium after controlling for the effect of sorting on worker characteristics of the narrow sample. Narrow sample refers to all employed wage/salary workers aged 20-55 - working in the private sector. Worker characteristics include age, age-squared, marital status, gender, and the years of schooling. The source of the population data is the Gridded Population of the World (GPW), v4.