Inter-American Development Bank

IDB WORKING PAPER SERIES Nº IDB-WP-01324

Agglomeration and Congestion in Latin America

Andrés Gómez-Lobo
Santiago Sánchez González
Vileydy González Mejia
Agustina Calatayud

Inter-American Development Bank
Transport Division

February 2022

Agglomeration and Congestion in Latin America

Andrés Gómez-Lobo
Santiago Sánchez González
Vileydy González Mejia
Agustina Calatayud

Inter-American Development Bank
Transport Division

February 2022

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Agglomeration and congestion in Latin America / Andrés Gómez-Lobo, Santiago Sánchez González, Vileydy González Mejia, Agustina Calatayud.
p. cm. — (IDB Working Paper Series; 1324)
Includes bibliographic references.
1. Traffic congestion-Latin America-Econometric models. 2. Urban transportation-Latin America-Econometrics models. 3. Traffic congestion-Economic aspects-Latin America. I. Gómez-Lobo, Andrés. II. Sánchez González, Santiago. III. González, Vileydy. IV. Calatayud, Agustina. V. Inter-American Development Bank. Transport Division. VI. Series.
IDB-WP-1324
JEL Codes: R41, R48
Keywords: automobile, agglomeration, congestion, economic productivity

http://www.iadb.org

Agglomeration and Congestion in Latin America[1]

Andrés Gómez-Lobo
Department of Economics, University of Chile

Santiago Sánchez González
Inter-American Development Bank

Vileydy González Mejia
Inter-American Development Bank

Agustina Calatayud
Inter-American Development Bank

Abstract

In this paper we explore the possible impact of urban congestion on agglomeration economies for a cross-section set of cities in Latin America. We use travel time data from Tom Tom to estimate wage regressions equations controlling for city size and congestion. We use population in each city in the 19^{th} and early 20^{th} century as instruments for current city size (measures by population). In our baseline estimates, we find an elasticity of wages to city size of 0.05, very similar to previous research in the region. When congestion is included in the estimation, we find that agglomeration economies are reduced. This holds even after using rain-days and average yearly rain as an instrument for congestion. Our results imply that congestion is a drag on economic productivity. This indirect cost of congestion is considerably larger economically than the direct cost measured as the loss of valuable time for citizens.

Keywords: Latin America, agglomeration economies, congestion, economic productivity

1. Introduction

Latin America is the most urbanized region in the developing world. In 2020 it was estimated that 81.2 percent of the region's inhabitants lived in cities, a share expected to increase to

[1] We would like to thank Juan Pablo Chauvin for very useful comments to an earlier draft. The usual disclaimers apply.

87.8 percent by 2050.[2] The region also boasts six megacities and five of the largest thirty cities of the world.[3]

There are many explanations for this urbanization trend. During the last century, the import substitution strategy adopted by countries in the region created an economic bias against agriculture and natural resource industries, fostering accelerated migration to cities where industrial jobs were available (Lattes, 1995; Lattes, et al. 2004). Rural violence in some countries has also spurred migration to cities (Morrison, 1993; Calderón-Mejía and Ibañez, 2016). But even when these factors are no longer present, large cities in the region have retained their magnetism to attract people from other areas of their respective countries and have continued to grow.

Undoubtedly, another key explanation for the urbanization of the region is agglomeration economies. These economies are defined as the positive externalities derived from the spatial concentration of economic activities, an idea that goes back at least to Marshall (1890). These benefits are ultimately derived from savings in transportation costs, be it for the movement of goods, people or ideas (Glaser, 2010). These economies imply that workers and firms are more productive in larger cities. They are also the main explanations for the existence of cities.

There is substantial evidence for agglomeration economies in the developed world with wages and TFP rising with city size, employment density or industries' spatial concentration.[4] However, evidence for developing countries is scarce. As reported in Combes and Gobillon (2015), Combes, et al. (2013) find economies of density in Chinese cities with a higher impact of density on wages compared to the received literature for developed countries. A similar finding is reported in Chauvin et al. (2017) for Chinese and Indian cities.

[2] UN Department of Economic and Social Affairs, Population Dynamics, World Urbanization Prospects: The 2018 Revision. https://population.un.org/wup/Download/.
[3] By Megacity we refer to cities with over 10 million people: São Paulo, México City, Río de Janeiro, Buenos Aires, Bogotá, and Lima. UN Department of Economic and Social Affairs, Population Dynamics, World Urbanization Prospects: The 2018 Revision. https://population.un.org/wup/Download/.
[4] For a recent review see Duranton and Puga (2020) and Combes and Gobillon (2015), and Rosenthal and Strange (2004) for an earlier review.

Therefore, economies of density may be even more important for cities in the developing world than in developed countries.

The measurement of agglomeration economies on urban wages in Latin America has been less studied. Duranton (2016) finds that city population increases wages in Colombia with an elasticity of 0.054, larger than the parameter estimates for developed countries (albeit smaller than those found for India and China cited above).[5] Interestingly, he finds evidence that agglomeration economies are stronger in the informal sector than the formal sector. A similar result was obtained by Bernedo and Patrick (2021) for Peru, who find a population wage elasticity between 0.06 and 0.08 for informal firms in Peru, while the same results was between 0.02 and 0.06 for formal firms. However, using data for Ecuador, Matano, et al (2020) find an overall elasticity of 0.071 of wages with respect to urban density, but this effect is mostly for formal sector wages with a much smaller and statistically insignificant effect for informal sector wages.

Barufi, et al. (2016) find a 0.05-0.09 employment density wage effect for Brazilian cities, although Chauvin, et al. (2017) find a much lower effect of 0.026 for Brazilian microregions. Silva and Azzoni (2021) also use Brazilian data but use a double fixed effects model to control for worker and firm heterogeneity, and a definition of city size using night light satellite data. They still find a density wage effect of 0.043 to 0.057. While Ehrl and Monasterios (2021) find wage spillover effects from skill concentration in Brazil.

Finally, in a paper methodologically similar to ours, Quintero and Roberts (2018) try to disentangle the effects of agglomeration economies, human capital externalities and market access explanations for productivity differences across cities using data from 16 Latin American countries. Although in the pooled model of all countries the elasticity of wages to population density is 0.057 ―comparable to previous results― once human capital and market access variables are included, they find no evidence for agglomeration economies.

[5] Bacolod, et al (2021) find a wage city population elasticity of 0.052 for college graduates in Colombia. When controls are included for individual ability, parental education and income, this elasticity falls to 0.025. They also find that individuals sort into cities according to talent at a pre-work stage (decision of where to attend college).

This paper adds to the literature by measuring agglomeration economies across a set of cities in different countries of Latin America. We use harmonized household survey data from 129 cities from 5 countries of the region to estimate the impact of population on average wages. We also consider potential biases in the estimations due to sorting effects (cities attract more educated or skilled workers) and reverse causality (a city grows because wages are higher).

For the first problem we condition on workers´ educational attainment and labor experience in each city. As in Duranton (2016), Chauvin, et al (2017) and Quintero and Roberts (2018) we do not have panel data, so we are unable to control for unobserved heterogeneity (as in Combes, et al, 2010; or D'Costa and Overman, 2014) or dynamic issues such as human capital accumulation in larger cities. However, our results are in line with regional estimates reported above. Future research should address this issue.

For the second problem we follow Glaeser and Maré (2001) and Combes, et al (2010) and use an instrumental variable approach using the population levels of each city from census data in the 19th Century and beginning of the 20th Century; the idea being that economic activities hundreds of years ago are not related to the economic activities found today nor their productivity in a city, albeit population size is probably correlated for historical reasons.

We find agglomeration effects very similar to previous research with an elasticity of 0.05 for wages with respect to city population size. When only formal sector workers are used in the estimation sample, this elasticity is somewhat lower (0.04) indicating that agglomeration economies seem to be stronger in the informal sector than the formal sector, something already noted for the case of Colombia (Duranton, 2016) and Peru (Bernedo and Patrick, 2021).

This study is not limited to the estimation of agglomeration economies in Latin America. Our main interest is to explore the impact that traffic congestion may have for the productivity of economic activities in a city. This manner, this study aligns with the Inter-American Development Bank's Vision 2025 as benefits of a traffic congestion reduction are not limited to productivity gains, but it will certainly support the reduction of greenhouse emission derived from urban transportation and support a sustainable mobility system in the urban areas of LAC. To this end, we augment the estimated models using excess travel time information from TomTom for a subsample of our cities where this information is available. We find that conditional on city size, more congestion reduces the productivity enhancing effect of agglomeration economies.[6] This is so even when we instrument congestion levels using the average number of rainy days and the yearly average amount of rain in each city.

This last finding is relevant for policy discussions in the region. A recent study by Catalayud, et al. (2021) found that traffic congestion caused a loss of 650 million hours in Mexico City and 700 million hours in São Paulo. Even in a relatively smaller city, San Salvador, traffic congestion generated an excess loss of 37 million hours in 2019. Similar results are found for other cities of the region such as Bogota and Santiago, Chile.

Therefore, traffic congestion generates important losses in time for inhabitants in Latin American cities. These time losses imply annual economic costs that range from 0.5% of GDP in Mexico City to 1.1% of a city's GDP for Buenos Aires, with cities like Santiago, Rio de Janeiro and Bogota in between. What we find in this study is that there is an added economic cost due to congestion distinct from the loss of hours for travelers. In so far as traffic congestion acts like a drag on productivity enhancing agglomeration economies, there is an additional cost when mobility is impaired in a city.[7] Using our estimated results we simulate a reduction of 5% in the excess waiting time (as a proportion of free flow time), our measure of congestion, in each city. This reduction would increase productivity by an amount equivalent to 0.5% of GDP over the 13 cities where this simulation can be undertaken. In the aggregate these economic benefits are close to 10 billion dollars in 2019 for these 13 cities. These indirect gains from reducing congestion are much higher that the direct gains obtained from reduced travel time.

---

[6] Interestingly, Quintero and Roberts (2018) do not find much evidence of agglomeration effects in their analysis using data for 16 Latin American countries. They attribute this result to "sub-optimal infrastructure that increases congestion under high densities, which in turn overwhelm agglomeration economies" (page 4-5). Our results lend evidence to their conjecture.

[7] Other economic costs of congestion are related to potentially higher accident rates, air, and noise pollution, and even the loss of productivity from higher sick leave behavior (Gómez-Lobo and Micco, 2021).

One important consequence of our finding is that mobility enhancing transport investments or interventions in Latin American cities will provide economic and social benefits that go beyond those estimated by conventional cost-benefit analysis that consider only the resource savings and time savings as benefits (Venables, 2007). Some countries, notably the UK and Australia, have revised their project evaluation methodologies to include these "wider economic benefits" of transport projects when agglomeration economies are present.

This paper is structured as follows. In the next section we discuss the different types of agglomeration economies discussed in the literature. Since this is a well known topic, the discussion is rather brief and interested readers can consult the cited references. We then discuss the relation between agglomeration economies and congestion, providing a simple yet easy to understand model to clarify why agglomeration economies imply congestion costs that go beyond the usual excess time costs. In that section we also review the literature related to agglomeration economies and congestion. Following that we present the empirical approach to estimate agglomeration economies and the model used in the present research. A section describing our data follows. Results are then presented, and the paper concludes with a discussion of the results and their implications.

2. Agglomeration economies and congestion

Agglomeration economies are usually classified into two types, localization economies and urbanization economies. The first refers to the spatial clustering of firms within an industry, generating economies that are external to each firm but internal to the industry. The second refers to economies that arise due to higher urban density in general. These are economies that are external to firms and to an industry but internal to a city. However, research has shown that the impact of these last economies varies by industry (e.g., Graham, 2007a).

Localization economies go back at least to the ideas of Marshall (1890), who presented three complementary explanations of why firms of the same trade may localize together: access to specialized labor, informational spillovers, and better input-output linkages.[8] If employment density increases overall productivity, then policies that induce higher employment or density will generate a positive externality on existing workers.[9]

The literature presents evidence on the increase in productivity derived from agglomeration. These will be reflected in higher TFP, labor productivity and wages. Just as agglomeration generates benefits —such as increased productivity— it also has associated costs, such as inter- and intra-urban displacement.

Already Ciccone and Hall (1996) recognize that there will be two forces at work determining labor productivity: agglomeration economies and congestion. Their estimated parameter relating county density with observed labor productivity will be the net effect of these two forces. Congestion in their model could cause this relationship to be negative. That is, higher density implies lower productivity. However, Ciccone and Hall (1996) refer to congestion in the general sense of decreasing returns lo labor density, not specifically due to traffic congestion. Their econometric results imply that the agglomeration effect dominates the congestion effect, although productivity would be higher if congestion is lower.

Lucas and Rossi-Hansberg (2002) present a model where the spatial arrangement of productive activities in a city will depend on a trade-off between the productivity gains from co-localization of firms with the additional commuting cost for workers who are displaced to the periphery due to rising land rents in the central parts of a city. Therefore, commuting costs will have an impact on the localization of firms and thus on agglomeration externalities that firms enjoy. In principle then, higher congestion, by increasing commuting costs, should have an impact on productivity. In essence, congestion increases the economically relevant distance (as opposed to geographic distance) between firms and thus can have a negative

[8] Chapter 1 of Brueckner (2011) presents an intuitive and detailed discussion of the different agglomeration forces that may give rise to large cities. Duranton (2016) notes that the idea of agglomeration economies goes back even to Adam Smith.

[9] Implicit in this statement is the assumption that there are non-linearities in the relation between density and agglomeration economies. Otherwise, the benefits of higher employment in one area will be exactly compensated by the lower employment in the areas where workers migrated from.

impact on productivity through a lower agglomeration effect, irrespective of the underlying source of the agglomeration externality.

Commuting costs due to higher congestion could also affect productivity directly by reducing work hours or the quality of work due to tiredness. For example, Gómez-Lobo and Micco (2021) find that longer commuting times in Santiago, Chile, are related with more sick-leave behavior by workers. Sick-leave could be due to direct health problems created by long commutes (as in Künn-Nelen, 2016) or from shirking behavior (as in Ross and Zenou (2008) or van Ommeren and Gutierrez-Puigarnau (2011)). They estimate a productivity loss of 33 million dollars a year due to sick leave by workers. They also find that underground metro expansion decreases sick leave absence by workers benefiting from these investments.

Another result is found by Zarate (2021) in Mexico City. Metro expansion in that city is associated with better access to formal jobs by poorer individuals living in the poorer neighborhoods further away from the city center. Formal jobs pay higher wages and therefore higher congestion would in principle exacerbate access problems for lower income individuals to formal labor markets, reducing their wages.

For all these reasons we expect to find a negative relationship between congestion and wages.[10] However, in this paper we focus on the relationship between agglomeration economies and congestion, although some of our empirical results below could also be related to the other explanations discussed above.

Before presenting the literature analyzing the relation between agglomeration economies and congestion it is instructive to illustrate the possible impacts of congestion in a simple model due to Venables (2007). This allows to clearly see the additional costs that congestion may impose on the economy besides the loss of time by travelers.

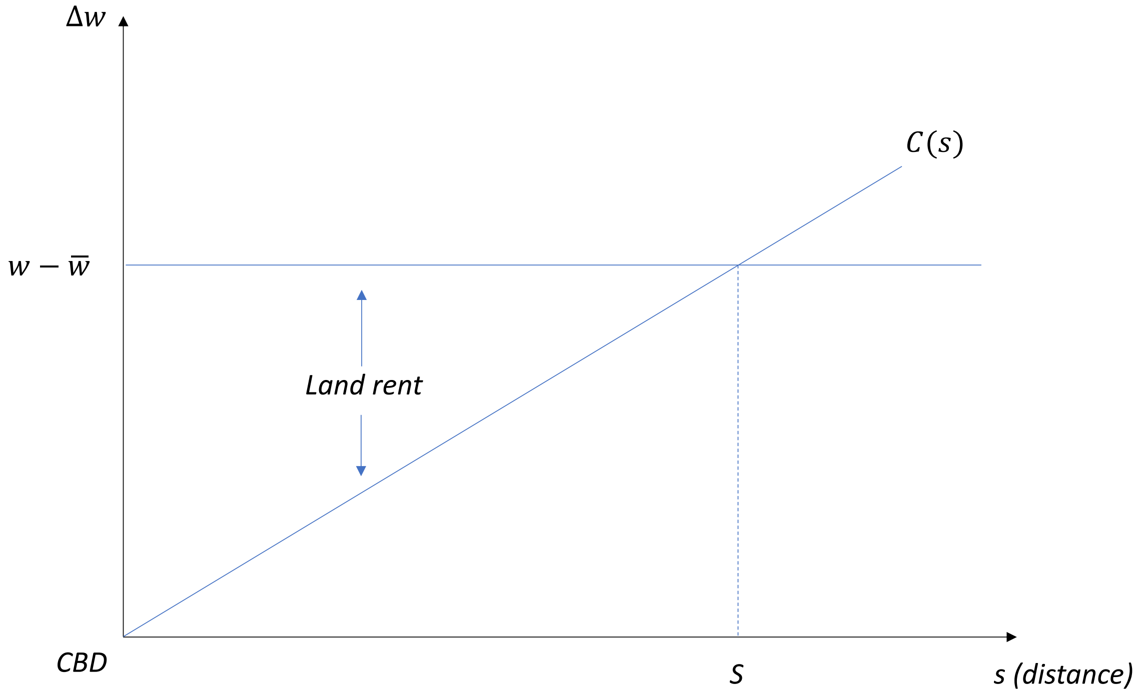
[10] There is a subtle point to be made here. The standard assumption in the urban economics literature of utility equalization across areas (otherwise workers would flock to the high real wage places) implies that if congestion is higher in a city and this produces disutility, nominal wages would have to increase to compensate and equate utility across space. However, prices, particularly land prices, may fall with congestion, as in the model presented next, and utility is equalized through this variable with a resulting negative relationship between nominal wages and congestion.

Following Venables (2007), we assume that agglomeration economies are related to the number of workers in a city, a particular measure of city size. As city size increases and so too the number of employments, the average wage rate is higher. We can express this relationship as *w(L(S))* where *w* is the average wage, which depends on the number of employments (L) which in turn is a function of city size *S*. We also assume that all employments are in the city center, or Central Business District (CBD).

For the sake of simplicity in presenting the intuition of the model, we further assume a very simple linear city where all employment is at the origin and commuting costs for workers increase as the distance between the city center and their residential location increases. Figure 1 shows a benchmark case where there are no agglomeration economies and wages are constant independent of city size (employment level). The wage differential between city wages and non-city (rural) wages ($\bar{w}$) is given by $\Delta w = w - \bar{w}$ and commuting costs as a function of distance is represented by the line *C(s)*, where *s* in the horizontal axis represents distance.

City size will extend to a distance where workers residing at that point are indifferent between working in the city and earning the wage premium $w - \bar{w}$ but incurring the commuting cost and working outside the city. In Figure 1 this point is at *S*. At that distance from the CBD commuting costs offset the wage differential and a worker residing at that point is indifferent between working in the city or outside the city. Individuals residing beyond the point *S* will not work in the city as commuting costs are higher than the wage premium. Also, note that in equilibrium workers residing at all distances from the CBD must obtain the same utility, otherwise there would be incentives to relocate to a more convenient residential location. Therefore, in equilibrium the difference between the wage premium and commuting costs are reflected in higher land rents. A worker residing closer to the CBD incurs lower commuting costs but pays a higher rent, while workers further out pay less rent but incur higher commuting costs. In this model all the benefits from agglomeration are capitalized into land prices and reaped by landowners.[11]

Figure 1: **City size without agglomeration economies and without congestion**



Source: Venable (2007).

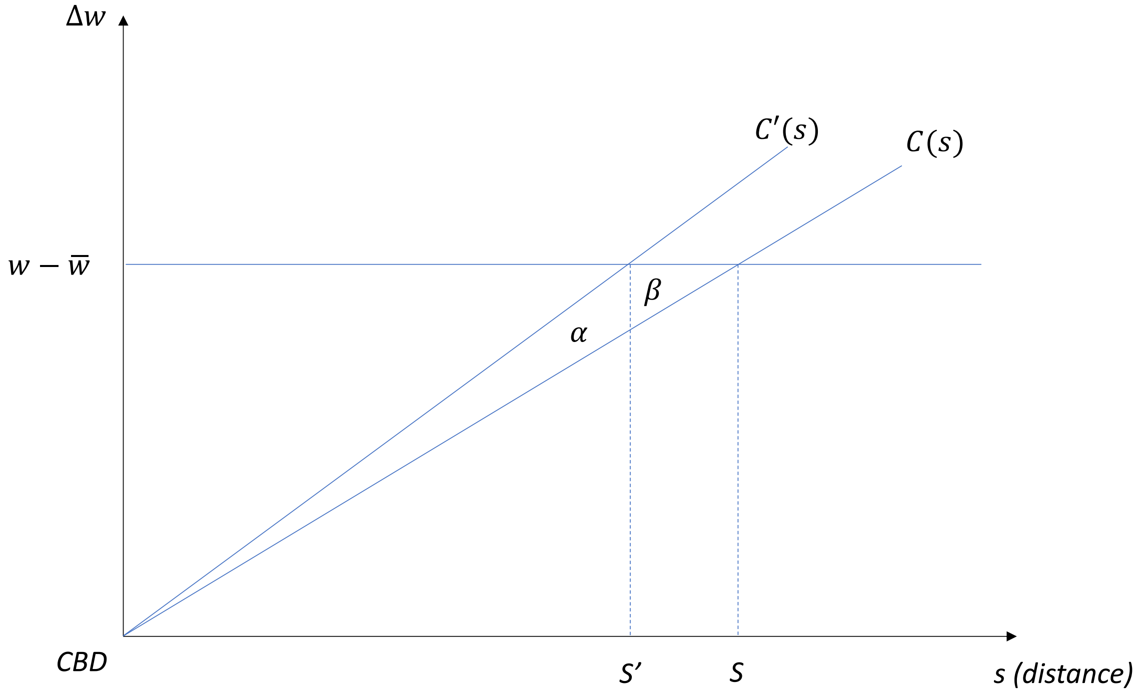
To see the impact of congestion, we can model this phenomenon as a rise in the commuting cost curve from *C(s)* to *C'(s)*. This could be due to rising private motorization rates, for example.[12] The effect is a reduction in city size from *S* to *S'*. The cost of congestion in this

[11] Glaeser (2010) reports that the correlation between density of cities in the United States in 1980 and housing price growth between 1980 and 2006 is 0.42. See also, Combes, et al. (2019) for the case of France. Duranton and Puga (2019) estimate that housing prices fall at the same rate from distance to the city center as travel cost increase with this distance for US metropolitan areas. More generally, Glaeser and Mare (2001) present evidence that the cost of living is higher in larger cities and that the real wage differential disappears when these local prices are considered. This is reasonable since an equilibrium condition is that utility of workers is the same across space. Otherwise, people would flock to the high real wage areas. Amenities may also differ among cities, and this could explain some real wage differential.

[12] Venables (2007) and Graham (2007b) suggest modelling the impact of congestion as a fall in the wage differential.

case is the additional commuting costs that workers must incur (area α) and the loss of the wage premium net of commuting costs that workers who no longer work in the city would have earned (area β). The first area is what is usually measured in studies that estimated the excess hours of travel caused by congestion, as in Catalayud, et al. (2021).[13] Area β is usually not estimated.[14]

Figure 2: **City size without agglomeration economies and with congestion**



Source: Venable (2007).

The above results change when agglomeration economies are introduced. These are modelled by a rising wage premium as city size, and thus the number of workers, increases, as in Figure 3. Now there are other effects of congestion. Assume an initial situation with commuting costs of *C(s)* and city size *S*. The initial wage premium is thus $w^0 - \bar{w}$. If rising congestion

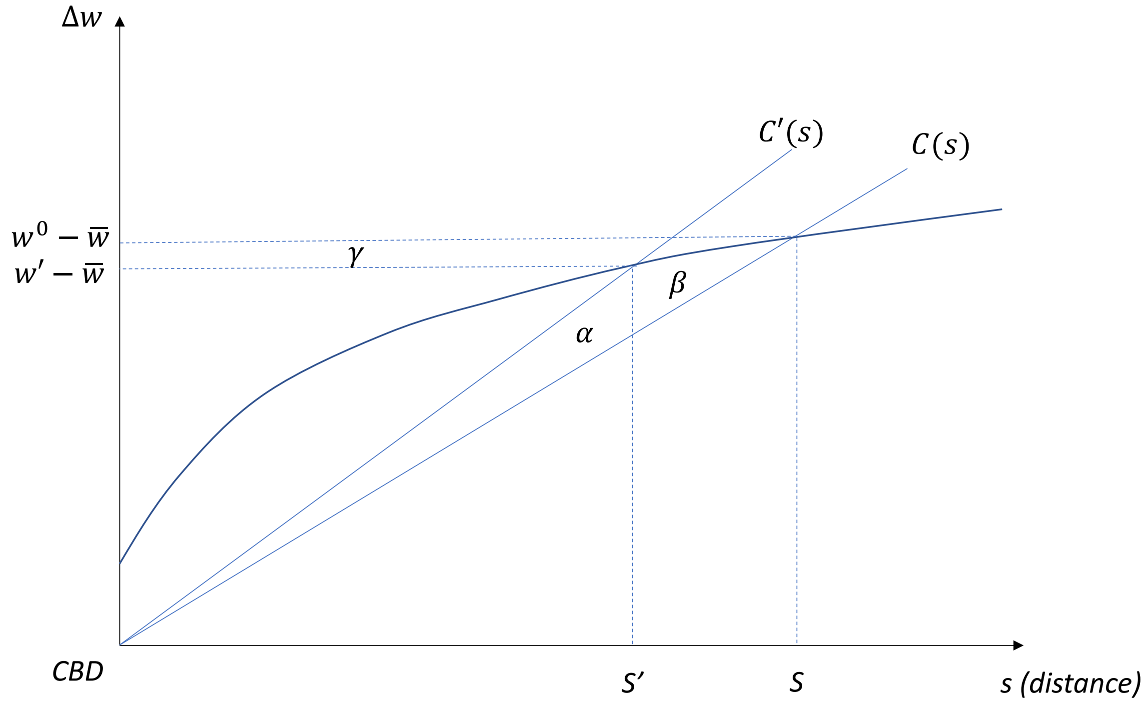
---

[13] It must be borne in mind that excess travel times will usually overestimate the social costs of congestion since the social optimal level of congestion is not zero. We thank Luis Rizzi for pointing this out to us. In this simple illustrative model, we do not make this distinction.

[14] However, the study by Zárate (2021) of the impact metro expansion in Mexico City can be assimilated to this effect, but for a decrease in commuting costs. In that study it is found that a new metro line that reduced travel times for low-income workers residing in neighborhoods in the outskirts of the city, increased access to formal jobs paying higher wages. Thus, lower commuting costs increased the effective size of the city in the sense that workers who previously worked in informal jobs near their residential area could now access higher paying formal jobs in the central part of the city. Therefore, the opposite effect would occur when commuting costs increase.

increases commuting costs to *C'(s)*, there is an additional cost of congestion given by area γ. Congestion causes city size to decrease from *S* to *S'*, reducing the benefits of agglomeration and reducing the wage premium to $w' - \overline{w}$. As mobility decreases in the city, the agglomeration economies related to lower transport costs and higher proximity between firms and workers are reduced, lowering productivity and wages. The important thing to note is that this is a distinct cost of congestion additional to the excess travel time cost.

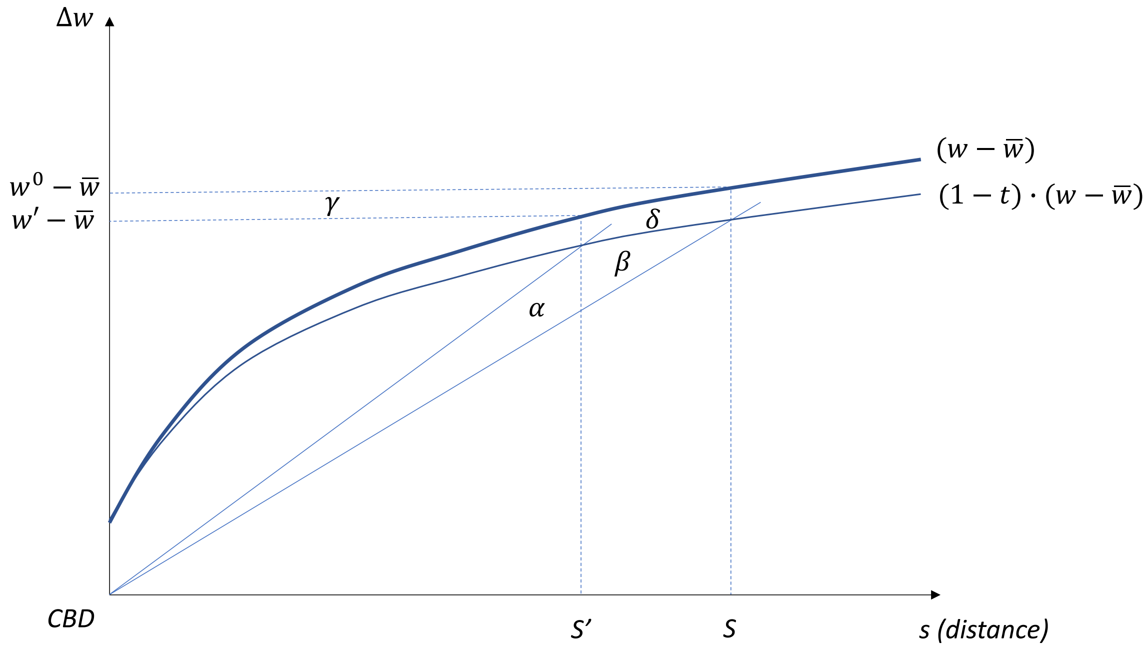
Figure 3: **City size with agglomeration economies and congestion**



Source: Venables (2007).

Venables (2007) also argues that when there is labor income taxation there is an additional benefit from agglomeration —and thus an additional cost to congestion— that is related to changes in tax revenues for the government. This is shown in Figure 4 where there is a marginal tax rate of *t* from workers' wages. Individuals' decision to work in the city or not will depend on the wage rate net of taxes, $(1 - t) \cdot (w - \overline{w})$, while the government earns as tax revenues the difference between the gross wage and the net wage. Congestion would now have four effects; the usual congestion costs due to commuting, given by area α and β, the impact on productivity due to the loss of agglomeration economies as city size is reduced from *S* to *S'*, plus a decrease in tax revenues for the government given by area δ. Thus, with income taxation there are two additional costs to congestion not considered by conventional measures of excess time costs.

Figure 4: **City size with agglomeration economies, income taxation and congestion**



Source: Venables (2007).

Graham (2007b) analysis the above effects using UK data and two measures of density, a distance measure of economic density and another measured based on the generalized cost of travel that considers travel times between zones. His results imply that congestion, by increasing the generalized cost of travel, effectively limits the productivity effects of agglomeration economies. Moreover, Graham (2007b) notes that estimates of agglomeration economies based on density measures that do not account for congestion will be downward biased. This is because there is a positive correlation between city size or density and congestion. Therefore, city size measures based solely on geographic distance or area will overstate the true size of a city, reducing the estimated productivity parameter. His empirical results support this conjecture. The estimated parameter for the productivity enhancing agglomeration effect is larger when a generalized cost of travel measure is used to define density compared to a purely geographic distance measure.

Brinkman (2016) uses a disaggregate model of a city based on Lucas and Rossi-Hansberg (2002) that reproduces stylized patterns of urban employment, residential location, and commercial and residential land prices for Columbus, Ohio, to study the effects of agglomeration and congestion. Agglomeration enhances production with declining effect as firms are located further apart. In this model, congestion cost at a given location is proportional to the number of commutes that pass through that location. Total commuting costs is then the sum of these costs for the locations that commuters must pass through to get to their jobs. Brinkman (2016) shows that higher congestion costs imply lower production, employment, and land rents. He also simulates the effect of an optimal congestion charge. Interestingly, with agglomeration economies a congestion charge will have two effects. On the one hand, it will reduce the congestion externality. But on the other hand, it will generate a more dispersed employment locational pattern reducing the productivity agglomeration effect. For the simulations and parameter values undertaken by Brinkman (2016) these two effects are of similar magnitude and in some cases the negative effect on productivity dominates the positive effect on congestion.[15]

Graham (2007b) also discusses that not all transport interventions will increase density and thus have a positive external agglomeration effect. Congestion charging is one such case. Congestion charging will reduce the time cost component of travel but will increase the financial component. The net effect will depend on the sensitivity of travel demand to cost for different trip purposes and on heterogeneity of value of time among different workers. The impact on agglomeration economies will depend on which type of workers drive the externalities and the net effect of a congestion charge on their overall cost of travel and, thus, on the effect on employment density.

[15] Sullivan (1983) also analyzes the impact of optimal congestion charging in a general equilibrium model of an urban setting, where land, housing and labor prices are endogenous as well as residential and firm location. Congestion charging leads to more concentrated residential location but less concentrated firm location. Since this model does not consider agglomeration economies (only external scale economies) the implication for our analysis is not straightforward. Safirova (2002) uses a general equilibrium model with agglomeration economies and congestion to analyze the impact of telecommuting. She finds that congestion charging decreases welfare for her simulation parameters. The reason is that the loss of agglomeration economies due to more telecommuting more than compensates for the lower congestion.

However, in both papers no mention is made of the differential effect of a congestion charge on different transport modes. Not only would a congestion charge induce more people to use public transport (lowering congestion) but it would improve travel times for all those already using public transport (at least in the case of buses) without having to pay for the charge. Therefore, the results of a congestion charge may be quite different from those found by Brinkman (2016) or discussed by Graham (2007b).

In a different vein, Sweet (2014) estimates the impact of road congestion on employment growth and labor productivity using a reduced form model with data from 88 US metropolitan statistical areas. He finds that congestion measured by Average Daily Traffic slows job growth and productivity growth per worker.

3. Model specification and estimation

Agglomeration effects have been measured using wages (Glaeser and Maré, 2001; Combes, et al., 2010, among others), average labor productivity (Ciccone and Hall, 1996) and TFP (Graham, 2007a, 2007b).[16] In this study we examine wage differences by city size (measured by population) among the countries of our database.

Duranton (2016) notes that in developing countries where labor market informality is high firm level data or formal sector wages may be inappropriate to estimate agglomeration economies given that it leaves out an important sector of the labor market. Therefore, we use household survey data for each country that includes both formal and informal labor earnings. We follow Combes and Gobillon (2015) and assume the following that firms have the following profit function:

$$\pi_{c,j} = p_{c,j} \cdot Q_{c,j} - w_{c,j} \cdot L_{c,j} - r_{c,j} \cdot K_{c,j} \qquad (1)$$

[16] Combes, et al. (2010) using French data show that estimated results are quite similar when using wages or TFP: Some authors have also examined potential impacts on employment and firm locational choice. On these studies see Combes and Gobillon (2015).

where c indexes a city in country j. For now, we assume all monetary units are measured in local currency. Below we will discuss this issue.

We also assume a Cobb-Douglas production function for each firm in each country given by:

$$Q_{c,j} = \frac{A_{c,j}}{\alpha^{\alpha} \cdot (1-\alpha)^{1-\alpha}} \cdot \left(s_{c,j} \cdot L_{c,j}\right)^{\alpha} \cdot K_{c,j}^{1-\alpha} \qquad (2)$$

where $Q_{c,j}$ is output in city c of country j, $A_{c,j}$ is total factor productivity, $s_{c,j}$ is labor skill, $L_{c,j}$ is the quantity of labor, $K_{c,j}$ is the quantity of capital and other inputs such as land and intermediate inputs, and α is a parameter with $0 < \alpha < 1$.

It is standard to show that if markets are competitive, profit maximization implies that the wage rate in each city is given by:

$$w_{c,j} = \left(\frac{p_{c,j} \cdot A_{c,j}}{r_{c,j}}\right)^{1/\alpha} \cdot s_{c,j} = B_{c,j} \cdot s_{c,j} \qquad (3)$$

where $w_{c,j}$ is the wage rate, $r_{c,j}$ is the cost of capital, and $B_{c,j} = \left(\frac{p_{c,j} \cdot A_{c,j}}{r_{c,j}}\right)^{1/\alpha}$. As discussed in Combes and Gobillon (2015), equation (3) contains two terms through which wages can be higher in city c of country j. Either labor skills $s_{c,j}$ are higher in city c, j or the term $B_{c,j}$ is higher.

Usually, agglomeration effects are modelled by assuming that the term $B_{c,j}$ is a function of city size or density, $B_{c,j}(N_{c,j})$, where $N_{c,j}$ is some measure of city size or density. Notice that the term $B_{c,j}$ includes several channels through which agglomeration economies can operate. For example, it could be that in large cities product prices are higher and capital costs are lower because transport costs are lower. Besides the pecuniary externalities, it could be that there are knowledge spillovers or the sharing of local public goods that make $A_{c,j}$ higher. The composite productivity term $B_{c,j}$ does not distinguish among these competing or complementary explanations. Also, if learning and the acquisition of skills is easier in larger cities, as in Glaeser and Maré (2001), then some of the effects operating through the $s_{c,j}$ term could also be attributed to agglomeration economies.

Assume now that nominal prices in country j are given by $x_{c,j} = PPP_j \cdot \tilde{x}_{c,j}$ where PPP_j is the Purchasing Power Parity exchange rate for country j, $\tilde{x}_j$ is the price x expressed in a common currency and $x \in (p, w, r)$.

Taking the average of equation (3) over all cities in country j, we obtain:

$$\overline{w}_j = \frac{1}{C_j}\sum_{c=1}^{C_j} w_{c,j} = PPP_j \cdot \overline{\widetilde{w}}_j = \frac{1}{C_j} \cdot \sum_{c=1}^{C_j} B_{c,j} \cdot s_{c,j} = \overline{Bs}_j$$

Dividing equation (3) by this last expression and taking logs we obtain:

$$ln\left(\frac{w_{c,j}}{\overline{w}_j}\right) = ln\left(\frac{\widetilde{w}_{c,j}}{\overline{\widetilde{w}}_j}\right) = ln\left(\frac{B_{c,j} \cdot s_{c,j}}{\overline{Bs}_j}\right)$$

From which we obtain the main equation we estimate:

$$\ln(w_{c,j}) - \ln(\overline{w}_j) = \mu_j + lnB_{c,j} + lns_{c,j} \qquad (4)$$

That is, the deviation of the log wage for a city from the log of the country's average wage is a function of country specific fixed effects, a term related to agglomeration effects and a term related to labor skills.[17]

Assuming a linear structure for the last two terms we arrive at:

[17] The log of the average wage by country can be subsumed into the country fixed effects without affecting the other parameters. Also note that the country fixed effects will also help to control for institutional differences, such as tax systems and other characteristics, between countries.

$$\ln(w_{c,j}) - \ln(\overline{w}_j) = \mu_j + \beta \cdot N_{c,j} + \gamma' X_{c,j} + \epsilon_{c,j} \qquad (5)$$

where $X_{c,j}$ is a vector of observable labor skills in city c of country j, β is a parameter, γ is a parameter vector conformable with X and $\varepsilon_{c,j}$ is an error term.

There are two econometrics issues that must be dealt with which before estimating equation (5) (Combes, et al., 2010; Combes and Gobillon, 2015). First, there could be reversed causality between wages and city size. For example, high wages in a locality may attract more workers, increasing density. Second, there could a sorting effect with more productive workers may end up in cities and this is reflected in higher wages.

The first problem has usually been resolved using instrumental variables. For example, Ciccone and Hall (1996) use the county population in the nineteenth century (from census information), the presence of a railroad at the end of that century and the distance to the seaboard as instruments for observed modern county density. Combes, et al. (2010) also use population from nineteenth century census as instruments in their study of the relationship between wages and TFP and density of the different labor markets in France. These last authors also use different measures of soil quality and geology as instruments for population size. Duranton (2016) also use census data from the nineteenth century census and geological characteristics of municipalities in their study for Colombia.

The idea behind these instruments is that population a century or more ago will be related to population today due to persistence in settlement patterns. However, the productive reasons that led to population concentration more than a century ago —early industrialization or agricultural production— are unrelated to current productive activities and thus productivity. Soil quality may also explain historical population concentration at a time when agriculture was one of the most important sectors (and employer) in the economy and thus correlated with current population, but not related to current productive activities where agriculture is no longer an important sector. The use of railroads and the closeness to the seaboard as instruments have a similar explanation.

In this paper we follow this same approach and use as instruments the population density in the 19th and early 20th century for the different cities of the database. The historical population figures were obtained from the earliest censuses in the five countries we study. We also have information of the population of several cities during colonial times. We use these as instruments in some estimations, although the number of observations is limited.

The second problem, sorting or labor quality problem can be partially dealt with by conditioning on observable labor skills for workers in each city. We use age and educational attainment of the workforce in each city. However, unobservable characteristics is still an issue. Since we do not have panel data, we cannot use worker fixed effects to control for unobservable characteristics. As in Duranton (2016), Chauvin, et al (2017) and Quintero and Roberts (2018) we must acknowledge this limitation of our data.[18] However, our results presented below are in line with previous studies and thus this may not be such a major concern. Future research with panel data could address this problem.[19]

When congestion is added as an additional explanatory variable in equation (5) there is an additional endogeneity issue to be dealt with. Since congestion will most probably be a function of city size, productivity shocks that change a city's population size will also increase congestion levels. Therefore, the error term of a wage equation may also be correlated with the congestion variable, generating a classical endogeneity problem. To tackle this problem, we use average rainfall and the number of rainy days in each city as additional instruments for congestion is some specifications.

4. Data

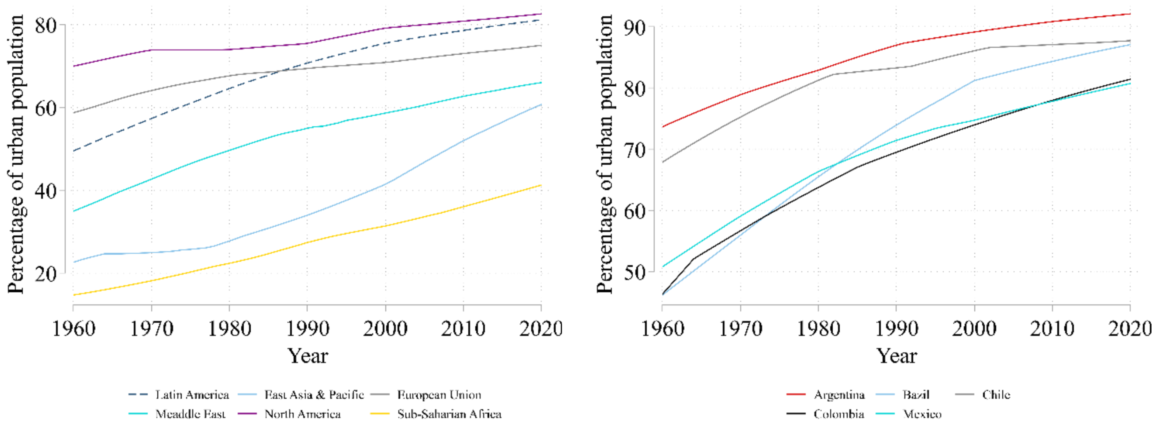
Latin America is currently the second region with the largest urbanization in the world. Figure 5.A presents the rate of urbanization during the last 60 years of Latin America

[18] We explored using the size rank of each city within a country as an additional control. The idea was that, conditional on absolute city size measured by population, the relative rank of each city may be relevant for workers sorting according to unobservable abilities. However, the results were not significant and the estimations including city rank variables are not reported below.
[19] Tom Tom congestion data is available for most cities in our sample from 2017 to 2020.

compared to other regions of the World. Latin America surpassed the European Union some decades ago and is now close to the urbanization rates of North America. Figure 5.B presents the urbanization rates for the five countries considered in the present study: Argentina, Brazil, Chile, Colombia, and Mexico.

Figure 5: Urbanization

A. Regions of the world B. Countries under analysis



Source: World Development Indicators, The World Bank.

For these five countries data was gathered from a set of different sources. First, information regarding wages, population, employment, education level, age, and economic sector, was retrieved from household-level data from each country's national institute of statistics:

- *Argentina*: Encuesta Permanente de Hogares (EPH, for its acronym in Spanish) – Second quarter 2019.
- *Brazil*: Pesqiusa Nacional por Amostra de Domicilios (PNAD, for its acronym in Portuguese) – Second quarter 2019.
- *Chile:* Encuesta Suplementaria de Ingreso (ESI, for its acronym in Spanish) – fourth quarter 2019.
- *Colombia*: Gran Encuesta Integrada de Hogares (GEIH, for its acronym in Spanish) – third quarter 2019.
- *México*: Encuesta Nacional de Ocupación y Empleo (ENOE, for its acronym in Spanish) – third quarter 2019.

This data was harmonized and combined. From this dataset, 129 urban areas from the 5 countries are observed (see: Table 1). The household surveys for Argentina, Brazil, Colombia, and Mexico are statistically representative at the major urban or metropolitan areas. On the other hand, Chile’s household survey is representative at county level while it allows to differentiate between urban and rural area. In all, the surveys have statistical representation for all the 129 cities or urban areas used in this study.

In general, the employment data retrieved from these surveys are divided into 9 economic sectors according to the International Standard Industrial Classification of All Economic Activities (ISIC). Data for Mexico is different since it classifies economic sectors according to the North American Industrial Classification System (NAICS). Those sectors that were not comparable to ISIC codification formed a set of economic sectors for this country. In total, the dataset contains over 650 thousand observations. Mexico has the largest number of cities included in the dataset with 36 metropolitan areas, while Argentina and Chile have data representing the largest proportion of the national population, registering over 60% of its total inhabitants. The second part of Table 1 presents the distribution of urban population in quartiles, namely, *Q1* is the total proportion of people that live in a city that belongs to one of the top 25% most populated city relative to the grand total. This can be understood as a proxy for decentralization. As seen, the countries with the most evenly distributed population are Mexico and Brazil. On the other hand, Chile tends to be the most centralized country.

Table 1: Household data description

Country	Number of urban areas	Population (thousands)			
		Minimum	Mean	Maximum	Total
Argentina	31	82	914	15,243	28,330
Brazil	27	298	3,125	21,712	84,374
Chile	12	157	970	6,999	11,637
Colombia	23	109	1,178	10,223	27,086
Mexico	36	220	1,312	17,143	47,234
	Q1	Q2	Q3	Q4	
Argentina	78.3	11.5	6.8	3.4	
Brazil	66.2	20.8	9.3	3.7	
Chile	77.3	10.6	7.4	4.8	
Colombia	76.0	12.8	7.7	3.6	
Mexico	67.7	14.6	11.6	6.2	

Source: Authors' elaboration with information from harmonized household survey of each country.
Note: Population figures are given in thousands

Furthermore, congestion data has been retrieved from TomTom data source[20]. This is an international index that estimate congestion in more than 400 cities for 57 countries around the world, among which are the 5 considered in this analysis. The index retrieved from this source, referred as congestion onwards, reports the average percentage increase in time that would take to do a trip compared to free-flow conditions. For example, if the index reports 50%, this means that, on average, it will take 90 minutes to do a trip that is done in 60 minutes under free-flow conditions. Although this is the most complete comparable congestion data available, the number of urban areas with information in our dataset is only 13[21]. Bogota is the city that registers the largest congestion according to the index, while Brasilia is the lowest congested metropolitan area (see Table 2).

[20] Data available at: https://www.tomtom.com/en_gb/traffic-index/.
[21] The congestion data for each country at the city level are: (i) Argentina has information available for Buenos Aires; (ii) Recife, Rio de Janeiro, Fortaleza, Salvador, Sao Paulo, Belo Horizonte, Porto Alegre, Curitiba, and Brasilia are the cities available for Brazil; (iii) Chile has information only for the capital city (Santiago de Chile). (iv) Colombia has information for Bogota; (v) Mexico has information for Mexico City.

Table 2: Congestion in 2019

City	Country	Congestion level 2019
Bogota	Colombia	68%
Mexico City	Mexico	52%
Recife	Brazil	50%
Rio de Janeiro	Brazil	46%
Sao Paulo	Brazil	45%
Santiago	Chile	44%
Salvador	Brazil	43%
Fortaleza	Brazil	37%
Belo Horizonte	Brazil	35%
Porto Alegre	Brazil	35%
Buenos Aires	Brazil	35%
Curitiba	Brazil	28%
Brasilia	Brazil	21%

Source: TomTom

Information regarding population in the 19th and 20th centuries for these cities has been retrieved from different sources. These include (i) academic publications such as Morse (1974), Borah (2021), Abundio (2017), Alcalá Ferráez (2015), Sánchez (2016), Toledo et al. (1992) and Vargas and Vargas (2016); (ii) historical statistical yearbooks; (iii) data from national statistics institutions, (iv) and other alternative sources.[22] Due to the heterogeneity of sources considered, the years do not coincide exactly. Therefore, to make the historical population comparable across metropolitan areas, the average of all recorded populations in 1850 – 1859 for each city was calculated and accordingly the same for 1900 – 1949. These last averages were used as instruments for current population in the estimations below. It is worth mentioning that when the same metropolitan area registers historic population in more than one source for the same year, the figure used was the one from the most reliable source according to the following order: (i) academic publications; (ii) historical statistical yearbooks; (iii) data from national statistics institutes; and (iv) other alternative sources.

Finally, to instrument the congestion variable, information regarding precipitation was obtained from Current Results weather and science facts, information which ultimately

[22] All of the specific sources are duly cited in Table A.1 of the Annex.

comes from the different national meteorological agencies. Average yearly precipitation and mean days of rain in the year for each city were used.

Following the discussion above, to estimate the effects of agglomeration on productivity, a variable measuring the wage rate of each individual relative to their country's average wage was constructed. This indicator is based on the hourly wage of each individual as follows:[23]

$$\ln(Indicator_i) = \ln\left(\frac{{W_i}/{h_i}}{\sum_i^{N_j}\gamma_i \cdot \left({W_i}/{h_i}\right) \Big/ \sum_i^{N_j}\gamma_i} * 100\right) \quad (6)$$

where subscripts *i* refers to individuals; N_j is the number of people employed in country *j* (that is, the number of observations in the dataset for each country); γ refers to the survey expansion factors; *h* stands for the number of hours effectively worked; and *W* are total labor earnings. Notice that this indicator variable is expressed in logarithm.

Descriptive statistics

Table 3 and Figure 1 show a brief description of the data by country. The average age and female participation in the working population is similar across countries, around 40 years and 43%. On the other hand, the size of the formal sector differs significantly from one country to another:[24] it is largest in Chile followed by Brazil, with 76% and 67%, respectively, while Mexico's formal sector is below 50%. Argentina is the country with the highest share of educated population. For instance, only 3% of its urban population has never assisted to an educational institution, while the regional average is 10% for this variable.

---

[23] For the econometric estimations below, it is not necessary to demean the individual wage rate since a country fixed effect is included in the model. However, for the descriptive statistics and graphs presented next it is useful to calculate this indicator.

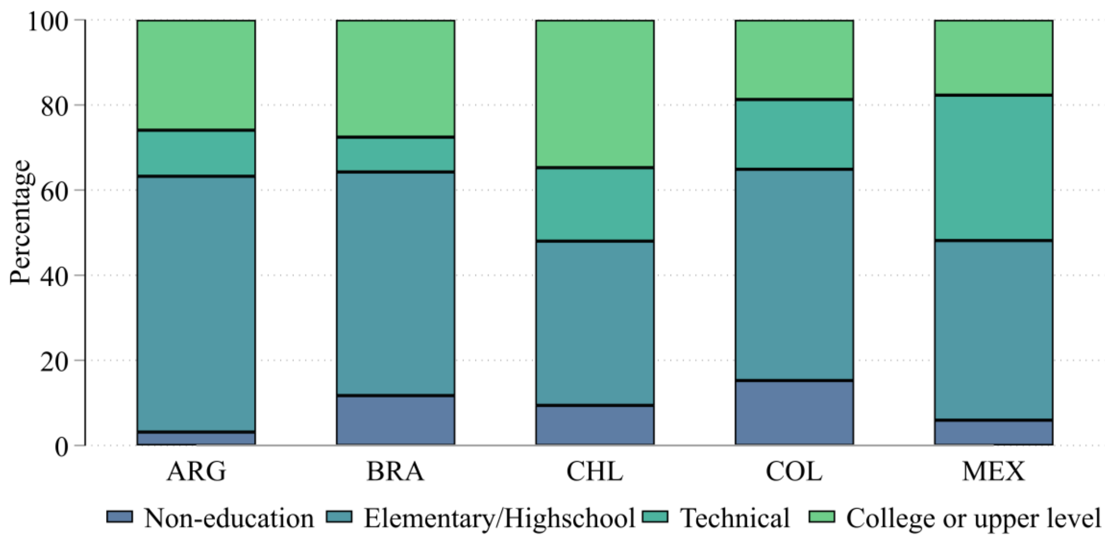
[24] Informal workers are those who either do not contribute to social security or who own an unincorporated enterprise where a separation of the financial activities of production and those of the owners is not possible.

Table 3: Descriptive statistic

Country	Women (%)	Average age	Formal (%)
Argentina	44.1	41.2	50.5
Brazil	45.7	40.2	67.2
Chile	43.1	42.4	75.6
Colombia	44.6	39.4	50.6
Mexico	41.6	39.4	47.0
Total	44.4	40.2	59.3

Source: Authors' elaboration with data from national household surveys
Note: Sample weights are used to calculate the average for each country. Column totals are also weighted averages.

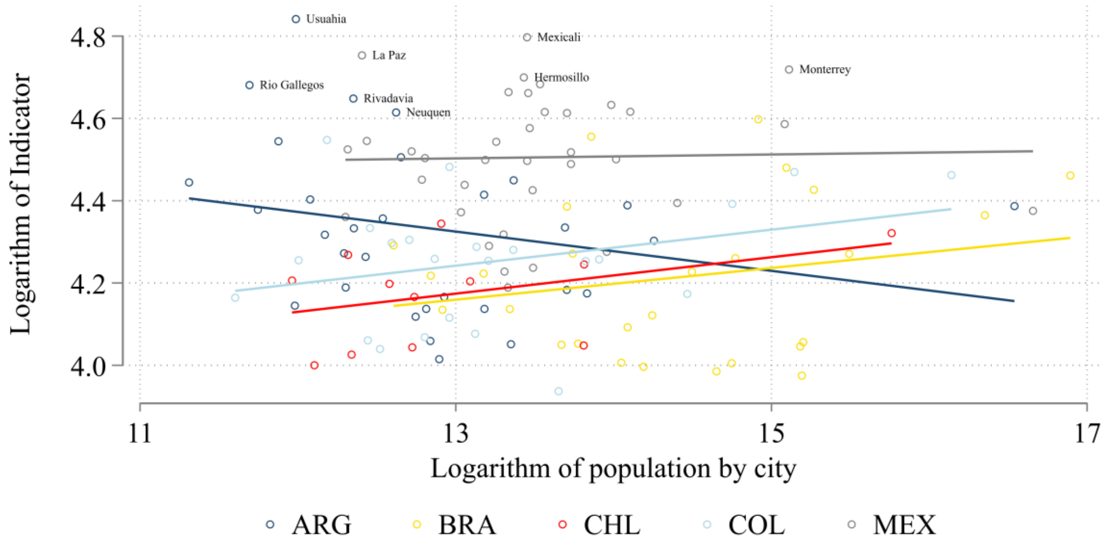
Figure 6: Descriptive statistics (education)



Source: Authors' elaboration with data from national household surveys

Figure A.1 in the Annex presents information on the relation between city population size and relative labor earnings by country and economic sector. Figure 3 presents the linear relation between the average relative wage indicator and population by metropolitan area for each country. Four out of the five countries present a positive relation between population size and average relative wage, and the slope is statistically significant for all countries that present a positive trend. On the other hand, only Argentina presents a negative slope, nonetheless, it is not significant. Most of the outliers, those cities that have an apparent dynamic that would be counterintuitive according to our hypothesis, are found in Argentina and Mexico, and have been labeled in Figure 7. These countries present specific metropolitan areas with high relative labor earnings in conurbations that are not particularly large. These cities are mainly found in the south for the case of Argentina and in the north for Mexico and these particularities can be explained by factors that our beyond the scope of this research.

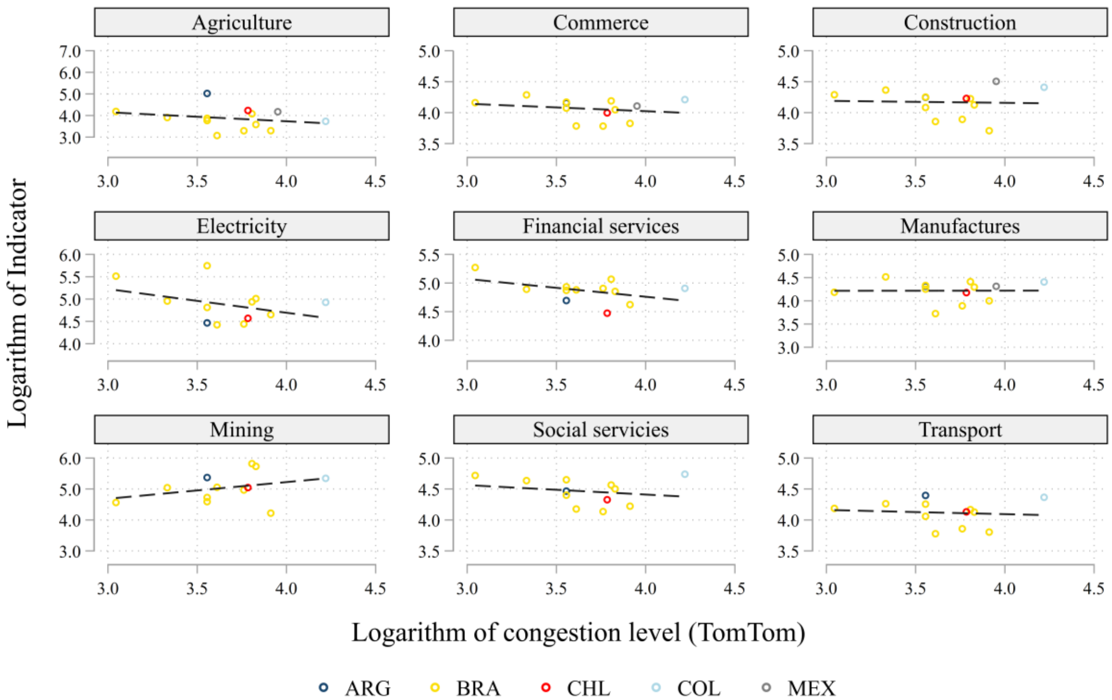
Figure 7: Indicator and population



Source: Authors' elaboration with data from national household surveys

One of the novelties of this article is the introduction of congestion to examine the effects of mobility on productivity. Figure 8 presents the relative labor productivity for each city in each economic sector with its corresponding congestion level. Except for Mining, the trend is always negative. More importantly, five out of the eight economic sectors present a significant coefficient.

Figure 8: Wages and congestion



Source: Authors' elaboration with data from national household surveys and TomTom

6. Results

Following equation (5) from Section 3 above, the most general model estimated is:

$$\ln\left(indicator_{ic(i)rp}\right) = \delta \ln\left(pop_{icp}\right) + \beta \ln\left(TomTom_{cp}\right) + X_{ic(i)rp}\theta + \varphi_r + \omega_p + \mu_{ic(i)rp} \qquad (7)$$

where the variable *indicator* and subscripts are as explain for equation (6). In addition, δ is the elasticity of productivity with respect to agglomeration; β is the corresponding elasticity regarding congestion; X is a matrix that contains individual level controls; while φ and ω refer to economic sector and country fixed effects; finally, μ stands for the clustered error term.

For this research, agglomeration is defined as the number of residents in the metropolitan area: in this sense, *Ln(pop)* refers to the natural logarithm of the population that reside in each city.

It is important to note that all the models consider economic sector and country fixed effects. Additionally, following Duranton (2016), we estimate the models for the full sample and for a subset of formal sector workers. All the models consider robust standard errors clustered by metropolitan area.

In some specifications we test for non-linear agglomeration effects by including the logarithm of population squared: $Ln(pop)^2$. *Ln(TomTom)* refers to the logarithm of the TomTom index, our measure of congestion.

As for individual level controls, we include gender where *Female* is a dummy variable that takes the value of 1 for women; *Ln(age)* refers to the natural logarithm of the individual's age, as a proxy for labor experience; *Ln(age)2,* accordingly, refers to the square of *Ln(age); Formal* takes the value of 0 if the worker belongs to the informal sector, which, for the case of Latin America, are those who either do not contribute to social security or those who own a unincorporated enterprise whose activities do not permit a separation between the financial results of production activities and that of its owner's; *Education* stands for educational level, which, to make it comparable across countries, classifies educational attainment into four categories (see Figure 2).

As already mentioned, all models include economic sector and country fixed effects to control for institutional factors (tax systems, labor laws, etc) of each country as well as any other variable that is country specific. Furthermore, all models consider robust errors clustered at the city level.

Table 4A presents the first set of regressions using the full sample and without including the congestion variable. The intention of these first set of models is to evidence the effect of agglomeration on productivity and assess the robustness of the results to alternative specifications. According to these estimates, being a woman significantly reduces wages compared to man with a gender wage gap of approximately 18%. *Ln(Age)* which for the purpose of this article is a proxy for labor experience, presents an increasing, but marginally decreasing effect of productivity, as expected. It is important to notice that belonging to the formal sector increases productivity by over 30%. As expected, higher educational attainment increases relative wages also. Since congestion is not included in these first set of models, data from all the urban areas can be used in the regressions, this is 129 metropolitan areas.

The coefficient of interest which refers to the elasticity of agglomeration on productivity, this is δ, ranges between 0.03 – 0.04 depending on the model specification. These figures are highly consistent with the results obtained by other researchers as discussed in the introduction. Model (*iv*) would indicate that there are no non-linear effects in the relation between city size and productivity.

Table 4.B presents results from the same models as above but restricting the sample to formal sector workers. The coefficients of the individual level controls are very similar to the above results using the full sample. However, the estimated agglomeration elasticity is smaller than with the full sample, by 16% to 25% depending on the specification. This is consistent with previous findings comparing developed to developing countries (Chauvin, Glaeser, and Tobio, 2013; Combes, D´emurger, and Shi, 2015). As in Duranton (2016) for the case of Colombia and Bernedo and Patrick (2021) for Peru, we find that agglomeration effects on productivity are stronger for informal sector workers than for formal sector workers.

Table 4.A: Linear regression

	(i)	(ii)	(iii)	(iv)
Ln(pop)	0.04**	0.04**	-0.30	0.03**
Ln(pop)^2			0.01	
Female		-0.19***	-0.19***	-0.18***
Ln(age)		5.18***	5.18***	1.56***
Ln(age)^2		-0.68***	-0.68***	-0.16***
Formal				0.31***
Non-education				--
Primary/Highschool				0.30***
Technical				0.57***
College or upper				1.05***
Constant	3.25***	-6.47***	-3.92	-0.52
Observations	261,381	261,381	261,381	261,381
R^2	0.08	0.11	0.12	0.32
Metro Area	129	129	129	129

* p<0.1; ** p<0.05; *** p<0.01. All regressions include country fixed effects and standard errors are clustered at the metropolitan area.

Table 4.B: Linear regression (Only formal workers)

	(v)	(vi)	(vii)	(viii)
Ln(pop)	0.03**	0.03**	-0.37	0.03**
Ln(pop)^2			0.01	
Female		-0.13***	-0.13***	-0.17***
Ln(age)		3.88***	3.88***	1.30***
Ln(age)^2		-0.48***	-0.48***	-0.11**
Non-education				--
Primary/Highschool				0.31***
Technical				0.62***
College or upper				1.13***
Constant	3.92***	-3.84***	-0.87	0.25
Observations	140,856	140,856	140,856	140,856
R^2	0.08	0.13	0.13	0.36
Metro Area	129	129	129	129

* p<0.1; ** p<0.05; *** p<0.01. All regressions include country fixed effects and standard errors are clustered at the metropolitan area.

As discussed above, there is a potential reverse causation problem in the models presented in Table 4A and 4B. Population changes are likely to be sensitive to changes in wages. In what follows we use an instrumental variable approach to address this problem.

The instruments that are used in the Two Stage Least Square (TSLS) estimation exposed in Table 5 are both the average population in the last half of the 19th century and the average population of the first half of the 20th century. As specified in the methodology section, all population data was retrieved from different sources, however, due to the lack of homogeneity of these sources, and to make these comparable, they were averaged over a 50-year period. When $ln(pop)^2$ is included in the specification the square of the historic population is also used as an instrument.

It is worth to mention that Tables 5.A and 5.B maintain the same logic as in Table 4, the sample includes all workers for the former and only formal sector workers for the latter. The last columns of the table present the R^2 of the first stage. Since all first stage models present an F-statistic greater than 10 and an R^2 over 0.80, it is likely that we have strong instrument (Cameron et. al, 2005; Stock & Yogo, 2002). Since we do not have historic population data for all cities, our sample decreases to 59 Metropolitan Areas or cities.

The results obtained in Table 5.A and 5.B are robust and consistent with what was obtained in the first set of models. The results for the individual level controls are very similar to results presented earlier. The elasticity of agglomeration obtained when instrumenting is a little larger. However, this effect is mainly driven by the cities included in the latter models since, when executing the models proposed in Table 4 only with those cities included in Table 5, the agglomeration effect tends to be similar with what is observed in the last set of models.

Table 5.A: Linear regression instrumenting population

	(i)	(ii)	(iii)	(iv)
Ln(pop)	0.05**	0.05*	-2.59***	0.04**
Ln(pop)2			0.09***	
Female		-0.19***	-0.19***	-0.17***
Ln(age)		5.07***	5.08***	1.40***
Ln(age)^2		-0.67***	-0.67***	-0.14***
Formal				0.31***
Non-education				--
Primary/Highschool				0.30***
Technical				0.57***
Collage or upper				1.07***
Constant	2.97***	-6.51***	13.43*	-0.44
Observations	149,324	149,324	149,324	149,324
R^2	0.07	0.11	0.10	0.32
Metro Area	59	59	59	59
First-stage's R^2				
Ln(pop)	0.82	0.82	0.85	0.82
Ln(pop)2			0.86	

* p<0.1; ** p<0.05; *** p<0.01

Notice that all the F-statistic of the first-stage regression are larger than 10. All regressions include country fixed effects and standard errors are clustered at the metropolitan area.

Table 5.B: Linear regression instrumenting population (Only formal workers)

	(v)	(vi)	(vii)	(viii)
Ln(pop)	0.04**	0.04*	-2.00***	0.04*
Ln(pop)^2			0.07***	
Female		-0.12***	-0.12***	-0.16***
Ln(age)		3.79***	3.76***	1.25***
ln(age)2		-0.47***	-0.46***	-0.11*
Non-education				--
Primary/Highschool				0.31***
Technical				0.62***
Collage or upper				1.15***
Constant	3.69***	-3.86***	11.62**	0.20
Observations	80,211	80,211	80,211	80,211
R^2	0.08	0.12	0.12	0.36
Metro Area	59	59	59	59
First-stage's R^2				
Ln(pop)	0.82	0.82	0.86	0.82
Ln(pop)2			0.87	

* p<0.1; ** p<0.05; *** p<0.01

Notice that all the F-statistic of the first-stage regression are larger than 10. All regressions include country fixed effects and standard errors are clustered at the metropolitan area.

The next set of models include congestion as an explanatory variable. Since only 13 cities in the sample have information regarding congestion, the models are now estimated on this limited sample.

The results are presented in Table 6.A and 6.B. Interestingly, the coefficient associated to *Ln(pop)* is still significant but becomes considerably larger when *Ln(TomTom)* is included. It is worth to mention that this is not driven by the excluded cities. As a robustness check, the models of Table 4 show similar results when only the cities that have information regarding congestion are included in the estimation. This is another novelty obtained in this research: congestion has a negative and significant impact on productivity. This is robust to different specifications of the model. Moreover, the effects of agglomeration are considerably larger than observed in the previous specifications led by the inclusion of congestion, meaning that agglomeration has a larger effect than previously estimated although it is commonly overshadowed by the inherent congestion that agglomeration produces. However, the results imply that for a given city size, lowering congestion levels results in higher labor productivity. That is, congestion is a drag on agglomeration economies.

Panel B, on the other hand, considers only the individuals that are in formal sector jobs. In concordance to what was obtained in the previous specifications, the elasticity of agglomeration is reduced compared to Panel A. Interestingly, the elasticity of congestion suffers a similar effect, meaning that congestion is a drag on productivity more for those in the informal sector.

It is relevant to highlight that, even though the coefficient of the congestion variable is considerably larger than the coefficient of the population variable, this does not imply that congestion dominates the positive agglomeration effect since the range of each variable is different. Average *Ln(pop)* is at least five times higher than the average of *Ln (TomTom)* (see Table 2). A final consideration, from models (iii) and (vii) of Table 6.A and 6.B, respectively, there does not seem to be non-linear effects in the relation between agglomeration and relative wages.

Table 6.A: Linear regression considering congestion

	(i)	(ii)	(iii)	(iv)
Ln(pop)	0.23***	0.23***	0.68	0.19***
Ln(pop)^2			-0.01	
Ln(TomTom)	-0.75***	-0.77***	-0.79***	-0.64***
Female		-0.20***	-0.20***	-0.19***
Ln(age)		5.23***	5.23***	1.68***
Ln(age)^2		-0.68***	-0.68***	-0.17***
Formal				0.28***
Non-education				--
Primary/Highschool				0.30***
Technical				0.61***
Collage or upper				1.11***
Constant	2.75***	-7.17***	-10.72	-1.31**
Observations	62,111	62,111	62,111	62,111
R^2	0.08	0.13	0.13	0.35
Metro Area	13	13	13	13

* p<0.1; ** p<0.05; *** p<0.01. All regressions include country fixed effects and standard errors are clustered at the metropolitan area.

Table 6.B: Linear regression considering congestion (Only formal workers)

	(v)	(vi)	(vii)	(viii)
Ln(pop)	0.17***	0.18***	-0.79	0.17***
Ln(pop)^2			0.03	
Ln(TomTom)	-0.62***	-0.66***	-0.63***	-0.61***
Female		-0.15***	-0.15***	-0.19***
Ln(age)		4.36***	4.36***	1.65***
Ln(age)^2		-0.53***	-0.53***	-0.15**
Non-education				--
Primary/Highschool				0.33***
Technical				0.69***
Collage or upper				1.21***
Constant	3.73***	-5.02***	2.69	-0.68
Observations	40,468	40,468	40,468	40,468
R^2	0.09	0.14	0.14	0.39
Metro Area	13	13	13	13

* p<0.1; ** p<0.05; *** p<0.01. All regressions include country fixed effects and standard errors are clustered at the metropolitan area.

Finally, due to the potential endogeneity of agglomeration discussed above, the models presented in Table 7 are estimated with TSLS using population in the 19th and 20th centuries as instruments. Furthermore, congestion is also likely to endogenous. First, because there is a relation between city size and congestion. Therefore, if population levels are endogenous, then we would expect congestion to be endogenous as well in the sense that this variable will be correlated with the error term. But there could be other channels for a reverse causation between congestion and productivity. For example, more productive cities may be able to afford better mobility infrastructure, such as Rapid Transit Buses (BRT), Metro and commuting rail. Therefore, we also instrument congestion in the models presented in Table 7. As instruments we use the average number of rainy days and millimeters of precipitation in each city, since it is a well-established fact that rain leads to a deterioration in driving conditions in urban settings and it is clearly exogenous with respect to productivity.

Since Brasilia and Fortaleza do not register historic population, these cities had to be removed from this specification resulting in 11 cities in the sample.

First, notice from Table 7 panel A and B that all R^2 of the first stages are large and all F-statistics are above 10, which suggests that these are strong instruments. Furthermore, all independent variables remain robust and consistent to previous specifications.

The elasticity of agglomeration on productivity remains consistent compared to what was observed in Table 6. Furthermore, the effects of congestion on productivity are still significant and consistent to the results evidenced previously. Finally, Panel B also shows robustness in the sense that the effects are slightly reduced compared to Panel A. According to these results, agglomeration significantly increases productivity, at least 0.20% for all workers or 0.17% for formal workers when population increases by a 1%.

Table 7.A: Linear regression instrumenting both population and congestion

	(i)	(ii)	(iii)	(iv)
Ln(pop)	0.23***	0.24***	2.58	0.20***
Ln(pop)^2			-0.07	
Ln(TomTom)	-0.64**	-0.69***	-1.00***	-0.62***
Female		-0.19***	-0.19***	-0.18***
Ln(age)		5.05***	5.05***	1.59***
Ln(age)^2		-0.65***	-0.65***	-0.16***
Formal				0.29***
Non-education				--
Primary/Highschool				0.30***
Technical				0.60***
Collage or upper				1.10***
Constant	2.23*	-7.26***	-25.13	-1.37
Observations	52,793	52,793	52,793	52,793
R^2	0.08	0.13	0.13	0.34
Metro Area	11	11	11	11
First-stage's R^2				
Ln(TomTom)	0.95	0.95	0.96	0.95
Ln(pop)	0.93	0.93	0.96	0.93
Ln(pop)2			0.96	

* p<0.1; ** p<0.05; *** p<0.01

Notice that all the F-statistic of the first-stage regression are larger than 10. All regressions include country fixed effects and standard errors are clustered at the metropolitan area.

Table 7.B: Linear regression instrumenting both population and congestion (Only formal workers)

	(v)	(vi)	(vii)	(viii)
Ln(pop)	0.17***	0.18***	0.07	0.17***
Ln(pop)^2			0.00	
Ln(TomTom)	-0.47***	-0.54***	-0.61***	-0.57***
Female		-0.14***	-0.14***	-0.18***
Ln(age)		4.15***	4.15***	1.63***
Ln(age)^2		-0.51***	-0.51***	-0.15**
Non-education				--
Primary/Highschool				0.33***
Technical				0.68***
Collage or upper				1.19***
Constant	3.27***	-4.99***	-3.93	-0.8
Observations	33,773	33,773	33,773	33,773
R^2	0.08	0.13	0.13	0.38
Metro Area	11	11	11	11
First-stage's R^2				
Ln(TomTom)	0.94	0.94	0.93	0.94
Ln(pop)	0.93	0.93	0.93	0.93
Ln(pop)2			0.93	

* p<0.1; ** p<0.05; *** p<0.01

Notice that all the F-statistic of the first-stage regression are larger than 10. All regressions include country fixed effects and standard errors are clustered at the metropolitan area.

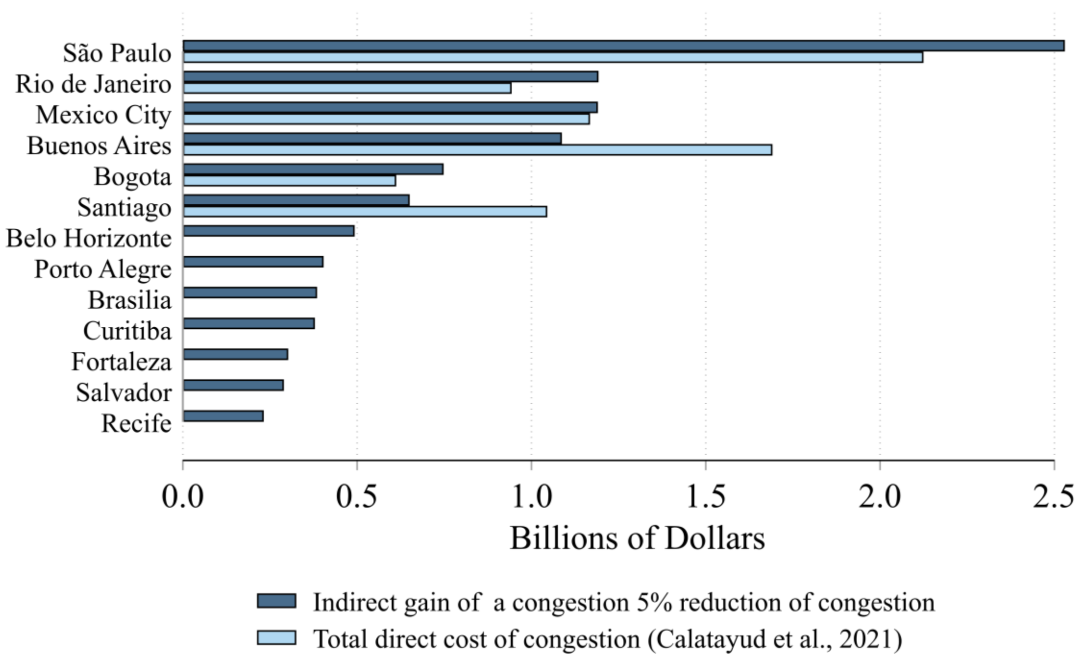
7. Valuing the productivity cost of congestion

In this section, we estimate the impact on productivity of a reduction in congestion. To do so, we simulated a 5% reduction in our congestion variable due, for example, to transit investments or other mobility improving policies within a city. Using the above econometric results, we estimate the impact of the mobility improvement on wages. We then sum the wage increase for all workers in each city as a measure of the aggregate productivity impact. We undertake this simulation for the 13 cities where information on congestion is available.

A 5% reduction in congestion would represent a regional average time savings equivalent to 2.1% of free-flow travel times, and up to 3.4% for the case of Bogota, the city with the largest relative congestion figure (see Table 2).

Using the parameter estimates from column (iv) of Table 7.A, the 5% reduction in congestion would represent almost 10 billion dollars in productivity gains for the 13 selected cities in 2019.[25] In total, 4 cities in the region would see economic benefits rise by more than 1 billion dollars, these are São Paulo (2.5 billion); Rio de Janeiro (1.2 billion); Mexico City (1.2 billion); and Bueno Aires (1.1 billion). The metropolitan areas with the smallest gain would be Fortaleza (302 million); Salvador (290 million); and Recife (232 million), nonetheless, this is mainly due to population size differences among cities. These results are of high relevance considering that they are almost as high as the total direct cost of congestion (Calatayud et al., 2021). For instance, the indirect productivity gain from a 5% reduction in congestion would represent 62% of the total cost of congestion in Santiago and 64% for Buenos Aires, and up to 120% for the cases of Bogota and São Paulo.

Figure 9: Productivity gain from 5% congestion reduction

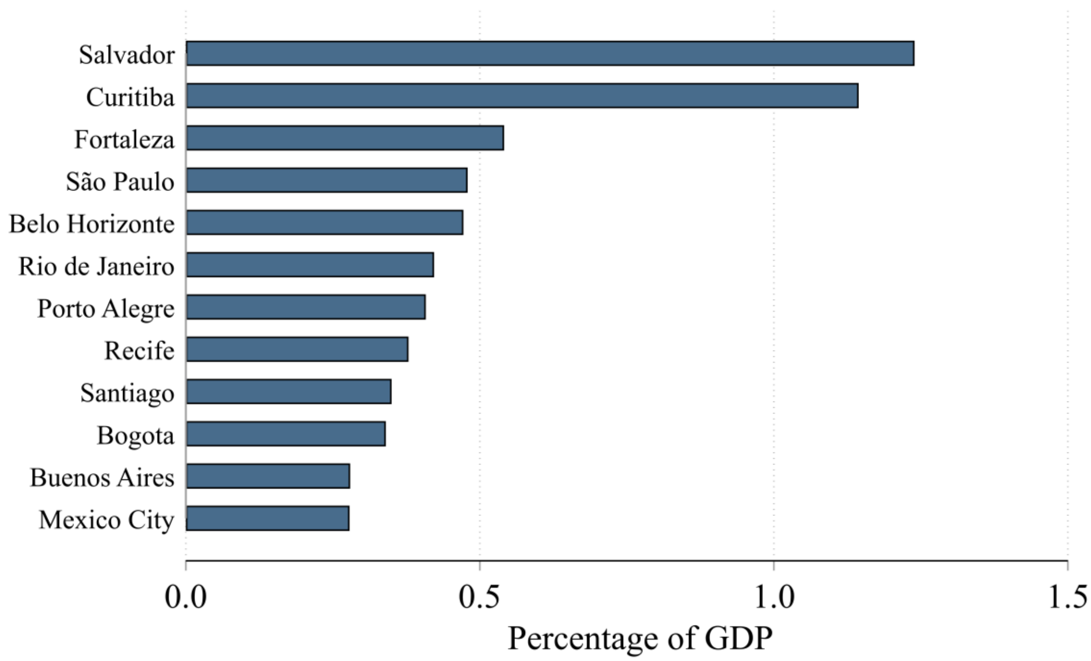


Source: Own elaboration based on Household surveys from national statistic institutes; and Calatayud et al., 2021

[25] This benefit does not consider the gains to landowners due to higher rents when congestion is lowered (see Figures 3 and 4 above). We thank Juan Pablo Chauvin for pointing this out to us.

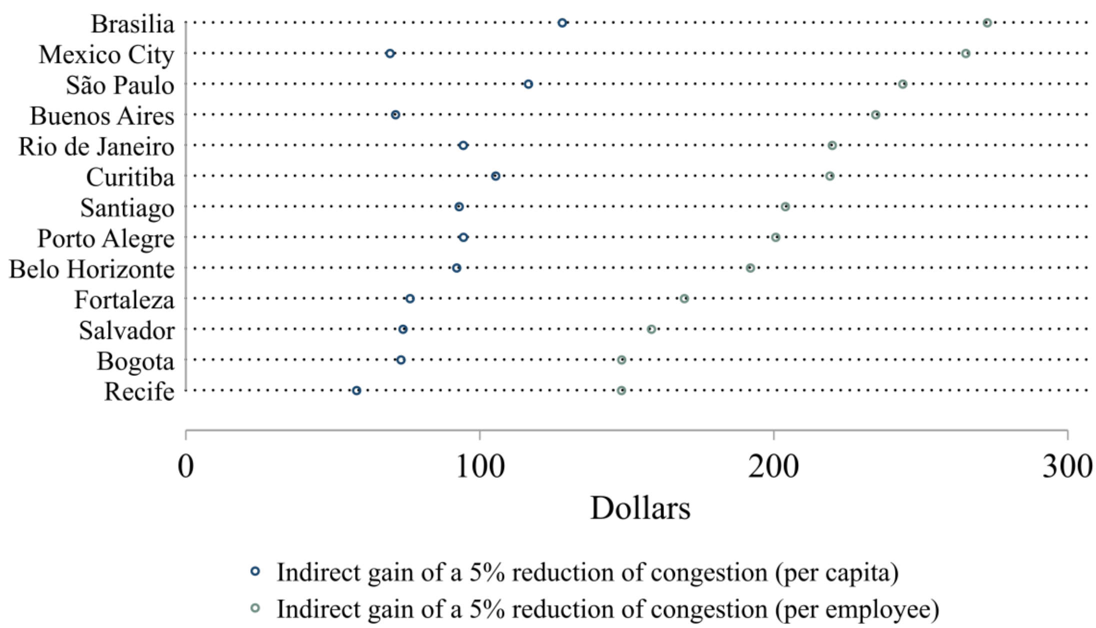
Figure 11 presents the results relative to the city's size in two dimensions, according to the number of residents and employed persons. Brasilia is the city that presents the largest gains both per capita (130 dollars) and per employee (272 dollars). Interestingly, Mexico City would have a relatively large gain per employee but low gain per capita from this reduction in congestion. On average, the annual productivity gain would be 88 dollars per person in the region and 206 per employee. On the other hand, the cities with the largest gains relative to its own economy would be Salvador and Curitiba representing more than 1.1% of GDP. This same figure would be at least 0.3% for the cases of Mexico and Buenos Aires. On average, the productivity gain represents 0.5% of GDP for the cities in the sample.

Figure 10: Productivity gain from 5% congestion reduction relative to GDP



Source: Own elaboration based on Household surveys from national statistic institutes; and Metroverse Growth Lab

Figure 11: Relative productivity gain from 5% congestion reduction



Source: Own elaboration based on Household surveys from national statistic institutes

8. Conclusions

This study evidences the benefits of agglomeration in five of the most populated countries in Latin America. Three novelties are included in this research. Firstly, consideration of a cross-country analysis using household surveys and a new proposed indicator to make them comparable. Inclusion of a measure of congestion to determine the real effects of agglomeration on productivity and the loss due to perturbation on mobility. Finally, instrumenting current population levels using population in the 19th and early 20th centuries and congestion using millimeters of precipitation and average number of rainy days in the year.

The results from this article are consistent with what has been evidenced in the literature for developing countries. The agglomeration elasticity was estimated to be about 0.04 when considering the whole sample of workers and 0.03 when only formal sector workers are included in the model. Additionally, these results remain robust and consistent to different set of control variables, specifications of the model, and instrumental variable estimation.

When considering congestion in the model, the effects of agglomeration rise significantly. The gross agglomeration effect is four times larger when the full sample of workers is used in the estimation, presenting an elasticity of at least 0.20. The coefficient when considering only formal workers in the sample also rises considerably, to at least 0.17. Congestion has the opposite effect implying a drag on productivity growth due to agglomeration. This last result holds even when instrumenting congestion.

Finally, to get a sense of the magnitude of the congestion effect, we simulated a decrease in 5% in congestion levels in each city where this information is available, maintaining city size constant. This could be the result of transit investments that improve mobility within a city, for example. According to our model, the results indicate that the productivity benefits would reach an average 0.5% of GDP among the 13 cities in the sample. This represents a net gain of at least 214 million dollars for Recife, and up to 2.5 billion dollars for São Paulo. On average these benefits amount to 206 dollars per employee in the region, much higher than the direct time cost savings from the 5% reduction in congestion.

The main limitation of this paper is that cross section data was used and thus we were unable to use worker fixed effects to control for sorting based on unobservable individual characteristics. Future research should attempt to address this limitation using panel data.

8. References

Alcalá Ferráez, C. (2015). Población en la ciudad de Campeche: fuentes y tendencias, 1810-1861. *Secuencia*, 92.

Bacolod, M., J. De la Roca and M.M Ferreyra (2021), 'In search of better opportunities: Sorting and agglomeration effects among young college graduates in Colombia', *Regional Science and Urban Economics*, 87, pp. 1-15.

Barufi, A. M. B., E. A. Haddad and P. Nijkamp, P. (2016), 'Industrial scope of agglomeration economies in Brazil', *The Annals of Regional Science,* 56(3), 707–755. https://doi.org/10.1007/s00168-016-0768-3

Bernedo Del Carpio, M. and C. Patrick (2021), 'Agglomeration and informality: Evidence from Peruvian establishments', *Journal of Regional Science*, 61(2), pp.442-471.

Brinkman, J.C. (2016), 'Congestion, agglomeration and the structure of cities', *Journal of Urban Economics*, 94, pp. 13-31.

Brueckner, J.K. (2001), *Lecture on Urban Economics*, MIT Press.

Bundio, S. (2017). Movilización geográfica humana como consecuencia del desarrollo turístico en Acapulco (1950-1980) [Universidad Autonoma de Guerrero].

Calatayud, A., S. Sánchez González, F. Bedoya Maya, F. Giraldez, J. M. Márquez. 2021. Congestión urbana en América Latina y el Caribe: características, costos. Washington, D.C.: BID.

Calderón-Mejía, V. and A.M. Ibañez (2016), `Labour market effects of migration-related supply shocks: evidence from internal refugees in Colombia', Journal of Economic Geography, 16(3), pp. 695-713.

Cameron, A. C., & Trivedi, P. K. (2005). Microeconometrics: methods and applications. Cambridge university press.

Chauvin, J. P., E. Glaeser, Y. Ma and K. Tobio (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. https://doi.org/10.1016/j.jue.2016.05.003.

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

Combes, P.-P., S. Démurger and S. Li (2013), 'Urbanisation and Migration Externalities in China', CEPR Discussion Paper No. DP9352, Available at SSRN: https://ssrn.com/abstract=2221752

Combes, P.-P., G. Duranton, L. Gobillon and S. Roux (2010), Chapter 1: Estimating Agglomeration Economies with History, Geology, and Worker Effects, in E.L. Glaeser (ed.), *Agglomeration Economies*, Chicago University Press.

Combes, P.-P., G. Duranton, and L. Gobillon (2019), 'The Costs of Agglomeration: House and Land Prices in French Cities', *Review of Economic Studies*, 86 (4), pp. 1556–89.

Combes, P.-P. and L. Gobillon (2015), Chapter 5: The Empirics of Agglomeration Economies, in

D'Costa, S. and H.G. Overman (2014), 'The urban wage growth premium: Sorting or learning?', *Regional Science and Urban Economics*, 48, pp. 168-179.

Duranton, G., J. V. Henderson and W. Strange (eds.), *Handbook of Regional and Urban Economics*, volume 5b, Amsterdam: Elsevier.

Duranton, G. (2016), 'Agglomeration effects in Colombia', *Journal of Regional Science*, 56(2), pp. 210-238.

Duranton, G. and D. Puga (2019), 'Urban Growth and Its Aggregate Implications', NBER Working Paper 26591.

Duranton, G. and D. Puga (2020), 'The Economics of Urban Density', *Journal of Economic Perspectives*, 34(3), pp. 3-26.

Ehrl, P. and L. Monasterio (2021), 'Spatial skill concentration agglomeration economies', *Journal of Regional Science*, 61(1), pp.140-161.

Glaeser, E.L. (2010), Introduction, in E.L. Glaeser (ed.), *Agglomeration Economies*, Chicago University Press.

Glaeser, E.L and D.C. Maré (2001), 'Cities and Skills', *Journal of Labor Economics*, 19(2), pp. 316-342.

Gomez-Lobo, A. and A. Micco (2021), 'Urban commuting distance and medical licenses: an empirical study of Santiago, Chile', mimeo, Department of Economics, University of Chile.

Graham, D.J. (2007a), 'Agglomeration, Productivity and Transport Investment', *Journal of Transport Economics and Policy*, 41(3), pp. 317–343.

Graham, D.J. (2007b), 'Variable returns to agglomeration and the effect of road traffic congestion', *Journal of Urban Economics*, 62, pp. 103–120.

Guevara, C. (2016), 'Growth Agglomeration Effects in Spatially Interdependent Latin American Regions', GATE Working No. 1611, 2016, Available at SSRN: https://ssrn.com/abstract=2741025 or http://dx.doi.org/10.2139/ssrn.2741025.

Künn-Nelen, A. (2015), 'Does commuting affect health?', *Health Economics*, 25, pp. 984-1004.

Lattes, A. (1995), 'Urbanización, crecimiento urbano y migraciones en América Latina', Pensamiento Iberoamericano, No 28, Chile, pp. 211-260.

Lattes, A., J. Rodríguez and M. Villa (2004), Population dynamics and urbanization in Latin America. Concepts and data limitations, in Champion, T. and H. Graeme (eds), New Forms of Urbanization: Beyond the Urban-Rural Dichotomy, Routledge, London.

Lopez, R. and J. Suedekum, (2009), 'Vertical industry relations, spillovers, and productivity: evidence from Chilean plants', *Journal of Regional Science*, 49, pp. 721–747.

Lucas, R.E., E. Rossi-Hansberg (2002), ‘On the internal structure of cities’, *Econometrica*, 70(4), pp. 1445–1476.

Marshall, A. (1890), *Principles of Economics*, Macmillan, London.

Matano, A., M. Obaco and V. Royuela (2020), ‘What drives the spatial wage premium in formal and informal labor markets? The case of Ecuador’, *Journal of Regional Science*, 60(4), pp.823-847.

Morrison, A. R. (1993), ‘Violence or Economics: What Drives Internal Migration in Guatemala?’, Economic Development and Cultural Change, 41(4), pp. 817-831.

Quintero, L. and M. Roberts (2018), ‘Explaining spatial variations in productivity: evidence from Latin America and the Caribbean’, World Bank Policy Research Working Paper, (8560).

Rosenthal, S.S. and W.C. Strange (2004), Evidence on the nature and sources of agglomeration economies, in Henderson, V. and J.F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, Vol. 4. North-Holland, Amsterdam, pp. 2119–2171.

Ross, S.L. and Y. Zenou (2008), ‘Are shirking and leisure substitutable? An empirical test of efficiency wages based on urban economic theory’, *Regional Science and Urban Economics*, 38(5), pp. 498-517.

Saito, H. and M. Gopinath (2009), ‘Plants' self-selection, agglomeration economies and regional productivity in Chile’, *Journal of Economic Geography*, 9(4), pp. 539-558.

Safirova, E. (2002), 'Telecommuting, traffic congestion, and agglomeration; a general equilibrium model’, *Journal of Urban Economics*, 52, pp. 26-52.

Sánchez, G. (2016). Crecimiento, modernización y desigualdad regional. La Belle Époque argentina. *Estudios Avanzados*, 25(42–67).

Silva, D. and C. Azzoni (2021), ‘Worker and firm heterogeneity, agglomeration, and wages in Brazil’, *Papers in Regional Science*, 2021, 1-27. DOI: 10.1111/pirs.12637.

Stock, J. H., and M. Yogo (2002), ‘Testing for weak instruments in linear IV regression’, National Bureau of Economic Research.

Sullivan, A.M. (1983), ‘The General Equilibrium Effects of Congestion Externalities’, Journal of Urban Economics, 14, pp. 80-104.

Sweet, M. (2014), ‘Traffic Congestion’s Economic Impacts: Evidence from US Metropolitan Regions’, Urban Studies, 51(10), pp. 2088-2110.

Toledo, S. P., Klein, H. S., & Pulido, L. E. (1992). La población de la ciudad de Zacatecas en 1857. *Historia Mexicana*, 42(1), 77–102.

Venables, A. (2007), ‘Evaluating Urban Transport Improvements: Cost-Benefit Analysis in the Presence of Agglomeration and Income Taxation’, Journal of Transport Economics and Policy, 41(2), pp. 173-188.

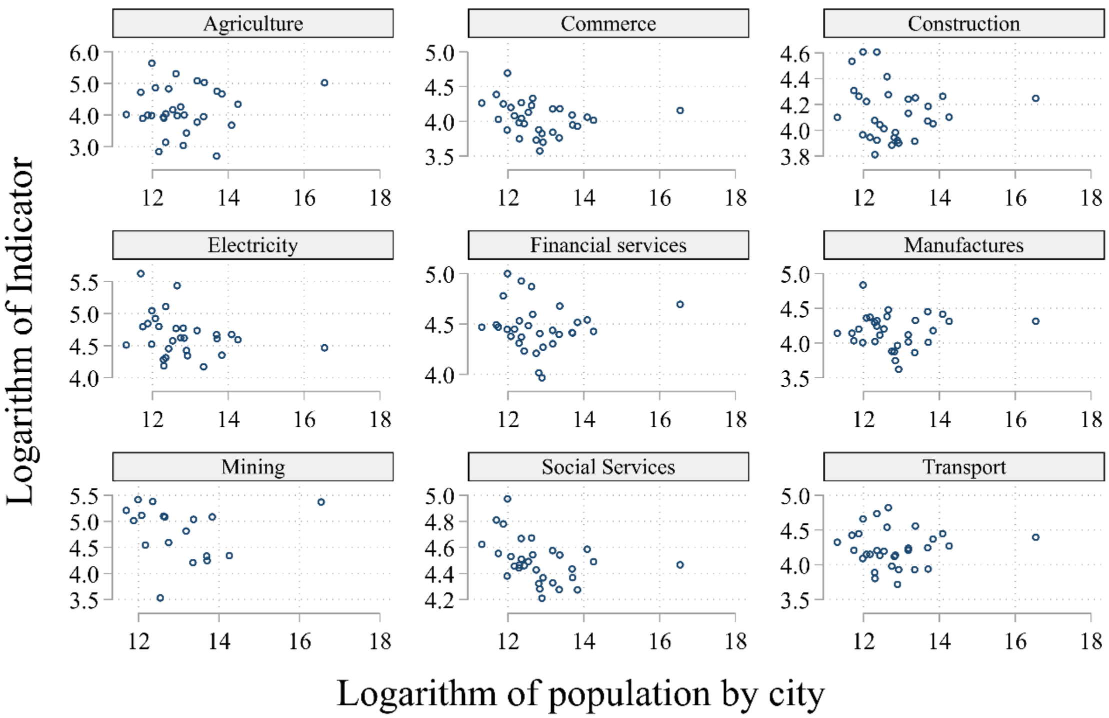
Van Ommeren, J.N. and E. Gutiérrez-i-Puigarnau (2011), ‘Are workers with long commute less productive? An empirical analysis of absenteeism’, *Regional Science and Urban Economics*, 41, pp. 1-8.

Vargas, G., & Vargas, G. (2016). Aceptabilidad social, forma urbana y sustentabilidad de barrios urbanos en Ciudad Juárez, Chihuahua. *Revista Oper*, 19.

Zárate, R.D. (2021), `Spatial Misallocation, Informality, and Transit Improvements: Evidence from Mexico City', mimeo, The World Bank, January.

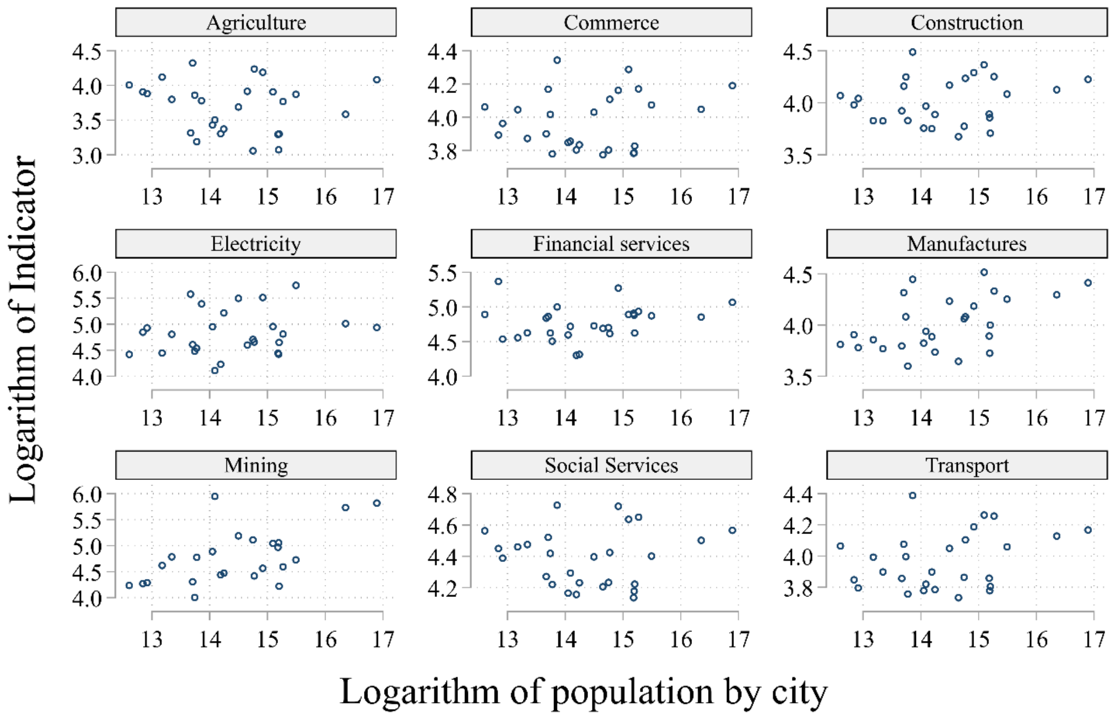
9. Annex

Figure A.1: Index and population size by country and economic sector
Figure A.1.1: Argentina



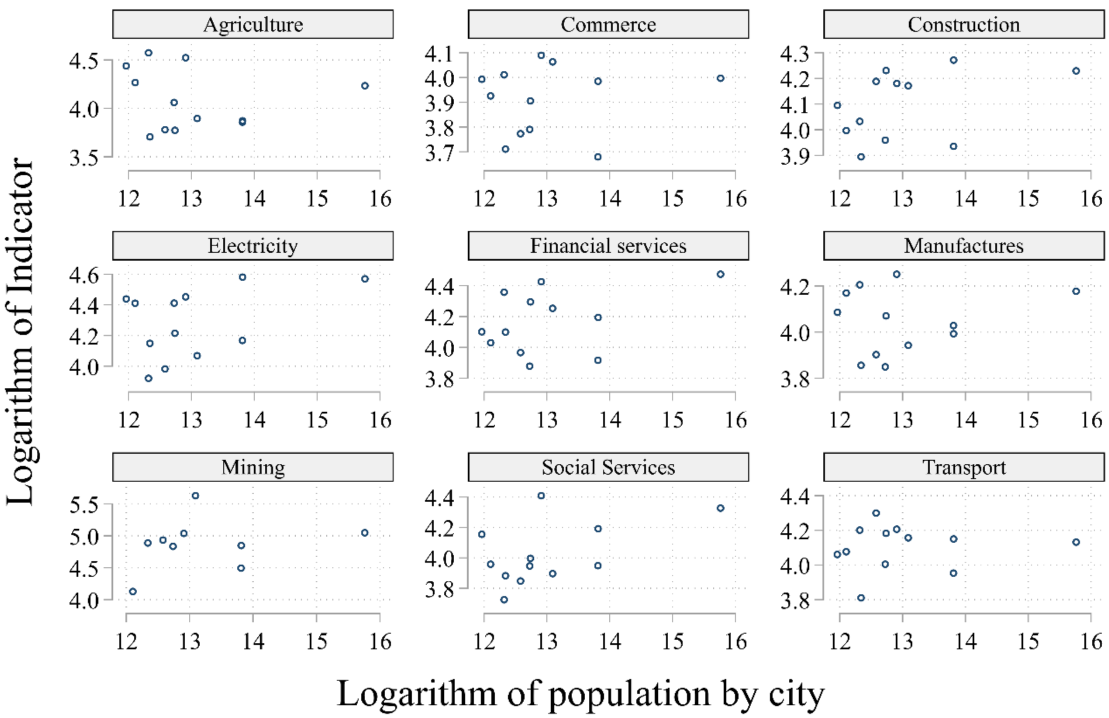
Source: Authors' elaboration with data from national household surveys.

Figure A.1.2: Brazil



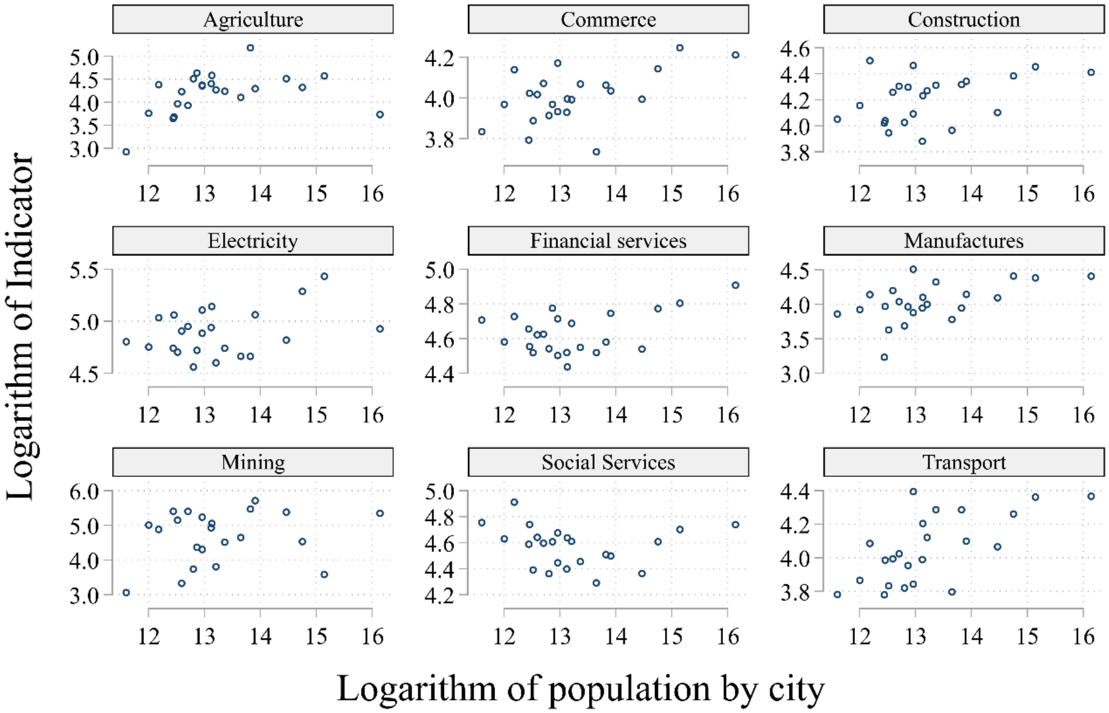
Source: Authors' elaboration with data from national household surveys.

Figure A.1.3: Chile



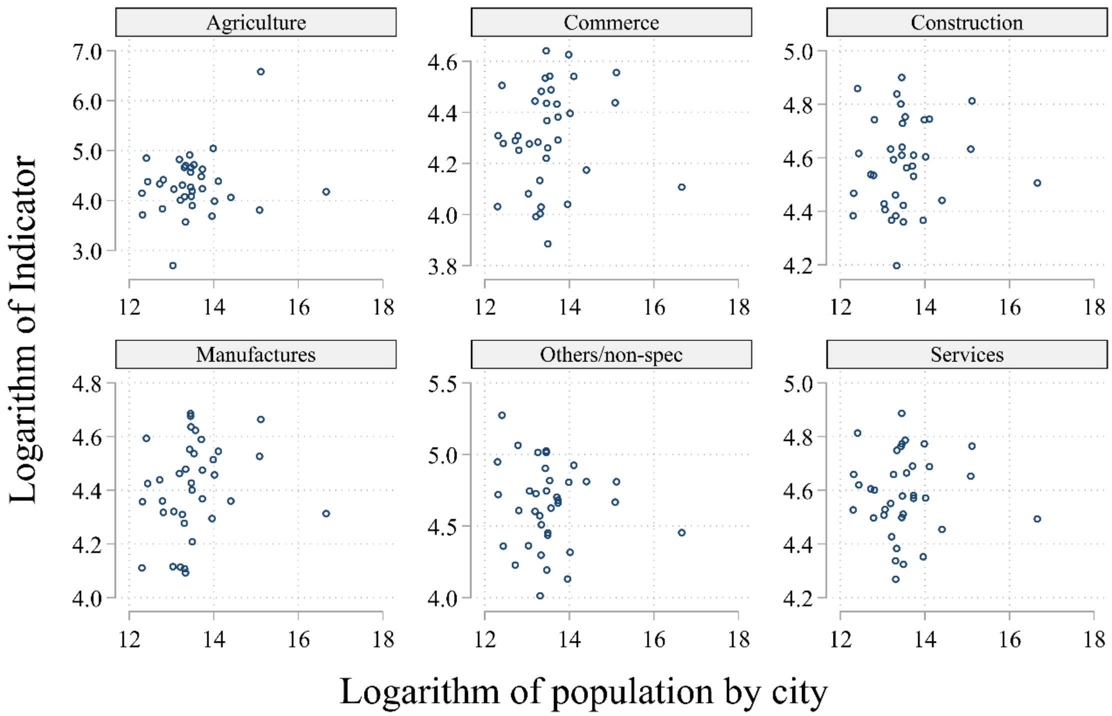
Source: Authors' elaboration with data from national household surveys.

Figure A.1.4: Colombia



Source: Authors' elaboration with data from national household surveys

Figure A.1.5: Mexico



Source: Authors' elaboration with data from national household surveys

Table A.1: Population Argentina

City	Year	Population	Year	Population	Source
Buenos Aires	1857	91000	1887	438000	(Morse, 1974)
	1869	178000	1895	678000	
Formosa	1914	1576000	1947	16500	(Gobierno de la Provincia de Formosa, 2020)
Jujuy	1880	44077	1914	77511	(Sánchez, 2016)
	1895	49713			
La Rioja	1880	56794	1914	79754	
	1895	69502			
Mar del Plata	1881	4030	1914	32940	(INDEC, 2010)
	1890	8639	1947	115000	
Neuquen	1914	2152	1947	7498	(KLOSTER, 1993)
Rawson	1881	1007			(INDEC, 2010)
Rio Gallegos	1895	150	1920	2912	(Cáceres, 2012)
	1912	1557	1947	5880	
Santa Rosa	1914	5487	1947	14623	
	1920	6383			
Usuahia	1947	2182			(INDEC, 2010)
Viedma	1910	3000			(INDEC, 2010)

Table A.2: Population Brazil

City	Year	Population	Year	Population	Source
Belo horizonte	1900	13472	1940	211377	(IBGE, 1950)
	1920	55563	1950	352724	
Cuiaba	1872	35987	1920	33678	(IBGE, 1950)
	1890	17815	1940	54394	
	1900	34393	1950	56204	
Curitiba	1854	15000	1900	50000	(Morse, 1974)
	1872	13000	1920	79000	
	1890	25000			
Fortaleza	1854	35000	1900	48000	
	1872	42000	1920	79000	
	1890	41000			
Porto Alegre	1854	30000	1900	74000	
	1872	44000	1920	180000	
	1890	52000			
Recife	1854	86000	1900	113000	
	1872	117000	1920	239000	
	1890	112000			
Rio de Janeiro	1854	186000	1900	688000	
	1872	267000	1920	1158000	
	1890	523000			
Salvador	1854	108000	1900	206000	
	1872	129000	1920	283000	
	1890	174000			
Sao Paulo	1854	26000	1900	240000	
	1872	31000	1920	579000	
	1890	65000			

Table A.3: Population Chile

City	Year	Population	Year	Population	Source
Chillan	1907	34269			(Espinoza, 2013)
Santiago	1865	115000	1895	256000	(Morse, 1974)
	1875	130000	1907	333000	
	1885	189000	1920	507000	

Table A.1.4: Population Colombia

City	Year	Population	Year	Population	Source
Armenia	1938	50383			(Banco de la República, 2020a)
Bogota	1851	40000	1905	100000	(Morse, 1974)
	1870	41000	1912	121000	
	1884	96000	1918	143000	
Bucaramanga	1938	51283	1938	51283	(Banco de la República, 2020b)
Cucuta	1851	6353	1918	29490	(Banco de la República, 2020c)
	1864	7345	1923	40151	
	1870	9226	1928	49279	
	1896	17475	1938	57248	
	1912	25955		25955	
Neiva	1851	7719	1918	29938	(Alcaldía de Neiva, 2011)
	1870	8332	1938	34294	
	1912	21852		21852	
Pereira	1870	633	1928	50699	(Alcaldía de Pereira, 2020)
	1905	19036	1938	60492	
	1918	23584		23584	
Popayan	1938	30038	1938	30038	(Banco de la República, 2020d)
Santa Marta	1851	4340	1912	8348	(Banco de la República, 2020e)
	1871	5742	1918	18040	
	1905	9568	1938	33245	
Sincelejo	1852	6046	1918	14722	(Viloria, 2001)
	1870	11336		11336	
Valledupar	1938	3339	1938	3339	(Garcia, 1999)

Table A.1.5: Population Mexico

City	Year	Population	Year	Population	Source
Acapulco	1900	4932	1930	6529	(Abundio, 2017)
	1910	5900	1940	9993	
	1921	5768	1950	28512	
Aguascalientes	1861	22543	1900	35052	(Rojas et al., 2020)
	1879	20327	1910	45198	
	1857	20000	1895	31169	(INEGI, 1985)
	1862	22534	1900	34982	
	1862	20907	1910	45198	
	1865	20000	1921	48041	
	1869	31842	1930	62244	
	1878	32000	1940	82234	
	1882	35000	1950	93363	
	1895	30872			
Campeche	1853	15357	1861	15197	(Alcalá Ferráez, 2015)
	1853	15000			
Cd. Juarez	1856	4342	1939	48881	(Vargas & Vargas, 2016)
	1894	7582	1950	122566	
	1911	11289			
Chihuahua	1895	18279	1930	45595	(INEGI, 1985)
	1900	30405	1940	56805	
	1910	39706	1950	87000	
	1921	37078			
	1859	14000	1895	18521	(INEGI, 1985)
	1862	12000	1900	30405	
	1869	12000	1910	39706	
	1870	10000	1921	37078	
	1882	16000	1930	45595	
	1882	28000	1940	56805	
	1884	25000	1950	86961	
	1895	18279			
Colima	1857	31774	1910	25148	
	1861	41974	1921	28326	
	1877	23579	1930	21117	
	1895	18977	1940	22601	
	1900	30698	1950	28658	
Cuernavaca	1877	12000	1921	7117	
	1888	6342	1930	8554	
	1895	8717	1940	14336	
	1900	9581	1950	30597	
	1910	12776			
Culiacan	1900	10380	1930	18202	(INEGI, 1985)
	1910	13527	1940	22025	
	1920	16034	1950	48936	
	1857	9647	1910	13527	(INEGI, 1985)

	1877	8000	1921	16034	
	1886	10000	1930	18202	
	1895	10487	1940	22025	
	1900	10380	1950	48963	
Durango	1855	16060	1900	31092	
	1856	12499	1910	31763	
	1859	17500	1921	39091	
	1862	16014	1930	35330	
	1869	12000	1940	33412	
	1893	24800	1950	59496	
	1895	26425			
Hermosillo	1877	8000	1921	14745	
	1890	7071	1930	19959	
	1895	8474	1940	18601	
	1900	10613	1950	43516	
	1910	14578			
La Paz	1857	1274	1910	5536	
	1861	2276	1921	7480	
	1877	1000	1930	8166	
	1895	4737	1940	10401	
	1900	5046	1950	13081	
Mexico City	1855	200000	1900	344000	(Morse, 1974)
	1862	210000	1910	471000	
	1877	230000	1921	615000	
	1884	300000			
Morelia	1852	25000	1895	33890	(INEGI, 1985)
	1857	25000	1895	32287	
	1862	12335	1900	37278	
	1869	25000	1910	40042	
	1882	25000	1921	31148	
	1882	25000	1930	39916	
	1884	24000	1940	44304	
	1890	26974	1950	63245	
Oaxaca	1855	24000	1884	28000	
	1857	25000	1889	29038	
	1861	28750	1895	32437	
	1863	24433	1900	35049	
	1865	24907	1910	38011	
	1868	19200	1921	27792	
	1877	26051	1930	33423	
	1880	30000	1940	29306	
	1881	27583	1950	46741	
	1882	27822			
Pachuca	1850	4000	1904	50981	
	1852	5442	1910	39009	
	1864	12000	1921	40802	
	1869	15000	1930	43023	
	1895	40487	1940	53354	

	1897 1900	40000 37487	1950	58650	
Queretaro	1854 1857 1861 1873 1877 1895	27456 27496 27492 27560 27580 34576	1900 1910 1921 1930 1940 1950	33152 33062 30073 32585 33629 49209	
San Luis Potosi	1857 1861 1877 1895 1900	10678 26841 34000 69050 61019	1910 1921 1930 1940 1950	68022 57353 74003 104481 162446	
Tlaxcala	1857 1861 1877 1889 1895 1900	3463 5634 4000 6761 1879 2715	1910 1921 1930 1940 1950	2812 2069 2403 3261 5071	
Toluca	1857 1861 1877 1895 1900	12000 18794 12000 2315 25940	1910 1921 1930 1940 1950	31023 34265 41234 43429 52068	
Tuxtla Gutierrez	1870 1877 1892 1895 1900	6963 10500 6000 10951 9395	1910 1921 1930 1940 1950	10239 12517 14849 15883 28280	
Villahermosa	1857 1861 1895 1900 1910	5500 7300 9604 10543 12327	1921 1930 1940 1950	15819 15395 25114 33587	
Zacatecas	1856 1856	16451 15427	1857 1858	21417 25005	(Toledo et al., 1992)