Agglomeration effects in Colombia

Gilles Duranton[*‡]
University of Pennsylvania

February 2014

Abstract: I estimate an elasticity of wages with respect to city population of about 5% for Colombian cities. This finding is robust to a number of econometric concerns. The second main finding is a negative effect of market access on wages. I find only mild evidence in favour of human capital externalities and no evidence of a complementarity between cities and skills. Despite precise estimates, I do not find stronger agglomeration effects for older workers as would be predicted by the existence of learning effects nor do find I sizeable effects of roads or amenities on wages.

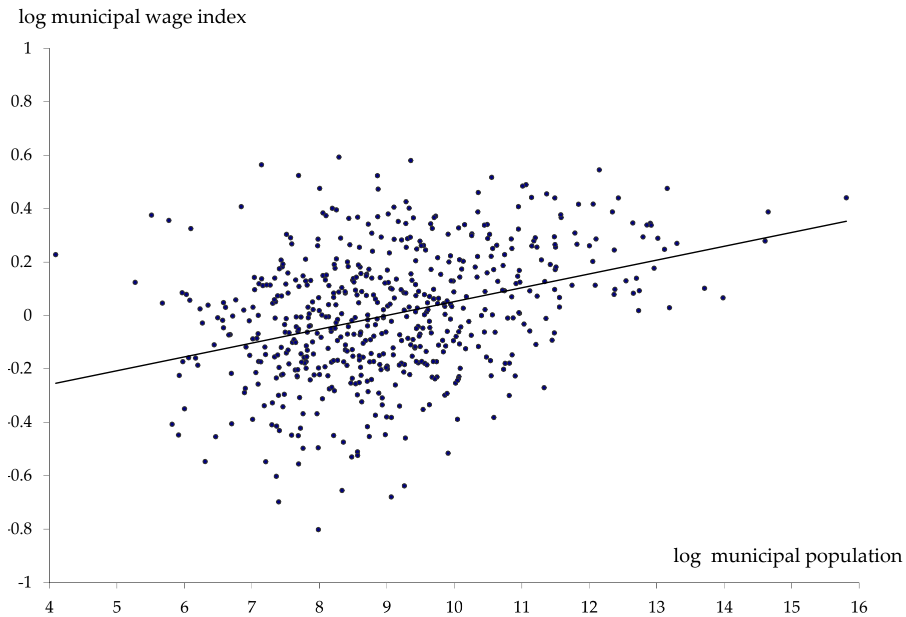
Key words: agglomeration, Colombia, city population, market access

JEL classification: R12, R23

[*]Thanks to Prottoy Akbar for terrific research assistance. I am grateful to Jose Antonio Pinzon, Diana Lopez, Adolfo Meisel, and Fabio Sanchez for providing me data and helping with them. Thanks also to Kristian Behrens and Thierry Mayer for feedback and comments. Financial support from the Colombian Ministry of Planning (DNP) and UN-Habitat is gratefully acknowledged.

[‡]Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia, PA 19104, USA (e-mail: `duranton@wharton.upenn.edu`; website: `https://real-estate.wharton.upenn.edu/profile/21470/`).

Figure 1: Population and wages in Colombian municipalities, 2008-2012



The vertical axis represents centered (natural) log municipal wages computed using 2008-2012 individual wage data after conditioning out year effects and individual characteristics. The horizontal axis represents log 2010 population in the urban part of municipalities. There are 583 municipalities. The slope of the regression line is 0.051 and the R-squared is 12%.

1. Introduction

In this paper I estimate agglomeration effects for cities in Colombia. My preferred estimate regarding the elasticity of wages with respect to city population is about 5%. Figure 1 provides a summary illustration for the recent years. The slope of the regression line in the figure is 5.1%. This finding is robust to a number of econometric concerns. My second main finding is a negative effect of market access on wages. Places with good external market access actually offer lower nominal wages. This is consistent with better market access allowing local workers to consume a broader variety of goods at a lower price coupled with a notion of spatial equilibrium leading to real wage equalisation. To my knowledge, this result is novel in the literature. I find only mild evidence in favour of human capital externalities in cities. My other findings are mostly negative. I find no evidence of a complementarity between cities and skills nor do I find evidence of stronger agglomeration effects for older workers as would be predicted by the existence of learning effects. I also find no effects of roads or amenities on wages. Given the precision of most estimates, these negative results should be thought as zeroes rather than inconclusive findings.

This investigation is important for several reasons. First, many surveys on urbanisation in developing countries lament the paucity of results about agglomeration economies in these countries (Henderson, 2005, Overman and Venables, 2005, Duranton, 2014*a*). This analysis hopes to be part of a broader knowledge base about urbanisation and its effects in developing countries. For instance, the elasticity of wages with respect to city population of 5% found for Colombia appears slightly larger than comparable estimates for cities in developed counties as discussed in Puga (2010). At the same time, this elasticity of 5% is also smaller than recent findings for less advanced developing countries like China or India (Combes, Démurger, and Shi, 2013, Chauvin, Glaeser, and Tobio, 2013). This perhaps suggests that agglomeration effects are stronger in less advanced countries. In turn, if this is confirmed, understanding this feature in greater depth should become an important research priority.

Second, this study wishes to take a *broad look* at agglomeration effects in Colombia. Most studies in the literature focus on specific dimensions of agglomeration economies or specific identification issues. For instance Ciccone and Hall (1996) and many others have focused on the possible endogeneity of city population. Rauch (1993), Moretti (2004) and many of their followers have explored the effect of city education on wages. Glaeser and Maré (2001) and De la Roca and Puga (2012) emphasised possible learning effects associated with larger cities. Glaeser and Resseger (2010) also suggest a complementarity between education and city population, etc. Although data and space limitations make impossible to take into account everything that previous literature has considered, it is important to develop a more integrated approach that brings together most of the issues highlighted by past research. This more integrated approach should also be useful to assess the robustness of many previous piecemeal findings and offer a broad assessment of agglomeration effects in Colombia.

Finally, knowing more about the benefits from cities is important for Colombia. As the Government of Colombia makes significant change to its system of municipal finance and ponders more broadly what sort of urban policy it intends to develop and implement (Samad, Lozano-Gracia, and Panman, 2012, Barco, 2013), it is important to know more about the economic benefit and costs of cities in Colombia. This paper is part of that enterprise.

To fulfill the objectives of this paper, one key tradeoff must be resolved. The need to consider many aspects of agglomeration benefits jointly precludes the use of some of the econometrically more advanced approaches that have been developed in the literature that are often extremely data-intensive and thus impossible to implement for most developing countries. For instance

Combes, Duranton, Gobillon, and Roux (2010) use a rich panel of workers to control for unobserved individual characteristics that may be correlated with location choices. Such rich data is not available in Colombia or in most developing countries. Hence, expanding our knowledge base on agglomeration effects in developing countries will require some methodological compromises. At the same time, serious identification concerns should not be ignored. Put differently, although this study does attempt to push the methodological frontier on the study of agglomeration effects, it endeavours to use the best available approaches given the data at hand.

Because comparability across countries is important, this study will focus mostly on widely available data.[1] It will also report both positive and negative results. Despite being often perceived unhelpful for publication, negative results should be reported. Otherwise it is impossible to know whether a relationship is absent because of a lack of result or because of a lack of investigation.

The rest of this paper is organised as follow. Section 2 describes the data. Section 3 discusses some simple agglomeration theory, its predictions, and the associated econometrics. Section 4 presents the main results regarding the elasticity of wages with respect to city population. Section 5 reports and discusses results about human capital externalities, market access, and infrastructure and amenities. Finally, section 6 concludes.

2. Data

To measure agglomeration effects, much of extant research uses large-scale administrative data either at the worker or at the firm level. This type of data is unfortunately inappropriate for cities in developing countries where half or more of the labour force is in the informal sector. I rely instead on the Colombian Great Integrated Household Survey (GEIH) administered by the Colombian National Statistical Agency (DANE) between 1996 and 2012 (with 2006 and 2007 missing). This survey provides detailed information about a large cross-section of Colombian households including age, gender, year of education, municipality of residence, number of hours worked, and labour income. The initial data contains 5,437,894 observations ranging from 44,709 in 1996 to 647,050 in 2102. After eliminating individuals with no labour income and the 0.5% of workers with the lowest and highest wages every year, there are 2,739,768 observations left. Note that this

[1] For instance, the innovative study of Greenstone, Hornbeck, and Moretti (2010) uses data from jurisdictions competing to host new large assembly plants. The ingenious design of the study circumvents serious identification concerns about unobserved local characteristics potentially correlated with employment changes. There is no hope of collecting such data for most countries. Nonetheless, there would be significant value replicating this study for a developing country but this represents a separate research project.

Table 1: Descriptive statistics

Panel a: Individual observations.	1996	2004	2012	all years	std. dev. (all years)
% male	62.67	57.13	55.30	56.80	49.53
Average age	35.55	37.71	39.03	38.06	13.31
Average years of schooling	7.52	8.50	9.40	8.88	4.71
Median municipal urban population	262,171	356,461	320,721	304,062	1,521,714
Panel b: 645 Municipalities.	p25	median	p75	mean	std. dev.
Total population (1993)	9,510	17,568	33,733	50,810	248,533
Total population (2005)	8,588	16,587	32,319	56,701	302,248
% pop. with postsecondary educ. (2005)	8.26	11.79	18.55	14.69	8.70
Market access index	30,941	123,036	280,977	201,870	239,181
Road length (km, 2012)	0	10.60	23.07	16.87	23.30
Distance to main road (km, 2012)	0	0	9.15	6.77	15.77

survey has been widely used by prior research, including Attanasio, Goldberg, and Pavcnik (2004) on the wage effects of trade reform or Angrist and Kugler (2008) on the effect of violence on wages, among many others. See panel a of table 1 for some descriptive statistics.

Several features of the geih are worth keeping in mind. Parts of the Colombian territory that were more recently administratively integrated with the rest of the country (the so called new departments) are not covered. This does not matter much for our purpose because these areas are extremely lightly populated and contain no significant city. In addition, this survey is not representative for smaller cities. Although this problem is likely to be much attenuated since I consider cities by size and not individually, I deal with it by either controlling for individual characteristics of workers or restricting our sample to the larger cities. Finally, the exact definition of labour income in the survey changed twice during the study period. There is a first change in 2000 in how hourly wages are computed and a second in 2005 when several other surveys were integrated into the National Household Survey to form the geih. These changes were hopefully minor and uncorrelated with the questions at hand. However, and because the number of surveyed households also greatly increased over time, the analysis is duplicated for pairs of years separately. See Departamento Administrativo Nacional de Estadística (2013) for further details about the geih.

This study complements the geih with a wealth of municipal-level information. The first source of information are the Colombian censuses of 1993 and 2005 for population by municipality with a breakdown between the rural and the urban part of each municipality. Colombia is composed of 1,119 municipalities. Colombian municipalities are rather large, with on average an area of 1,018 squared kilometres. This equivalent is to a disk with radius 18 kilometres. In most cases, municipalities are organised around one major settlement, referred to as the 'head' or urban part of the municipality. I complement the 1993 and 2005 population figures with DANE's municipal population estimates for 2010. The Colombian census also contains a wealth of socio-economic and education variables. Among them, the municipal share of the workforce with post secondary education is of primary interest.

To construct instruments for current population, I use population data collected from the 1843, 1870, and 1938 censuses. Although the first national census of Colombia dates back to 1812, the 1843 census, although probably subject to undercounting, is the first useable one (Bushnell, 1993). The details of the instrumentation strategy are postponed until later.

Other municipal characteristics are coming from several sources. Sánchez and Nuñez (2000) gathered a large number of geographical characteristics for Colombian municipalities including climate, availability of natural resources such as gold, coal, or oil, altitude, soils, or availability of water among others. This source is complemented with further municipal data collected by the National Planning Ministry (DNP) which cover land area, many socio-economic indicators, and some public finance variables. I also use a rich set of amenity and tourism variables collected by Meisel Roca and Pérez (2012) containing information about museums, public libraries, hotels, restaurants, and airport arrivals and departures.

I also use various measures of roads and market access for Colombian municipalities. Roads data are further described in Duranton (2013*a*) where I also compute market access variables. These data cover the Colombian road network from the country's colonial era to the current day network and various market access variables measured either directly using wages in Colombian municipalities and travel time distance from Google Map or indirectly using a large-scale internal trade survey. See below for further details. Panel B of table 1 provides some descriptive statistics for the Colombian municipalities represented in the study.

While most of the analysis here uses the population of the urban part of Colombian municipalities, in some regressions I use total municipal population. I also sometimes aggregate municipalities into metropolitan areas (MA) as in Duranton (2013*b*) where metropolitan areas are defined using inter-municipal commuting patterns.

3. Theory and estimation

The proposition that larger cities enjoy a productive advantage dates back to Smith (1776), who provided an insightful comparison between workers in Edinburgh and farmers in the Scottish highlands, and Marshall (1890), who argued that this productive advantage occurred because of more intensive input-output linkages, thicker local labour markets, and knowledge spillovers. Modern economic theory has provided a variety of microeconomic foundations for agglomeration economies (see Duranton and Puga, 2004, for a summary). Despite relying on different mechanisms, theories of agglomeration in production all predict that output increases more than proportionately with employment for a group of spatially proximate workers.

While there are different ways to define a given 'group of spatially proximate workers' such as the workers in narrowly defined sectors working within walking distance of each other, I will follow much of the recent literature and consider all workers in a city as defining the relevant group to measure agglomeration effects in my baseline analysis. To measure efficiency, I will also follow a dominant tradition in the literature and measure it through wages. The prediction is then that, because of agglomeration effects, wages should increase with city population.

More formally, I estimate variants of the following regression:

$$\log w_{ic(i)t} = \alpha \log Pop_{c(i)} + X_{it}\beta + A_{c(i)t}\gamma + \delta_t + \epsilon_{ic(i)t}\,, \tag{1}$$

where the dependent variable is the (natural) log of the hourly wage of person i who lives in city $c(i)$ in year t. The coefficient of interest is α, the elasticity of wages with respect to city population. The regression also includes a vector of individual characteristics X, a vector of city characteristics A, year indicators δ, and an error term ϵ.

Ideally one would like to compare the same workers across different, exogenously assigned locations. This is unfortunately impossible. Instead, I can only compare the wages of different workers who chose to be where they are.[2] There are two main sources of bias in the estimation of equation (1). They both find their source in the fact that our explanatory variable of interest $\log Pop_{c(i)}$ is indexed by $c(i)$, that is the city c is *chosen* by worker i.

The first source bias arises from the possibility that some city effects not included in the vector of city characteristics $A_{c(i)t}$ are correlated with city population and wages — a missing variable problem. Alternatively, workers may prefer to locate in cities where wages are higher — a reverse

[2] This discussion builds on Duranton (2014*a*). See also Combes and Gobillon (2014) for further details.

causation problem. Econometrically, the two problems are equivalent. I deal with them in two different manners. First, I use an extensive set of local control variables capturing the geography of Colombian municipalities, the education of their workforce, their infrastructure, their proximity to markets, their amenities, etc. Second, I implement a standard instrumental variable (IV) strategy and instrument current city population with long lags of population from 1843, 1870, and 1938 (Ciccone and Hall, 1996, Combes *et al.*, 2010). Given that past population is expressed in logs, this is equivalent to using past populations and long term population growth.

This IV strategy is valid if past populations are correlated with current wages only through current population. While this exclusion restriction is plausible, one can imagine a number of possible violations caused by alternative links between past population and current wages including, for instance, better market access, which may be extremely persistent, or some other geographical characteristics. These variables could affect wages and population at the same time and thus defeat the purpose of the instruments. To minimise potential problems, I consider a wide variety of possible control variables to preclude such correlations. An alternative is to use other instruments. In the spirit of Combes *et al.* (2010), I use soil characteristics, namely soil erodibility and fertility, which predict current population but are not obviously linked to wages in the largest settlements where agriculture only plays a minor role.

The second main identification problem associated with the estimation of equation (1) is a possible correlation between the measure of city population and unobserved individual characteristics. More productive workers may stand to gain more from working in cities. This correlation between individual effects and city population may bias the estimate of the population elasticity α in equation (1). A possible strategy is to rely on a panel of workers and estimate α by comparing the same workers across several locations (e.g., Combes *et al.*, 2010). This strategy is not available here in absence of household identifiers. An alternative solution is to control for an extensive set of individual characteristics as in Bacolod, Blum, and Strange (2009) or Glaeser and Resseger (2010).

Much of the recent literature has been concerned with the heterogeneity of workers in their ability to generate and benefit from agglomeration effects. I explore many of these possible heterogeneous effects of agglomeration here.

The first form of heterogeneity regards the possibility of non-linear agglomeration effects. The literature has often focused on estimating a constant elasticity of wages (or any other measure of local efficiency) with respect to city population. A constant elasticity is often predicted by simple models of agglomeration economies (Duranton and Puga, 2004) but this prediction is mainly driven by the desire of theorists to rely on simple functional forms and keep their model tractable. Agglomeration benefits that differ with city size are interesting in themselves. They are also of importance for policy. As noted by Glaeser and Gottlieb (2008) and Moretti (2011), even though agglomeration effects rely on market failures, the distribution of population across cities need not be inefficient. With a constant agglomeration elasticity, relocating workers across cities is neutral from an efficiency perspective since agglomeration gains in the city of arrival are fully offset by agglomeration losses in the city of departure.

There may also be some heterogeneity in agglomeration effects over time. The first half of the study period starting in 1996 was a time of considerable civil unrest in many parts of Colombia. Although armed groups were mostly present in rural areas, smaller cities were also greatly affected. In the second part of the study period after 2004, levels of civil unrest were much lower. This may have had a powerful effect on Colombian cities. More generally, as countries develop agglomeration benefits may change as well. It may also be the case that agglomeration benefits are altered by technological progress which may alleviate or reinforce the need to co-locate and the benefits from such co-location. I assess time changes in agglomeration effects in what follows.

Different groups of workers may also benefit from agglomeration differently. There is a long standing conjecture that cities and skills may complement each other. There is some evidence of such effects being at play in the US (Bacolod *et al.*, 2009, Glaeser and Resseger, 2010). Whether a positive interaction between city population and a worker's education is also at work in Colombia is an open question. Because labour markets for male and female workers often differ considerably, particularly in a country like Colombia, urbanization may have different effects on male and female workers and more generally affect workers in different demographic groups differently. An analysis of individual heterogeneity in agglomeration effects is proposed below.

Another question of specific interest here is the possible difference in agglomeration effects between young and old workers. There is some evidence (e.g., Peri, 2002, Wheeler, 2006) that wage growth is stronger in larger cities in the US. There is also evidence that the length of tenure in large cities is also closely related to the ability of workers to reap the benefits of agglomeration (De la Roca and Puga, 2012). These outcomes are consistent with theories that emphasise greater learning in larger cities (Glaeser, 1999). Although the absence of a panel of worker prevents a direct treatment of the issue of urban learning, note that these theories predict higher benefits from cities for workers that have lived there longer, that is older workers. Hence the issue of urban learning can be tackled indirectly.

There may be some heterogeneity, not only in the ability of workers to benefit from agglomeration, but also in their propensity to generate such effects. Following Rauch (1993) and Moretti (2004), a significant body of literature emphasises that agglomeration benefits may be generated mostly by skilled workers. Although the literature has focused on the correlation between average education (and more specifically the share of workers with higher education) and city population, it is still unclear whether a measure of city education should be viewed as an alternative to city population to estimate the same agglomeration effects or as a way to capture other agglomeration effects. The analysis below will hopefully shed lights on these issues.

Going back to potential confounding factors that might be correlated with city size and explain wages, three of them are of particular relevance and interest: market access, the local road infrastructure, and amenities. Starting with market access, it is possible to view it as a form of agglomeration effects taking place at a broader spatial scale. In the original specification of Krugman (1980), better market access allows firms to export more and at a lower costs. Put differently, better market access implies a stronger demand for the output of local firms and this increases the value of the marginal product of their labour and can thus be expected to lead to an increase in local wages. In a refinement of this idea, Krugman and Venables (1995) consider the importance of market access in a model where they add input-output linkages across locations. Better access to outside markets makes local firms more productive by allowing them to access a broader variety of goods at a cheaper price. In turn, this makes local firms more productive and, again, should lead to higher wages. Following Redding and Venables (2004) there has been a lot of work highlighting the importance of market access including in particular the assessment of Fally, Paillacar, and Terra (2010) for Brazilian regions.

Adding to these two effects that increase the nominal wage, better market access also increases the real wage by allowing local consumers to consume a broader variety of goods at a lower price. Handbury and Weinstein (2013) provide strong evidence that larger cities (with perhaps better market access) provide a larger variety of grocery goods at a lower price.

The immobility of labour is an important assumption imposed by Krugman (1980) or Krugman and Venables (1995) in their models and much of the follow-up work. In the context of cities within a country, we need to allow for some labour mobility. Interestingly, labour mobility leads to very different predictions regarding the effects of market access. This is because labour mobility implies the equalisation of utility, often measured by real wages, in simple models. Then, if the wage divided by the local price index is equal across cities, cities with better market access (which enjoy a lower price index) must also offer lower nominal wages.[3] I explore these issues further in section 5.2.

There is also a recent literature interested in the effects of urban infrastructure, following in particular Baum-Snow (2007). This literature has looked at the effect of transportation infrastructure on a wide variety of outcomes including suburbanisation, the population growth of cities, or urban export and import patterns (see Redding and Turner, 2014, for a recent survey). Although some work has been devoted to the effects of transportation on rural wages (Michaels, 2008, Donaldson and Hornbeck, 2013), very little attention has been devoted to urban wages. In light of the discussion above, the expected effect of the road infrastructure (or urban wages) is ambiguous. More roads in a city may improve market access and thus have a negative effect on wages. The local transportation infrastructure may also act as an argument in the city production function and have a positive effect on wages. Finally, roads may affect the attractiveness of cities and may thus indirectly affect wages at the spatial equilibrium.

Related to this last point, we should expect any amenity to be potentially reflected in wages through, again, a spatial equilibrium mechanism (Rosen, 1979, Roback, 1982). Any city with more amenities will, at the spatial equilibrium, be a more expensive place to live and offer lower wages. Higher housing prices and lower wages are the two main margins that we expect to be at work to equate utility across cities (Glaeser and Gottlieb, 2009).

4. Results about city population

4.1 Baseline results

Table 2 reports the main OLS results regarding the elasticity of wages with respect to city population. In absence of any other control, the specification in column 1 reports an elasticity of 0.11 (or 11%). This indicates that in the cross-section of Colombian cities between 1996 and 2012, a 10% larger population is associated with 1.1% higher wages. Column 2 adds individual characteristics

[3] Although in absence of labour mobility, better market access raises wages as argued above, labour mobility, by increasing labour supply reduces wages. With free labour mobility, utility equalisation must thus imply lower nominal wages.

Table 2: Agglomeration effects, OLS specifications for all workers

	(1) Only pop.	(2) Indiv.	(3) Geog.	(4) Access	(5) Muni. pop.	(6) MA	(7) Non-lin.	(8) Comp.
log population	0.11^a	0.054^a	0.060^a	0.055^a	0.060^a	0.046^a	0.071	0.054^a
	(0.0051)	(0.0045)	(0.013)	(0.0062)	(0.0053)	(0.0070)	(0.045)	(0.0046)
log population^2							-0.00072	
							(0.0020)	
log market access				-0.13^a				
				(0.037)				
log market access^2				0.0068^a				
				(0.0022)				
Educ. $\times$ log pop.								-0.0086^c
								(0.0045)
Observations	2,738,494	2,735,518	2,712,456	1,873,803	2,735,518	2,735,518	2,735,518	2,735,518
R^2	0.10	0.40	0.41	0.41	0.40	0.36	0.40	0.40
Municipalities	645	645	618	337	645	572	645	645

Notes: OLS regressions with year indicators in all columns. Robust standard errors clustered at the municipality level in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. The dependent variable is a log wage in all columns. Individual characteristics included in columns 2 to 8 are: gender, 26 education indicators, and imputed experience and its square. Geographic characteristics in columns 3 include three regional indicators (Andean, Carribean, and Oriental), the log of the urban area, two climate indicators (hot and moderate), a series of indicators variables for natural resources (coal, gold, and oil), the log of natural water flows, a precipitation index, and the log of altitude and its square. For a given municipality, market access is measured by multiplying 2005 population by average municipal wage for all trading partners with population above 40,000, dividing by Google map time distance, and summing. The log of market access and its square are included in column 4. Its exact computation is explained in section 5.2. Column 5 replicates column 2 using total municipal population. Column 6 uses metropolitan areas instead of municipalities.

as control variables. This divides the coefficient on log city population by two at 0.054. This implies that about half the relationship between wages and city population is explained by larger cities hosting more educated workers. This is consistent with the greater representation of more educated workers in larger cities. Regressing the share of workers with postsecondary education on city population in 2010 yields a highly significant coefficient of 0.05 with a R^2 of 47%.[4]

While the rest of this section is devoted to asserting the robustness of this finding and examining

[4] This coefficient of 0.05 may seem small but only small differences are needed here given the large effects of individual education on wages. For instance, workers with 18 years of education have an hourly wage more than six times as high relative to workers with only six years of education. Small differences in the education composition of the workforce across cities are enough to explain away about half the raw elasticity of wages with respect to city population of 0.11. Behrens, Duranton, and Robert-Nicoud (2014) estimate similar relationships for US cities. Using aggregate data, they estimate an elasticity of earnings with respect to city population of 0.082 without conditioning for the share of population with a post-secondary education degree and 0.051 after conditioning for this. They also estimate an elasticity of the share of educated workers with respect to city population of about 0.068. Hence the drop in the elasticity of city earnings with respect to population drops by a lower amount in the US when controlling for city population despite a slightly faster increase in the number of college educated workers as city size increases. This is to be expected given the much high college premium in Colombia and the fact that Behrens *et al.* (2014) use aggregate data.

the possible heterogeneity that underlies it, a few comments are in order at this stage. First, a city population elasticity of wages of 5.4% is quantitatively important. Comparing a small municipality with an urban population of, say, 5,000 to Bogota with a population above seven million, this elasticity implies a wage difference of nearly 50%. While city population is not the sole factor accounting for spatial wage disparities in Colombia, it is certainly a prime driver.

Second, this elasticity of 5.4% is somewhat larger than comparable findings for cities in more developed countries. Using a similar specification, Glaeser and Resseger (2010) report an elasticity of 3.4% while, Combes *et al.* (2010) report an elasticity of 3.3% using a more data intensive approach. On the other hand, Combes *et al.* (2013) for Chinese cities and Chauvin *et al.* (2013) for Indian cities report much higher elasticities of 10 to 20%. Third, the drop in the coefficient on city population when controlling for individual characteristics is not unusual. As in most countries where this type of approach has been implemented, workforce composition effects account for a sizeable fraction of spatial wage disparities (Combes and Gobillon, 2014).

Column 3 further includes a number of geographic variables including regional indicators, land area, climate and precipitation, indicators for natural resources, water flows, and altitude. The coefficient on city population for this specification is marginally higher. Note also that adding geographic controls does not increase the explanatory power of the regression by much. Although detailed results are not reported here, the coefficients on most geographic controls are not statistically significant. Wages are higher in the Andean and Oriental regions of Colombia relative to the Caribbean and Pacific regions. In this regression, climate and the presence of oil also have a significant coefficient but this relationship is not robust.

Relative to the specification of column 2, column 4 adds log market access and its square. In this regression market access is computed by summing the income of other municipalities discounted by the distance to the municipality under consideration. Income is measured using the same wage data and distance is measured in travel hours since in a mountainous country like Colombia the Euclidian distance between two places is sometimes a poor indicator of the true travel cost between these places. Note also that the summation is performed across all trade partners as measured in the data used in Duranton (2013*a*). These typically do not include nearby municipalities. Interestingly the coefficient on log market access is negative. The coefficient on the squared market access term is positive but no municipality has a high enough market access for the overall effect to be positive. For now, only note that introducing market access does not affect the coefficient on city population. The effect of market access is subject to greater scrutiny below.

Column 5 duplicates column 2 but uses total municipal population instead of the population of the urban part of a municipality. The coefficient on log population increases marginally to 0.06. Column 6 duplicates column 2 again but uses metropolitan areas instead of municipalities. The coefficient on city population drops slightly to 0.046. This drop is small. It is caused by the fact that for small areas, the municipality and the 'metropolitan area' coincide whereas large metropolitan areas are formed of several municipalities. As a result, the dispersion of the dependent variable, population, increases whereas the dispersion of the main explanatory variable, mean wages, is marginally reduced. This mechanically reduces the coefficient on population. Reassuringly the effect remains small. Interestingly, the R^2 of teh regression also declines somewhat in this specification. This is perhaps a weak indication that it is better to work with municipalities than with metropolitan areas in the case of Colombia.

4.2 Heterogeneity in agglomeration effects

To start investigating possible heterogeneities in agglomeration effects, column 7 of table 2 attempts to detect non-linearities by adding the square of log city population as dependent variable to the specification of column 2. The point estimate for the coefficient on log city population is marginally higher at 0.07 but insignificant while the coefficient on the quadratic term is insignificant and close to zero. This absence of non-linearities in agglomeration effects in Colombia is confirmed by table 10 in appendix. This table duplicates table 2 but considers only individuals living in municipalities with more than 20,000 inhabitants in 1993. This divides the number of municipalities by more than two. Despite this much lower number of cities, the coefficients on log city population reported in table 10 are very close to those of table 2.

Finally, column 8 adds an interaction term to the specification of column 2: the product of the worker's number of year of schooling by log city population. The coefficient on this interaction term is small, negative, and marginally significant. This finding of higher agglomeration returns to less educated workers is not robust. Adding further controls typically makes the coefficient on the interaction term insignificant. This negative or insignificant interaction effect is nonetheless in contrast with results from extant literature which typically highlight the existence of higher returns to cities for more educated workers (see Wheeler, 2001, Bacolod *et al.*, 2009, Glaeser and Resseger, 2010, among others). A complementarity between city population and individual skills is also often alleged to be an important factor behind the over-representation of more skilled workers in larger cities (Behrens *et al.*, 2014). This over-representation occurs in most countries, including Colombia as argued in footnote 4. There are two explanations for the more than proportional presence of more educated workers in larger Colombian cities. The first is that more educated workers are more sensitive to the urban amenities provided by large cities. The second is that highly educated workers in Colombia are not mobile and more likely to originate from larger cities. Unfortunately, there is no strong evidence in the data in favour of either of these two explanations.

To pursue this investigation on possible heterogenous effects of agglomeration on different types of workers, table 11 in the appendix duplicates table 2 for prime-aged males only. The coefficients on log city population are similar in both tables. They never differ by more than 0.005 and are never statistically different across the two tables for the same specification. Table 12 replicates the same exercise for prime-aged females. Comparing the coefficients on log city size across specifications in tables 11 and 12 shows very little differences in the benefits from agglomeration accruing to men and women and there is no tendency for one gender to experience systematically higher coefficients. The coefficients are typically within 0.005 of each other, which corresponds to a standard deviation or less. Hence, despite considerable differences in both the supply and demand for labour between males and females, both genders appear to benefit equally from agglomeration effects.

Next, I turn to the comparison of workers of different age groups. Table 13 duplicates table 2 for male workers aged 20 to 30 whereas table 14 does the same thing for workers aged 40 to 50. The differences between the two tables and with table 2 are small. If anything, younger workers appear to benefit more from agglomeration with a coefficient that is between 0.005 and 0.02 higher. This difference is sometimes statistically significant at 5%.

This result is somewhat surprising and at odds with a lot of extant literature. It is often assumed that workers learn more in larger cities (Glaeser, 1999, Duranton and Puga, 2004). Indirect evidence of this has been proposed by Glaeser and Maré (2001) and Combes, Duranton, Gobillon, and Roux (2012) among others. They show that workers retain some benefits from agglomeration after they leave their city. In a slightly different spirit, Peri (2002) and Wheeler (2006), among others, shows that wage growth is stronger for younger workers in larger cities. In the most sophisticated treatment of the issue to-date, De la Roca and Puga (2012) show that workers benefit more from agglomeration as their tenure in a city gets longer. The learning in cities argument typically implies that agglomeration effects should be stronger in larger cities for older workers.[5] This is clearly not the case in Colombia.

Agglomeration effects might also be heterogeneous over time. Many popular discussions underscore either beliefs in how agglomeration effects are becoming irrelevant in an era of cheap telecommunications or, instead, how proximity increasingly matters. While a full analysis of this issue is clearly beyond the scope of this paper, it is interesting to look for any possible time heterogeneity behind the results reported so far. Table 15 in appendix duplicates the specification in table 2 for eight separate time periods, namely pairs of consecutive years since 1996 except for 2012 which is considered alone in the last column. Some caution is needed in interpreting these results since the sampling frame of the data and how wages are defined have varied since 1996. Overall the changes in the coefficient on city population are limited and coincide with changes in data collection. The estimated elasticity of wages with respect to city size is about 0.06-0.07 during the 1996-1999 period, it is then lower around 0.04-0.05 during the 2000-2003 period. They are finally extremely stable around 0.055 since 2004. While the changes in the GEIH may well explain these minor differences, it could also be the case that benefits from agglomeration became lower early in the century, at a time of considerable migration to some cities caused by civil unrest in some parts of the country.

4.3 Further identification of agglomeration effects

Now that the stability of agglomeration effects over time is established, it is tempting to attempt to identify the elasticity of city wages with respect to city population using city population and wage changes over time. In practice, this implies estimating equation (1) using municipality fixed effects. To rely solely on cross-sectional variation and avoid using any time variation, the results reported in table (2) and all its variants imputed the same population to all municipalities for all years using an average of 1993, 2005, and 2010 population data. To estimate equation (1) using only the time variation in the data, a different approach is needed. Urban municipal population for 1996 to 2004 can be created by extrapolation using 1993 and 2005 population data. Population for 2008, 2009, 2011, and 2012 can also be imputed using a similar approach.

[5]Unless older urban workers massively move to rural areas and are replaced by old workers from rural areas, in which case older workers in cities would be less skilled. This situation is deeply implausible.

Because of these imputations for the key explanatory variable, city population, estimating equation (1) with city fixed effects suffers from a major measurement problem. The other identification issue that afflicts this estimation is that variations in population are likely to be sensitive to changes in wages. More specifically, in a fixed-effect or first-difference setting, population is likely to be 'more endogenous' than in a pure cross-sectional estimation.

This said, adding municipal fixed effects to the baseline specification of table 2 column 2 estimates an insignificant coefficient of 0.04 for city population. (Results are not reported here.) Using only the years for which population is better measured and using overall municipal population, which is perhaps better measured than population of the urban part, implies a coefficient of 0.081 significant at 1%. Many other specifications estimate coefficients close to zero with large standard errors. After extensive experimentations, my two main conclusions are that fixed effect specifications do not contradict cross-sectional estimations but that they do not provide enough precision to be really useful.

Table 3: Agglomeration effects, TSLS specifications

	(1) Only pop.	(2) Indiv.	(3) Geog.	(4) Access	(5) Muni. pop.	(6) MA	(7) Non-lin.	(8) Comp.
log population	0.092^a	0.048^a	0.047^a	0.056^a	0.053^a	0.042^a	-0.12	0.077^a
	(0.011)	(0.011)	(0.029)	(0.012)	(0.012)	(0.014)	(0.10)	(0.014)
log population^2							0.0068^c	
							(0.0039)	
log market access				-0.12^a				
				(0.044)				
log market access^2				0.0061^b				
				(0.0026)				
Educ. $\times$ log pop.								-0.71
								(0.55)
Overid. p-value	0.38	0.25	0	0.050	0.25	0.33	0.39	0.27
First-stage Stat.	101	97.2	108	71.2	106	132	14.1	8.35
Obs.	2,029,310	2,026,723	2,015,610	1,444,433	2,026,723	2,063,528	2,026,723	2,026,723
Municipalities	316	316	311	165	316	264	316	316

Notes: This table duplicates table 2 using two-stage least-square (TSLS). In all columns, log population is instrumented by the corresponding log populations for 1843, 1870, 1938. The square of these instruments are also used in column 6. The log population term interacted with education in column 8 is a projection from log population in 1843, 1870, and 1938. Unlike in table 2, the set of geography controls in column 3 does not include the log of land area which makes the instruments weak since thinly populated places in the past led to the creation of large municipalities.

To alleviate worries about identification effects, I now turn to TSLS estimations and instrument contemporaneous populations with corresponding 1843, 1870, and 1938 populations. As argued above, these three instruments will avoid possible biases caused by reverse causality. The validity of the instruments rely on these three past populations being uncorrelated with unobserved drivers of contemporaneous wages conditional on the other controls. While a good case can be made for this (Combes *et al.*, 2010), it is by no means totally decisive. This type of IV estimation is probably best thought of as further evidence supportive of the main result and part of a series of robustness checks rather than a decisive approach to estimation.

The first series of IV results is reported in table 3. This table replicates table 2 instrumenting contemporaneous population by the three long historical lags described in the previous paragraph. In column 1, the coefficient on log city population is marginally lower than its OLS counterpart. Given the absence of individual characteristics, the informational content of this regression is probably limited. In column 2, the TSLS elasticity of wages with respect to city population is 0.048 instead of 0.054 for its OLS counterpart. This difference is small, less than one standard deviation, and insignificant. The ancillary tests show that the instruments are strong and imply similar coefficients. Although the OLS-TSLS difference obtained with this specification is not significant, it appears again with the same sign and magnitude in most of the specifications discussed below. This is suggestive of perhaps a small upward bias in the OLS coefficients of table 2. Note also that estimating a marginally lower coefficient for city population with this sort of instrument is a standard finding for this type of regression (e.g. Combes *et al.*, 2010).

Turning to the subsequent columns of table 3, note that introducing a large number of geographic controls makes no change to the coefficient on city population in column 3. Columns 4 and 5 estimate slightly higher coefficients whereas, in keeping with OLS results, column 6 estimates a slighly lower coefficient. Column 7 estimates a negative coefficient for city population. The coefficient on squared log population is however higher than its OLS counterpart and significant. These two estimates imply a marginal agglomeration effect of 0.048 (as in column 2) for a city of 240,000. The coefficient on city population in column 8 is also close to its corresponding OLS coefficient.

The main conclusion to be drawn from table 3 is the marginally lower agglomeration elasticity with TSLS. To assess the robustness of this result further, table 4 experiments with the instrumentation strategy and the estimation technique. Using the same specification as column 2 of

Table 4: Agglomeration effects, robustness to the choice of instruments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log population	0.050[a]	0.050[a]	0.050[a]	0.045	0.057[a]	0.048[a]	0.049[a]	0.048[a]
	(0.017)	(0.0090)	(0.0069)	(0.030)	(0.0078)	(0.011)	(0.011)	(0.011)
Instruments:								
1843 pop.	Y	N	N	N	Y	Y	Y	Y
1870 pop.	N	N	N	N	N	N	Y	Y
1938 pop.	N	Y	N	N	Y	Y	Y	Y
Tourism	N	N	Y	N	Y	N	N	N
Soil	N	N	N	Y	Y	N	N	N
Overid. p-value	.	.	0.100	0.0054	0.013	0.90	0.25	0.25
First-stage Stat.	13.3	193	130	1.45	198	144	97.2	97.2
Obs.	2,049,671	2,565,796	2,735,518	2,723,582	2,029,698	2,040,811	2,026,723	2,026,723
Municipalities	349	490	645	637	325	330	316	316

Notes: Regressions with year indicators and individual characteristics in all columns (TSLS in columns 1-6, GMM in column 7, and LIML in column 8). Tourism instruments are the log of one plus the number of air passenger arrivals in 1998 and 2009 and the log of one plus the number of hotels in 2005. Soil instruments are soil erosion and a measure of soil fertility (called 'aptitude'). The dependent variable is a log wage in all columns.

table 2, column 1 of table 4 instruments contemporaneous population with only 1848 population while column 2 uses only 1938 population. The coefficient on population is exactly the same in both columns. Column 3 instruments population with the number of air passenger arrival and the number of 2005 hotels. While it is obvious that larger cities will have a larger airport and more hotels, hotels are also more prevalent in nicer places that may have historically attracted a greater population. This still estimates the same coefficient. Column 4 uses two soil characteristics in the spirit of Combes *et al.* (2010). These two soils instruments are unfortunately weak. As a result, the coefficient on city population is insignificant but remains close to previous estimates. Similar estimates are obtained again in columns 5 and 6 using either all instruments together or only 1843 and 1938 populations. Finally columns 7 and 8 use 1843, 1870, and 1938 populations as in table 3 but estimate the regression with the generalised method of moments (IV-GMM) and with limited information maximum likelihood (LIML) instead of TSLS. The results are again the same using these two alternative estimation techniques. Overall, it appears that the instrumented coefficients on city population are generally statistically similar to our baseline OLS of 0.055 but with a marginally lower point estimate. This conclusion is generally robust to the choices of specification, instruments, and estimation technique.

5. Results about human capital externalities, market access, and infrastructure and amenities

I now extend the previous approach and focus on four sets of variables: city human capital, market access, infrastructure, and amenities. The results that follow should be viewed as providing robustness checks for our main results so far. It is also an exploration into alternative ways to measure and think about agglomeration effects.

5.1 Human capital externalities

Human capital externalities arise from the notion that having more education or skills in a city generates social benefits to urban workers over and beyond the already large private benefits of education or skills. Hence, there are benefits in being surrounded by more educated or skilled workers. Put differently, the notion of human capital externalities recognises that perhaps not all workers contribute to agglomeration effects equally. In particular, more skilled or educated workers may contribute more. On the other hand, the agglomeration effects associated with city population implicitly assume that all workers contribute equally to agglomeration benefits. The key distinction is thus that human capital externalities differ from the agglomeration effects measured so far as they concern the skill composition of the local workforce, not its absolute size.

Before going any further, it is important to recognise a key tension between the 'agglomeration' literature and the 'human capital externality' literature. The key explanatory variable of interest (city size or share/number of educated/skilled worker) of each literature is typically ignored (missing) in the other.[6] To reconcile these two strands of literature, I include both a measure of city scale (city population) just like in previous regressions and some measure of the share of more educated workers. This allows me to account for the skill composition of the workforce at the same time as its scale. To measure city skill/education, I use the share of workers with some university education since the importance of education at the upper end is underscored by extant results in the literature.

[6]The pioneering efforts in the estimation of human capital externalities are due to Rauch (1993) and Moretti (2004). Their findings of strong external returns to education have been reproduced in many countries. See Duranton (2014*a*) for a short review and discussion of the issue for developing countries.

Table 5: Agglomeration and human capital externalities, OLS

	(1) Only pop.	(2) Indiv.	(3) Geog.	(4) Access	(5) Muni. pop.	(6) MA	(7) Non-lin.	(8) Comp.
log population	0.057^a	0.052^a	0.055^a	0.060^a	0.047^a	0.035^a	0.068	0.053^a
	(0.011)	(0.0089)	(0.013)	(0.014)	(0.0096)	(0.010)	(0.049)	(0.0089)
share educated	1.43^a	0.050	0.47^a	-0.14	0.30	0.35	0.036	0.047
	(0.33)	(0.26)	(0.12)	(0.35)	(0.23)	(0.22)	(0.27)	(0.26)
log population^2							-0.00066	
							(0.0020)	
log market access				-0.13^a				
				(0.042)				
log market access^2				0.0070^a				
				(0.0025)				
Educ × log pop.								-0.0082^c
								(0.0043)
Observations	2,737,326	2,734,355	2,711,293	1,872,901	2,734,355	2,734,355	2,734,355	2,734,355
R^2	0.10	0.40	0.41	0.41	0.40	0.36	0.40	0.40
Municipalities	643	643	616	336	643	570	643	643

Notes: This table duplicates table 2 adding the share of population with some university education.

Table 5 duplicates table 2 adding the share of inhabitants with some university education, which can be loosely referred to as city education. Recall that individual education is controlled for so that the coefficient on city population really measures an external effect of education.[7]

Two results emerge. First, the coefficient on city population is unaffected by the introduction of this alternative measure of agglomeration. Its significance and its magnitude are the same as before. This reinforces all the conclusions drawn above. Second, the results regarding human capital externalities are mixed. The coefficient on city education is significant only in columns 1 and 3. In column 3, this coefficient is large but of the same magnitude as uncovered by Moretti (2004) in the US. In columns 5 and 6, the magnitude remains the same but the coefficient is insignificant. In other columns, the coefficient on city education is much smaller (if not negative) and insignificant.

When the same regressions as in table 5 are estimated without city population, the coefficient on city education is high and highly significant. However in the horse race between city population and city education, city population emerges as a clear winner. Since both population and education are measured from the same source (the Colombian censuses of 1993 and 2005 and more recent

[7]There are of course serious identification worries. The main one is that more educated workers flock to cities with higher wages. As will become clear in the next paragraph, this worry is moot in the Colombian case.

population projections), it is hard to argue that this result is caused by a measurement issue where the effect of city education would be captured by population.

To verify that these results were not driven by any of the choices made when estimating the regressions in table 5, I made a number of checks. Maybe the share of educated workers in city population is not used with an appropriate functional form. Table 16 in appendix duplicates the regression of table 5 using the log share of university educated workers instead of the level. The results are the same. Maybe the explanatory variable is not appropriate. I experimented with various alternative measures of post-secondary education going from the more restrictive (university degree) to the least restrictive (any form of post-secondary education) and obtained similar results. Using average years of schooling does not give stronger results either. Maybe using cross-sectional variation over a long period of time is not appropriate. Using only data for 2008-2010 with 2010 education data does not improve the results. Maybe using data for a large number of municipalities is not appropriate as human capital externalities may occur only in larger cities. Using municipalities with a population above 20,000 in 1993, which drastically reduces the number of municipalities as shown by table 10 does not yield stronger results. From this, I conclude that the evidence on human capital externalities in Colombia is mixed.

5.2 *Market access*

As argued above, the effects of market access on wages are potentially ambiguous and need not be positive as often believed. To understand the reasons for this possible ambiguity, consider the following heuristic. Indirect utility in municipality i is given by $(w_i - H_i)/P_i$ where H_i is expenditure on housing and local transportation and P_i is the local price index for consumption goods. In this simple example, indirect utility is given by real income once the cost of local housing and transportation has been paid. Better market access can have four effects. The first two are higher demand for the output from local firms (because buyers can buy at a lower price inclusive of transportation costs) and higher productivity (because the firm can purchase intermediate inputs at a lower price). These two effects are expected to translate into higher wages. On the other hand, better market access also implies lower prices for consumption goods. Finally, better market access may lead to population growth and thus higher local housing and transportation costs. At a simple spatial equilibrium with labour mobility, indirect utility $(w_i - H_i)/P_i$ is equalised across municipalities. As a result, better market access can imply higher or lower wages depending on which effects dominates and what the main adjustment margin is. In particular, if better market access primarily means a lower price index locally, wages will need to adjust downwards for indirect utility to be equalised across municipalities.

Table 6: Agglomeration effects, OLS specifications for workers in large municipalities

	'Theory' market access			Income market access			Population market access		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
log population	0.044[a]	0.040[a]	0.059[a]	0.055[a]	0.060[a]	0.071[a]	0.054[a]	0.052[a]	0.052[a]
	(0.0064)	(0.0087)	(0.013)	(0.0062)	(0.014)	(0.020)	(0.0057)	(0.0092)	(0.014)
log market access	-0.079[b]	-0.080[a]	-0.016	-0.13[a]	-0.13[a]	-0.010	-0.11	-0.10	-0.0067
	(0.031)	(0.030)	(0.018)	(0.037)	(0.042)	(0.025)	(0.096)	(0.099)	(0.046)
log market access^2	-0.028[a]	-0.027[b]	-0.019[a]	0.0068[a]	0.0070[a]	0.00074	0.0041	0.0040	-0.000030
	(0.011)	(0.011)	(0.0066)	(0.0022)	(0.0025)	(0.0014)	(0.0040)	(0.0041)	(0.0019)
share educ .	N	Y	Y	N	Y	Y	N	Y	Y
Geography	N	N	Y	N	N	Y	N	N	Y
Observations	2,608,300	2,607,137	2,586,549	1,873,803	1,872,901	1,870,884	2,573,455	2,573,455	2,559,814
R^2	0.40	0.40	0.41	0.41	0.41	0.41	0.40	0.40	0.41
Municipalities	630	628	603	337	336	329	470	470	457

Notes: OLS regressions with year indicators and individual characteristics in all columns. Robust standard errors clustered at the municipality level in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. The dependent variable is a log wage in all columns. Geographic characteristics are the same as in table 2. Market access in columns 1 to 3 is computed as in Duranton, Morrow, and Turner (2014) using the fixed effects estimated from a trade gravity regression. Market access in columns 4 to 6 is computed for municipality j as $\sum_j Pop_j \overline{w}_j / d_{ij}$ where the j are all the trade partners of municipality j identified in the shipment data used in Duranton (2013*a*), Pop_j is population of municipality j, $\overline{w}_j$ is the wage of municipality j net of individual characteristics estimated as above, and d_{ij} is road distance between i and j. Market access in columns 7 to 9 is computed for municipality j as $\sum_j Pop_j / d_{ij}$.

Table 6 reports results for three series of regressions with three different measures of market access. Columns 1 to 3 use a theory-based measure of market access as computed in Duranton *et al.* (2014), itself a simplified version of the one proposed by Redding and Venables (2004). Column 1 adds this variable to the baseline specification of column 2 in table 2. Column 2 also adds the log share of educated workers. Column 3 further includes geographic controls and market access. The same patterns is repeated in columns 4 to 6 with a measure of market access based on the aggregate incomes of a municipality's trade partners. Unlike the measure of market access used in columns 1 to 3, this one is constructed directly from local incomes. As made clear below, it is less subject to sampling errors. It main drawback is that it does not account for local prices unlike the theory-driven measure of market access estimated from trade data used in columns 1 to 3. Hence, this alternative measure of market access should be thought as an *ad hoc* counterpart of the previous one. Finally, columns 7 to 9 use an even simpler measure of market access where only population is considered.

In columns 1 and 2, the coefficient on market access and on its square are both negative and significant. In column 3, both coefficients remain negative but only that on squared market access is significant.The more ad-hoc market access variable used in columns 4 to 6 that weights population by a measure of income has negative and significant coefficients in two of the three specifications. The positive coefficient on the squared term never makes the marginal effect of market access positive over the range of observed market access. The simpler market access variable used in columns 7 to 9 is always insignificant. Note that this is due to larger standard errors. The point estimates are generally consistent with those obtained with the other two measures of market access.

It is important to note that market access is computed over a municipality's trading partners. The shipment data from 2011 used to identify a municipality's trading partners were collected by verifying the content of travelling trucks along Colombia's major roads. This implies both sampling errors and systematic errors since trade between nearby municipalities taking place on local roads is not recorded.[8] In practice, trading partners are cities with a population above 40,000 that are in most cases not neighbour municipalities. For instance, the shipment data used to compute market access does not record any trade between the capital city Bogota and its immediate suburbs. As explained below, this selection on trade partners that are not neighbours may be an advantage.

With this in mind, the next step is to compare the results of table 6 with more standard measures where market access is computed using all other municipalities and not only trade partners. Table 17 in appendix reports variants of column 1 of table 6 using alternative measures of market access for each municipality based on population, local labour incomes, or local GDP across all other municipalities, not just trade partners. In all cases, the coefficient on market access is insignificant and poorly estimated. I experimented with a variety of other specifications and obtained similar outcomes.

Another alternative is to compute the market access of a municipality including the municipality itself in the computation. Using the aggregate income of a municipality when computing its market access raises an obvious endogeneity problem. Because municipalities are weighted

[8] The primary objective of the data is to collect information about traffic along major roads in Colombia, not to measure trade across places.

by inverse distance in the computation of market access, own income will often play a dominant role in the computation of a municipality's market access. When regressing log wages on log market access, we come close to regressing these individual wages on average city wages and city population. Unsurprisingly, the coefficient on market access in this case is positive and significant. To avoid regressing a wage on a wage, it is probably better to compute market access using population only. The drawback here is that a population only market access is highly correlated with own city population which also appears in the regression. Dropping local population from the regression usually leads to strong positive coefficients for the market access term. Again, this is unsurprising since a market access term which include a municipality's own population captures the variation in wages previously captured by the local population variable.

Although subject to some caveats and of limited robustness, the market access results of table 6 are in stark contrast with results in extant literature where a positive effect of market access on wages is usually found. The most advanced work on the issue is by Fally *et al.* (2010) who assess the importance of market access for Brazilian regions. Their approach shares some similarities with what is done here. In a first regression, they compute a local wage conditioning out the effect of individual characteristics very much like here. In a second regression, they regress bilateral trade between Brazilian jurisdictions on distance and fixed effects for importers and exporters. Although details differ, this parallels what is done in Duranton (2013*a*) and used here to compute market access. Finally they regress local wages on market access and other controls in the same way as in table 6. They find evidence of a robust positive effect of market access. This is of course an obvious contrast with the results of table 6.

The coexistence of negative, insignificant, and positive results may seem contradictory. Perhaps it is not. First, the positive results of Fally *et al.* (2010) and much of the extant literature arises most likely because they consider a municipality's own population or income to compute market access. Here, I separate between what the market literature might call external market access (i.e., market access computed on other municipality) and 'internal' market access (i.e., a municipality's own population). The 'internal market access' label is somewhat misleading we need to keep in mind that there are other benefits from agglomeration that do not percolate through input-output linkages but instead are associated with thick local labour markets and learning spillovers. The literature that attempts to assess the effects of market access on wages usually fails to separate between different possible benefits from agglomeration.[9]

Keeping this distinction in mind between external and 'internal' market access, we can go back to the simple equilibrium condition stated above $(w_i - H_i)/P_i =$ constant. Then, it is clear from the results obtained above that a larger local population raises local wages. As shown in Duranton (2014*b*) larger cities are also more expensive places to live in due to higher housing and transportation costs. In the context of this simple spatial equilibrium, equilibration is probably occurring through the adjustment of urban costs when it comes to 'internal' market access (i.e., city population). On the other hand, a higher 'external' market access may predominantly affect the consumption side (i.e., by lowering the price index P) and may thus lead to lower nominal wages especially if the population response is weak and does not lead to higher urban costs.

The second important key to understanding this diversity of results regarding market access is that standard measures of market access imply that close neighbours share similar market accesses (their distances to other municipalities are roughly similar). We also expect these places to offer similar wages because the 'first-nature' determinants of wages are going to be similar and because workers may be able to arbitrage more easily between nearby locations through commuting. Hence, a strong positive correlation between wages and market access is expected. This correlation however may not be driven by the input output linkages emphasised by the models in the spirit of Krugman (1980) or Krugman and Venables (1995). In table 6, market access is computed using information on trading patterns across municipalities where short distance trade is not taken into account. While this may raise some issues of selection into trade and measurement error as mentioned above, this has nonetheless the great advantage of ignoring market access being computed using information from close neighbours which is likely correlated with other phenomenon that will affect the wage such as sharing the same local labour market, etc.

5.3 *Transportation infrastructure*

Related to market access, the local transportation infrastructure may also matter. In a related paper (Duranton, 2013*a*), I show that better provision of roads or better accessibility to the main road network of the country has a sizeable effect on both the volume and the value of exports

[9] Redding and Sturm (2008) is an exception since they look at German cities close to the Iron curtain. These cities lost access following the erection of the Iron Curtain in the early 1950s without initially losing their own agglomeration benefits. It must also be said that separating between the different channels through which agglomeration benefits percolate is an extremely challenging enterprise (Puga, 2010).

of Colombian municipalities. Whether roads also have an effect on wages is, of course, another question.

Table 7: Agglomeration effects and roads, OLS specifications

	Road index 1			Road index 2		
	(1)	(2)	(3)	(4)	(5)	(6)
log population	0.051[a]	0.049[a]	0.065[a]	0.050[a]	0.049[a]	0.065[a]
	(0.0091)	(0.010)	(0.017)	(0.0100)	(0.011)	(0.017)
Roads	0.0038	0.0037	0.0066	0.0053	0.0052	0.0096[c]
	(0.0074)	(0.0076)	(0.0040)	(0.011)	(0.011)	(0.0058)
share educ .	N	Y	Y	N	Y	Y
Geography+access	N	N	Y	N	N	Y
Observations	2,735,518	2,734,355	1,870,884	2,735,518	2,734,355	1,870,884
R^2	0.40	0.40	0.41	0.40	0.40	0.41
Municipalities	645	643	329	645	643	329

Notes: OLS regressions with year indicators and individual characteristics in all columns. Robust standard errors clustered at the municipality level in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. The dependent variable is log wage in all columns. Geographic characteristics are the same as in table 2. Market access is as in table 2. The road index in columns 1 to 3 is sum the log of 1 + the number of major national roads in the municipality, the log of 1 + the number of exits in and out of the municipality, the log of 1+ kilometers of national roads minus the log of 1 + kilometers of distance to the closest national road if there are no road in the municipality. The road index in columns 4 to 6 only sums the first terms of road index 1.

Table 7 reports results for two series of regressions using different measures of roads. Columns 1 to 3 use an index summing the number of national roads, their number of entries into the municipality, the mileage and, for municipalities with no national roads, the (negative) distance to the closest national road. Columns 4 to 6 use a similar index that does not consider the distance to the closest road. The first result in table 7 is that the estimated elasticity of wages with respect to city population is not affected by the inclusion of road variable. The second main finding in table 7 is that the road variable is significant in only one occasion. Because of the way the two road indices used in table 7 are constructed, the roads coefficients can be loosely interpreted as elasticities. They are always less than 1%. Importantly, the standard errors around these coefficients are also extremely small. As a result, we can typically rule out OLS elasticities as low as 3%. I experimented with these findings using the measures of roads that compose these indices one at a time and found similar results.

Following extant literature on the effects of roads on urban outcomes (Baum-Snow, 2007, Duranton and Turner, 2012), the worry is that road allocation may be endogenous. When assessing the effects of roads on the population growth of US cities, Duranton and Turner (2012) provide evidence of such endogeneity and provide instrumental variable results that suggest much stronger effects of roads than can be computed directly from the cross-section using OLS. Here, I follow the identification strategy used in Duranton and Turner (2012) and instrument the road indices used above by the corresponding indices computed from the 1938 road network and colonial roads ('caminos reales'). Because population is also potentially endogenous and correlated with current roads, I also instrument population by the same population lags as in table 3. These instrument will rule out any form of reverse causality. However, old roads may also be correlated with contemporaneous characteristics of Colombian municipalities that determine their wages. For this reason, I also estimate IV regressions where the share of university educated workers and geographical characteristics are also included.

Table 8: Agglomeration effects and roads, LIML specifications

	Road index 1			Road index 2		
	(1)	(2)	(3)	(4)	(5)	(6)
log population	0.015	-0.0012	0.018	0.015	-0.00016	0.020
	(0.023)	(0.028)	(0.023)	(0.020)	(0.027)	(0.022)
Roads	0.037[c]	0.037[c]	0.031[a]	0.044[b]	0.039[c]	0.040[a]
	(0.020)	(0.021)	(0.0082)	(0.020)	(0.023)	(0.013)
log share educ .	N	Y	Y	N	Y	Y
Geography+access	N	N	Y	N	N	Y
Overid. p-value	0.11	0.074	0	0.14	0.089	0
First-stage Stat.	4.00	3.83	6.73	5.19	4.87	4.50
Observations	2,026,723	2,026,723	1,443,955	2,026,723	2,026,723	1,443,955
Municipalities	316	316	164	316	316	164

Notes: This table duplicates table 8 using limited information maximum likelihood LIML. In all columns, log population is instrumented by the corresponding log populations for 1843, 1870, 1938. The road index is instrumented by the corresponding indices computed from the 1938 road network and from the colonisation roads ('caminos reales').

The results are reported in table 8. Unfortunately, the instruments are weak and much caution is needed when interpreting these results. It can nonetheless be noted that all six specifications estimate a significant coefficient on the road index between 0.03 and 0.05. On the other hand, the coefficient on city population becomes insignificant. Interestingly, the sum of the coefficients on the road index and city population remains the same as before at about 0.05-0.06. This suggests that the regressions cannot identify the effects of roads and population separately when using instrumental variables. This is due to the fact that past populations and past road networks predict current roads better than current population. If we think that the true elasticity of wages with respect to city population is around 0.05, that only leaves a tiny positive effect of roads on wages.

Remember also that whatever effect of roads on wages is measured, this is an overall effect. As shown in Duranton (2013*a*) roads in a municipality foster exports. In itself, this should have a positive effect on wages. Roads may also lead to higher wages because they improve local productivity by facilitating local trades. At the same time, roads may be thought of as improving market access. In line with the results of the previous section, a better access, which lowers the price of imported goods, may lead to lower wages if labour is mobile across places. Hence documenting a zero or very small coefficient is not synonymous with the absence of an effect. Roads may have countervailing effects on wages.

5.4 *Amenities*

Models of spatial equilibrium involving utility equalization across locations usually imply that amenities will be capitalised into property prices through a higher demand for land and through wages. To live in locations with better amenities, workers are willing to accept a combination of higher rents and lower wages. The channel through which amenities capitalise into property price is straightforward. Since housing is in limited supply in any location, a greater demand for housing will lead to higher prices. The mechanism through which amenities capitalise into wages is indirect. Firms that face higher land prices use less land, which, in turn, reduces the marginal product of labour and thus the wage (Roback, 1982, Glaeser and Gottlieb, 2009).

There is a large literature using data from developed countries that argues that climate is an important amenity (Rappaport, 2007, Cheshire and Magrini, 2006). There are major differences in climate between Colombian cities caused mainly by differences in their elevation. This said, from the regressions estimated in section 4 the effect of climate variables on wages in Colombia is weak and it is unclear whether the climate affects wages indirectly through the channel just described or directly as a factor of production.

Instead, table 9 performs three series of regressions with three types of non-climate amenity variables. Columns 1 to 3 consider leisure amenities: libraries, museums, and hotels and restaurants which are either amenities in themselves or proxies for amenities. Column 1 adds these three variables to the baseline specification of column 2 in table 2. Column 2 also adds the log share of educated workers. Column 3 further includes geographic controls and market access. The same

Table 9: Agglomeration effects and amenities, OLS specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
log population	0.067^a	0.075^a	0.12	0.043^a	0.043^a	0.13^a	0.053^a	0.054^a	0.066^a
	(0.012)	(0.016)	(0.074)	(0.012)	(0.016)	(0.042)	(0.0062)	(0.0098)	(0.018)
log libraries p.c.	0.058^b	0.061^b	-0.038						
	(0.025)	(0.025)	(0.031)						
log museum p.c.	-0.0026	0.0015	0.0076						
	(0.027)	(0.027)	(0.012)						
log hospitality p.c.	-0.047^c	-0.045	-0.0042						
	(0.027)	(0.028)	(0.041)						
log homicide p.c.				0.012	0.012	-0.015			
				(0.025)	(0.034)	(0.014)			
log violence 1							0.086^b	0.087^b	-0.019
							(0.036)	(0.038)	(0.021)
log violence 2							-0.053	-0.054	0.020
							(0.035)	(0.037)	(0.017)
share educ .	N	Y	Y	N	Y	Y	N	Y	Y
Geography+access	N	N	Y	N	N	Y	N	N	Y
Observations	2,326,863	2,326,863	1,637,316	2,340,482	2,340,482	1,639,045	2,735,289	2,734,126	1,870,884
R^2	0.40	0.40	0.41	0.40	0.40	0.41	0.40	0.40	0.41
Municipalities	102	102	39	118	118	44	644	642	329

Notes: OLS regressions with year indicators and individual characteristics in all columns. Robust standard errors clustered at the municipality level in parentheses. *a*, *b*, *c*: significant at 1%, 5%, 10%. The dependent variable is log wage in all columns. Geographic characteristics are the same as in table 2 and also include log market access. Library per capita is computed as the 1+ the number of libraries in 2009 divided by 2005 population. One is added to accommodate the log transformation. Museum per capita is computed as the 1+ the number of museums in 1998 + the number of museums in 2009 divided by 2005 population. Hospitality per capita is computed as the 1+ the number of restaurants in 2005 +20× the number of hotels in 2005 divided by 2005 population. Homicide per capita is computed as the 1+ the number of homicides in 2002 + the number of homicides in 2009 divided by 2005 population. Violence 1 is computed as 1+ the sum over 1993 to 2005 of all operations by the army and any of the unofficial armed groups. Violence 2 is 1+ the sum of all deaths attributed to unofficial armed groups over 1993-2005.

patterns is repeated in columns 4 to 6 with the log number of homicide per capita and in columns 7 to 9 with two variables associated with political violence between 1993 and 2005: the number of military operations plus attacks by unofficial armed groups and the number of death attributed to armed groups.

The patterns are weak. Among the three leisure activities, the coefficient for the number of libraries is only significant with a positive sign in columns 1 and 2, whereas hospitality has a negative significant coefficient only in column 1. The sign of the coefficient on the number of libraries may be viewed as 'perverse' since we expect good amenities to be associated with lower wages at the spatial equilibrium. However, this is probably due to reverse causation where richer municipalities can afford more libraries. In columns 4 to 6, the number of homicides is never significant. Finally, the first political violence variable has a positive coefficient that is significant at 5% in columns 7 and 8 but becomes insignificant in column 9. The other political variable is always insignificant.

Overall, the patterns are weak since none of the amenity variables considered here (nor in experimentations not reported here) appears to have a robust effect. This is consistent with amenities playing no discernible role in urban growth and Colombian households spending only a minimal fraction of their income on leisures (Meisel Roca and Pérez, 2012). Because the standard errors on the amenity coefficients are small, it is better to think about these results as being zeroes that are fairly precisely estimated rather than as having a complete lack of pattern. It is also noteworthy that this precision is obtained despite the amenity variables used in columns 1 to 6 being available for only a small subset of municipalities. There is of course the worry that these coefficients are biased by reverse causality. While this remains a possibility, it would require two things. First, the effect of all positive amenities should be overestimated (i.e., be zero instead of negative) because they are caused by greater local incomes. Second, the effect of negative amenities should be underestimated because they are caused by local poverty. It is hard to imagine opposite biases affecting positive and negative amenities with just this exact pattern.

To go deeper on this, it is possible that amenities affect more educated individuals more either because they are more mobile or because they value these amenities more, positively or negatively. Duplicating table 9 by adding some interactions between amenities and education does not change anything and the coefficients on the interaction terms are erratic in their sign and lack precision.

On the other hand, the coefficient on log city size remains positive, significant, and of the same magnitude as previously in all columns except column 3 where significance is lacking with only 39 municipalities (and essentially 17 explanatory variables at the municipality level).

6. Conclusions

This paper provides a broad exploration of agglomeration benefits in Colombian cities. The main finding is an elasticity of wages with respect to population of about 5%. This coefficient is extremely robust to the inclusion of a wide variety of controls and to the use of now standard instrumental variable strategies. This elasticity suggests sizeable gains from 'urbanisation'. Moving from a city with 10,000 inhabitants to Bogota with more than 7 million is associated with about

40% higher wages. This elasticity of about 5% is somewhat larger than the same elasticity estimated across US and French cities but smaller than for Chinese and Indian cities.

My other findings are either weaker or negative. I found only weak evidence regarding human capital externalities. The share of more educated workers in a city appears to capture the same effects as city population and, as soon as both variables are included together, the share of more educated workers turns out less robust. I also found suggestive and novel evidence of a negative effect of market access on wages. This effect is probably best understood in spatial equilibrium framework with labour mobility. The evidence about roads suggests no effect of roads on wages (or tiny positive effects at best). There is also no effect of urban amenities on wages.

These findings should be useful for two reasons. First, they provide a rather comprehensive documentation of agglomeration effects for Colombian cities. This should be helpful at a time when the country rediscovers its cities and is starting to rethink the organisation of its territory after having considered it for many years mainly though military lenses.

This broad set of findings, positive and negative, should also be useful to develop a broader knowledge base about agglomeration effects in developing countries. There is a now a lot of consistent evidence about agglomeration economies in more advanced countries. The evidence for developing countries still lags behind despite their crucial importance at a time when many of these countries are fast urbanising.

References

Angrist, Joshua D. and Adriana D. Kugler. 2008. Rural windfall or a new resource curse? Coca, income, and civil conflict in colombia. *Review of Economics and Statistics* 90(2):191–215.

Attanasio, Orazio, Pinelopi K. Goldberg, and Nina Pavcnik. 2004. Trade reforms and wage inequality in Colombia. *Journal of Development Economics* 74(2):331–366.

Bacolod, Marigee, Bernardo S. Blum, and William C. Strange. 2009. Skills in the city. *Journal of Urban Economics* 65(2):136–153.

Barco, Carolina. 2013. *Misión para el fortalecimiento del sistema de ciudades de Colombia. Documento técnico de soporte política del sistema de ciudades de Colombia*. Report for the Colombian National Planning Ministry (Departamento Nacional de Planeación, Dirección de Desarrollo Urbano).

Baum-Snow, Nathaniel. 2007. Did highways cause suburbanization? *Quarterly Journal of Economics* 122(2):775–805.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2014. Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy* 122:(forthcoming).

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

Chauvin, Juan Pablo, Edward L. Glaeser, and Kristina Tobio. 2013. Urban economics in the US and India. Manuscript in progress, Harvard University.

Cheshire, Paul C. and Stefano Magrini. 2006. Population growth in european cities: weather matters-but only nationally. *Regional Studies* 40(1):23–37.

Ciccone, Antonio and Robert E. Hall. 1996. Productivity and the density of economic activity. *American Economic Review* 86(1):54–70.

Combes, Pierre-Philippe, Sylvie Démurger, and Li Shi. 2013. Urbanisation and migration externalities in China. Processed, Aix-Marseille School of Economics.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2010. Estimating agglomeration economies with history, geology, and worker effects. In Edward L. Glaeser (ed.) *The Economics of Agglomeration*. Cambridge (MA): National Bureau of Economic Research, 15–65.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2012. Sorting and local wage and skill distributions in France. *Regional Science and Urban Economics* 42(6):913–930.

Combes, Pierre-Philippe and Laurent Gobillon. 2014. The empirics of agglomeration economies. Chapter in progress for the *Handbook of Regional and Urban Economics*.

De la Roca, Jorge and Diego Puga. 2012. The dynamic earnings premium of dense cities. Processed, CEMFI and IMDEA Social Sciences.

Departamento Administrativo Nacional de Estadística. 2013. *Great Integrated Household Survey (GEIH) Methodology*. Bogota, Colombia: DANE.

Donaldson, Dave and Richard Hornbeck. 2013. Railroads and American economic growth: A "market access" approach. Processed, Harvard University and MIT.

Duranton, Gilles. 2013*a*. Roads and trade between Colombian cities. Manuscript in progress, Wharton School, University of Pennsylvania.

Duranton, Gilles. 2013*b*. Delineating metropolitan areas: Measuring spatial labour market networks through commuting patterns. Mimeographed, Wharton School, University of Pennsylvania.

Duranton, Gilles. 2014*a*. Growing through cities in developing countries. *World Bank Research Observer.* Manuscript written for a special issue on 'jobs and development'.

Duranton, Gilles. 2014*b*. Urban costs in Colombian cities. Manuscript in progress, Wharton School, University of Pennsylvania.

Duranton, Gilles, Peter M. Morrow, and Matthew A. Turner. 2014. Roads and trade: Evidence from the US. *Review of Economic Studies* (forthcoming).

Duranton, Gilles and Diego Puga. 2004. Micro-foundations of urban agglomeration economies. In J. Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2063–2117.

Duranton, Gilles and Matthew A. Turner. 2012. Urban growth and transportation. *Review of Economic Studies* 79(4):1407–1440.

Fally, Thibault, Rodrigo Paillacar, and Cristina Terra. 2010. Economic geography and wages in Brazil: Evidence from micro-data. *Journal of Development Economics* 91(1):155–168.

Glaeser, Edward L. 1999. Learning in cities. *Journal of Urban Economics* 46(2):254–277.

Glaeser, Edward L. and Joshua D. Gottlieb. 2008. The economics of place-making policies. *Brookings Papers on Economic Activity* (1):155–253.

Glaeser, Edward L. and Joshua D. Gottlieb. 2009. The wealth of cities: Agglomeration economies and spatial equilibrium in the United States. *Journal of Economic Literature* 47(4):983–1028.

Glaeser, Edward L. and David C. Maré. 2001. Cities and skills. *Journal of Labor Economics* 19(2):316–342.

Glaeser, Edward L. and Matthew R. Resseger. 2010. The complementarity between cities and skills. *Journal of Regional Science* 50(1):221–244.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. Identifying agglomeration spillovers: Evidence from winners and losers of large plants openings. *Journal of Political Economy* 118(3):536–598.

Handbury, Jessie H. and David E. Weinstein. 2013. Is new economic geography right? Evidence from price data. Processed, Columbia University.

Henderson, J. Vernon. 2005. Urbanization and growth. In Philippe Aghion and Steven N. Durlauf (eds.) *Handbook of Economic Growth*, volume 1B. Amsterdam: North-Holland, 1543–1591.

Krugman, Paul. 1980. Scale economies, product differentiation, and the patterns of trade. *American Economic Review* 70(5):950–959.

Krugman, Paul R. and Anthony J. Venables. 1995. Globalization and the inequality of nations. *Quarterly Journal of Economics* 110(4):857–880.

Marshall, Alfred. 1890. *Principles of Economics*. London: Macmillan.

Meisel Roca, Adolfo and Gerson Javier Pérez. 2012. Las ciudades colombianas y sus attractivos. Report for the Urban Development Division of the National Planning Ministry (DNP-DDU).

Michaels, Guy. 2008. The effect of trade on the demand for skill - Evidence from the Interstate Highway System. *Review of Economics and Statistics* 90(4):683–701.

Moretti, Enrico. 2004. Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics* 121(1):175–212.

Moretti, Enrico. 2011. Local labor markets. In Orley Ashenfelter and David Card (eds.) *Handbook of Labor Economics*, volume 4B. Amsterdam: North-Holland, 1237–1313.

Overman, Henry G. and Anthony J. Venables. 2005. Cities in the developing world. Report, Department for International Development (UK).

Peri, Giovanni. 2002. Young workers, learning, and agglomerations. *Journal of Urban Economics* 52(3):582–607.

Puga, Diego. 2010. The magnitude and causes of agglomeration economies. *Journal of Regional Science* 50(1):203–219.

Rappaport, Jordan. 2007. Moving to nice weather. *Regional Science and Urban Economics* 37(3):375–398.

Rauch, James E. 1993. Productivity gains from geographic concentration of human capital: Evidence from the cities. *Journal of Urban Economics* 34(3):380–400.

Redding, Stephen and Anthony J. Venables. 2004. Economic geography and international inequality. *Journal of International Economics* 62(1):63–82.

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

Redding, Stephen J. and Matthew A. Turner. 2014. Transportation costs and the spatial organization of economic activity. Chapter in progress for the *Handbook of Regional and Urban Economics*.

Roback, Jennifer. 1982. Wages, rents and the quality of life. *Journal of Political Economy* 90(6):1257–1278.

Rosen, Sherwin. 1979. Wage-based indexes of urban quality of life. In Peter Mieszkowski and Mahlon Straszheim (eds.) *Current Issues in Urban Economics*. Baltimore: Johns Hopkins University Press, 74–104.

Samad, Taimur, Nancy Lozano-Gracia, and Alexandra Panman. 2012. *Colombia Urbanization Review. Amplifying the Gains from the Urban Transition*. Washington (DC): World Bank.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: Printed for W. Strahan, and T. Cadell.

Sánchez, Fabio and Jairo Nuñez. 2000. Geography and economic development: A municipal approach for Colombia. *Revista de Planeación y Desarrollo* 31(3/4):141–156.

Wheeler, Christopher H. 2001. Search, sorting, and urban agglomeration. *Journal of Labor Economics* 19(4):879–899.

Wheeler, Christopher H. 2006. Cities and the growth of wages among young workers: Evidence from the NLSY. *Journal of Urban Economics* 60(2):162–184.

Appendix A. Supplementary results

Table 10: Agglomeration effects, OLS specifications for workers in large municipalities

	(1) Only pop.	(2) Indiv.	(3) Geog.	(4) Access	(5) Muni. pop.	(6) MA	(7) Non-lin.	(8) Comp.
log population	0.11[a]	0.056[a]	0.049[a]	0.061[a]	0.058[a]	0.049[a]	0.11	0.056[a]
	(0.0070)	(0.0058)	(0.014)	(0.0075)	(0.0069)	(0.0093)	(0.072)	(0.0059)
log population^2							-0.0021	
							(0.0030)	
log market access				-0.13[a]				
				(0.039)				
log market access^2				0.0069[a]				
				(0.0023)				
Educ $\times$ log pop.								-0.011[c]
								(0.0059)
Observations	2,568,305	2,565,516	2,550,545	1,779,787	2,565,516	2,565,516	2,565,516	2,565,516
R^2	0.09	0.40	0.41	0.41	0.40	0.36	0.40	0.40
Municipalities	296	296	289	153	296	247	296	296

Notes: This table duplicates table 2 in the main text but samples only workers in municipalities with population above 20,000 in 1993.

Table 11: Agglomeration effects, OLS specifications for prime-aged males

	(1) Only pop.	(2) Indiv.	(3) Geog.	(4) Access	(5) Muni. pop.	(6) MA	(7) Non-lin.	(8) Comp.
log population	0.11[a]	0.051[a]	0.061[a]	0.049[a]	0.055[a]	0.044[a]	0.11[b]	0.051[a]
	(0.0053)	(0.0042)	(0.015)	(0.0059)	(0.0050)	(0.0070)	(0.047)	(0.0042)
log population^2							-0.0026	
							(0.0021)	
log market access				-0.14[a]				
				(0.073)				
log market access^2				0.0073[a]				
				(0.0022)				
Educ $\times$ log pop.								0.0081
								(0.0052)
Observations	1,128,897	1,127,543	1,117,168	758,149	1,127,543	1,127,543	1,127,543	1,127,543
R^2	0.13	0.42	0.42	0.42	0.42	0.35	0.42	0.42
Municipalities	644	644	617	336	644	571	644	644

Notes: This table duplicates table 2 in the main text but samples only prime-aged males (25-55).

Table 12: Agglomeration effects, OLS specifications for prime-aged females

	(1) Only pop.	(2) Indiv.	(3) Geog.	(4) Access	(5) Muni. pop.	(6) MA	(7) Non-lin.	(8) Comp.
log population	0.098^a	0.057^a	0.046^a	0.063^a	0.063^a	0.048^a	0.054	0.058^a
	(0.0064)	(0.0054)	(0.013)	(0.0064)	(0.0062)	(0.0078)	(0.048)	(0.0055)
log population^2							0.00014	
							(0.0021)	
log market access				-0.12^a				
				(0.036)				
log market access^2				0.0066^a				
				(0.0021)				
Educ $\times$ log pop.								-0.023^a
								(0.0049)
Observations	912,906	912,072	905,807	638,212	912,072	912,072	912,072	912,072
R^2	0.08	0.40	0.41	0.40	0.40	0.36	0.40	0.40
Municipalities	642	642	615	335	642	569	642	642

Notes: This table duplicates table 2 in the main text but samples only prime-aged females (25-55).

Table 13: Agglomeration effects, OLS specifications for males aged 20 to 30

	(1) Only pop.	(2) Indiv.	(3) Geog.	(4) Access	(5) Muni. pop.	(6) MA	(7) Non-lin.	(8) Comp.
log population	0.10^a	0.056^a	0.063^a	0.055^a	0.061^a	0.049^a	0.059	0.056^a
	(0.0047)	(0.0049)	(0.016)	(0.0075)	(0.0058)	(0.0077)	(0.049)	(0.0049)
log population^2							-0.00015	
							(0.0021)	
log market access				-0.16^a				
				(0.046)				
log market access^2				0.0087^a				
				(0.0027)				
Educ $\times$ log pop.								-0.0090
								(0.0074)
Observations 438,213	437,832	433,628	295,806	437,832	437,832	437,832	437,832	
R^2	0.16	0.35	0.37	0.36	0.35	0.26	0.35	0.35
Municipalities	641	641	614	334	641	568	641	641

Notes: This table duplicates table 2 in the main text but samples only males aged 20 to 30.

Table 14: Agglomeration effects, OLS specifications for males aged 40 to 50

	(1) Only pop.	(2) Indiv.	(3) Geog.	(4) Access	(5) Muni. pop.	(6) MA	(7) Non-lin.	(8) Comp.
log population	0.12^a	0.050^a	0.058^a	0.048^a	0.053^a	0.042^a	0.15^a	0.049^a
	(0.0060)	(0.0040)	(0.016)	(0.0051)	(0.0047)	(0.0067)	(0.048)	(0.0040)
log population^2							-0.0042^c	
							(0.0021)	
log market access				-0.14^a				
				(0.035)				
log market access^2				0.0071^a				
				(0.0021)				
Educ $\times$ log pop.								0.011^c
								(0.0066)
Observations	367,348	366,861	363,480	246,049	366,861	366,861	366,861	366,861
R^2	0.11	0.43	0.44	0.43	0.43	0.37	0.43	0.43
Municipalities	641	641	614	334	641	568	641	641

Notes: This table duplicates table 2 in the main text but samples only males aged 40 to 50.

Table 15: Agglomeration effects, OLS for different years

	(1) 1996-1997	(2) 1998-1999	(3) 2000-2001	(4) 2002-2003	(5) 2004-2005	(6) 2008-2009	(7) 2010-2011	(8) 2012
log population	0.070^a	0.061^a	0.047^a	0.040^a	0.055^a	0.056^a	0.056^a	0.055^a
	(0.0056)	(0.0067)	(0.0044)	(0.0048)	(0.0046)	(0.0065)	(0.0065)	(0.0074)
Observations	90,950	89,024	239,130	341,230	411,916	602,326	639,090	322,567
R^2	0.37	0.35	0.35	0.37	0.38	0.38	0.37	0.36
Municipalities	229	201	218	269	375	568	430	427

Notes: This table duplicates the specification of table 2 column in the main text across different time periods.

Table 16: Agglomeration and human capital externalities (measured in log), OLS

	(1) Only pop.	(2) Indiv.	(3) Geog.	(4) Access	(5) Muni. pop.	(6) MA	(7) Non-lin.	(8) Comp.
log population	0.055[a]	0.052[a]	0.055[a]	0.061[a]	0.047[a]	0.035[a]	0.068	0.052[a]
	(0.011)	(0.0088)	(0.013)	(0.014)	(0.0096)	(0.011)	(0.050)	(0.0089)
log share educ.	1.76[a]	0.068	0.59[a]	-0.19	0.38	0.42	0.047	0.064
	(0.38)	(0.30)	(0.14)	(0.42)	(0.27)	(0.26)	(0.31)	(0.30)
log population^2							-0.00064	
							(0.0020)	
log market access				-0.13[a]				
				(0.042)				
log market access^2				0.0071[a]				
				(0.0025)				
Educ $\times$ log pop.								-0.0081[c]
								(0.0043)
Observations	2,737,326	2,734,355	2,711,293	1,872,901	2,734,355	2,734,355	2,734,355	2,734,355
R^2	0.11	0.40	0.41	0.41	0.40	0.36	0.40	0.40
Municipalities	643	643	616	336	643	570	643	643

Notes: This table duplicates table 2 adding the log share of population with some university education.

Table 17: Agglomeration effects, OLS specifications for workers in large municipalities

Market access:	2011 w (1)	2004 w (2)	2010 pop (3)	1993 pop (4)	2009 GDP (5)	2010 pop (6)	1993 pop (7)	2009 GDP (8)
log population	0.054[a]	0.055[a]	0.052[a]	0.053[a]	0.053[a]	0.056[a]	0.056[a]	0.055[a]
	(0.0056)	(0.0055)	(0.0049)	(0.0046)	(0.0035)	(0.0091)	(0.0092)	(0.0091)
log market access	1.69	1.86	0.59	0.44	1.15[c]	-0.012	-0.0084	-0.019
	(1.38)	(1.39)	(1.02)	(1.09)	(0.67)	(0.055)	(0.053)	(0.059)
log market access^2	-0.030	-0.032	-0.018	-0.012	-0.034	0.00052	0.00035	0.00083
	(0.027)	(0.026)	(0.042)	(0.045)	(0.023)	(0.0032)	(0.0032)	(0.0027)
Observations	2,713,205	2,713,205	2,713,205	2,713,205	2,713,205	2,288,367	2,288,367	2,288,367
R^2	0.41	0.41	0.41	0.41	0.41	0.40	0.40	0.40
Municipalities	616	616	616	616	616	124	124	124

Notes: OLS regressions with year indicators and individual characteristics in all columns. Robust standard errors clustered at the municipality level in parentheses. a, b, c: significant at 1%, 5%, 10%. The dependent variable is log wage in all columns. Market access is computed discounting by Euclidian distances in columns 1 to 5 and road distance in column 6 to 9. Only municipalities with population above 40,000 in 2010 are considered in the computation of market access in all columns. Only workers in municipalities with population above 40,000 are considered in columns 6 to 8. Market access is computed using 2011 wages in column, 2004 wages in column 2, 2010 population in column 3, 1993 population in column 4, 2009 GDP in column 5, and again 2010 population, 1993 population and 2009 GDP in column 6, 7, and 8, respectively.