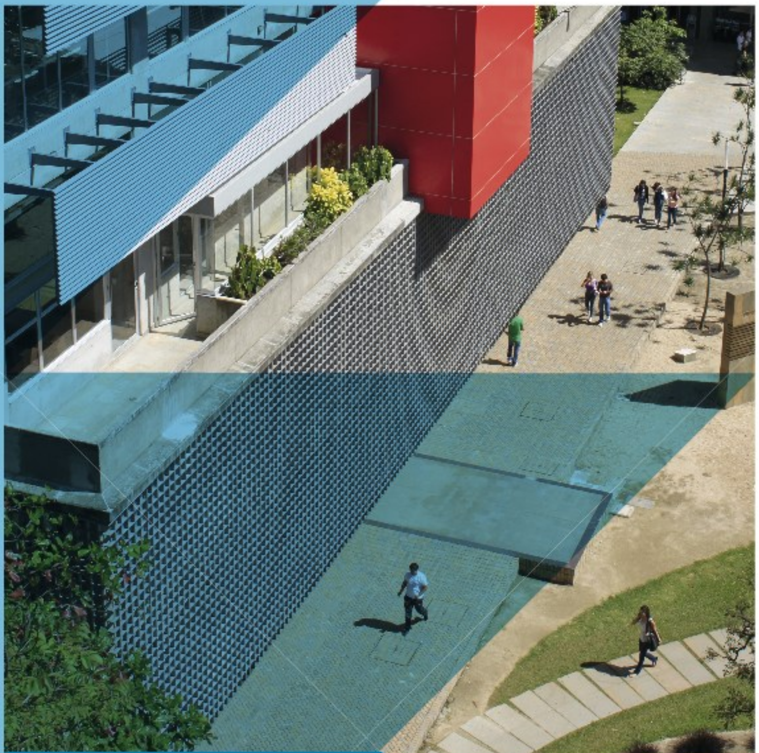


Escuela de Economía y Finanzas

Documentos de trabajo

Economía y Finanzas

Centro de Investigación Económicas y Financieras

No. 18-01
2018

Agglomeration Economies in the Presence of an Informal Sector The Colombian Case

Gustavo A. García

cief
Centro de Investigaciones Económicas y Financieras
Universidad EAFIT

Agglomeration Economies in the Presence of an Informal Sector
The Colombian Case

Gustavo A. García[*]
Research in Spatial Economics (RiSE-group)
Department of Economics, Universidad EAFIT
Medellín, Colombia
ggarci24@eafit.edu.co

Abstract

This paper analyzes the relationship between agglomeration economies and productivity in the context of a developing country while taking into account the marked presence of an informal sector. Using data from Colombia, we investigate the effect of agglomeration economies on formal and informal productivity. We examine whether the informal sector achieves benefits from agglomeration economies as well as whether there are differences between the formal and informal sectors in terms of agglomeration returns. We find that agglomeration economies, measured by the density of local employment, have a significantly positive effect on productivity in the informal sector, while there is little effect in the formal sector. We estimate an elasticity of wages with respect to employment density of approximately 2% for the informal sector, which implies that informal workers in denser areas will earn approximately 11% more than those in less dense areas.

Keywords: Agglomeration economies, informal sector, Colombia
JEL classification: R12, J46, R23, J31

[*] This paper has benefited from conference and seminars at the 62nd Annual North American Meetings of the Regional Science Association International (NARSC-2015), First Network on Microeconomics - AFADECO, Universidad EAFIT Economics Department Seminar, and CEDE seminar at Universidad de Los Andes, and especially Harris Selod, Rosella Nicolini, José Luis Roig and three anonymous referees for their helpful comments and suggestions on an early version of the paper. This paper has benefited from the support of Universidad de Los Andes and Colciencias program "Es Tiempo de Volver" during my postdoctoral research fellow at the Universidad de Los Andes.

1. Introduction

The pace and content of urbanization have crucial implications for developing economies. Among the key benefits of urbanization are gains from agglomeration. There are benefits to location externalities that arise from a dense network of production and market access links that increase productivity and decrease the unit costs for each firm in the network (Fujita et al., 1999). It can be posited, however, that the magnitude of agglomeration economies depends on the types of workers and industries as well as on the period and country analyzed. In this sense, it is important to understand whether agglomeration economies produce similar benefits for developing countries compared to those previously demonstrated for developed countries (see, for example, Ciccone and Hall (1996), Rosenthal and Strange (2008a), Melo et al. (2009), Melo and Graham (2009)).

Despite the rapid and continuing pace of urbanization in developing countries, formalization seems to have stalled, or at least it does not appear to be increasing as quickly as might be expected given these countries growth rates. The formal sector in developing economies is only responsible for a certain share of urban employment and growth. The informal sector, on the other hand, plays a large role in such economies, which constitutes an important difference between developing and developed economies (Schneider and Enste, 2000). According to estimates by Jütting and De Laiglesia (2009), over 55% of non-agricultural employment in developing countries is neither regulated nor protected by the state (informal activities). With regard to the size of the informal economy as measured as a percentage of GDP, Schneider et al. (2010) show that the shadow economy in developing countries accounts for approximately 40% of GDP. This discernible presence of the informal sector can affect the extent (or quality) of agglomeration economies, and the effects of urbanization could be just as likely to be found in the outcomes for the informal sector as for the formal sector.

Given the significant differences in the economic characteristics of the informal and formal sectors (productivity, profitability, and size), there are different perspectives regarding how the informal sector contributes to and benefits from agglomeration economies. For instance, Annez and Buckely (2009) argue that the informal sector is unproductive and increases the costs for the formal sector, thus crowding out agglomeration economies. In contrast, Overman and Venables (2005) and Moreno-Monroy (2012) state that the informal sector also contributes to and benefits from agglomeration economies via its interaction with the formal sector throughout different value chains, whereby the informal sector both obtains inputs from the formal sector and supplies it with intermediate or final goods and services. As noted by Duranton (2009), there are intense links between the formal and informal sectors, which suggests that agglomeration effects are generated within both sectors, with benefits that accrue to both. According to Overman and Venables (2005), the existence of an informal sector can affect the benefits of agglomeration economies in two ways. On the one hand, the presence of an informal sector can drive up urban costs and crowd out the formal sector, but, on the other hand, the informal sector also contributes to agglomeration economies. In this sense, the informal sector comprises small-scale producers that establish important networks that contribute to the formation of clusters. Furthermore, as in the formal sector, the informal sector can achieve benefits from the productivity effects associated with the concentration of activity and employment.

In this paper, we investigate the effect of agglomeration economies on formal and informal productivity, and we analyze whether the formal or informal sector achieves greater benefits from the diversity of activities and the spillovers associated with urbanization economies. For the analysis, we used worker data for Colombia from the 2008-2014 period. The empirical analysis is based on regressions of individual worker wage rates, as a measurement of labor productivity, on employment density, which in turn is a measure of urban agglomeration. We calculated the elasticity of wages with respect to density for the formal and informal sectors while controlling for several socioeconomic, socio-demographic and regional characteristics. These regressions comprise instrumental variable estimates to correct for the endogeneity attributable to the reverse causality between wages and agglomeration.

The purpose of this study is to provide new evidence on urbanization and its effects on developing countries while more closely considering the reality in such countries where formal and informal activities co-exist. To date few studies have analyzed the agglomeration effects in developing countries while taking into account the presence of the informal sector.

The remainder of this paper is organized as follows: Section 2 presents the literature review. The theoretical framework is described in Section 3. Section 4 defines the empirical model used. In Section 5, we present the data sources used in the analysis. Section 6 statistically documents the relationship between agglomeration and wages while taking into account the existence of the informal sector. Sections 7 and 8 discuss the results, and Section 9 concludes.

2. Literature review

In this section, we provide a brief outline of the prior empirical results on the presence of agglomeration economies in some developing countries. The related literature is scarce, and the extant studies generally find that agglomeration economies have a significant effect on productivity, that is much higher compared to estimates for developed countries.[1]

Developing countries are characterized by certain structural conditions such as economic and political instability, high rates of unemployment and underemployment, shallow markets, and low levels of industrial and infrastructure development. These conditions can affect the magnitude and quality of the external economies associated with agglomeration economies. Several studies have found quantitative evidence of both localization and urban effects. Henderson (1986), in his seminal work analyzing the role of localization and urbanization economies in productivity in the metropolitan areas of Brazil, finds that localization

[1] The two benchmark studies that use aggregate data for the US, Ciccone and Hall (1996) and Rosenthal and Strange (2008a) for the years 1988 and 2000, respectively, report values for the elasticity of productivity with respect to density of approximately 4%-5%. For French cities, Combes et al. (2008) and then Combes et al. (2010) use individual data to estimate the effect of density on wages as a measure of productivity. They find an elasticity of wages with respect to density of approximately 3%. Using individual data for Italy, the UK and the Netherlands, Mion and Naticchioni (2009), D'Costa and Overman (2014), and Groot et al. (2014) obtain values of 1%, 1.6% and 2.1%, respectively. The previous studies are focused on static agglomeration effects, but more recent research is moving toward a dynamic framework (Glaeser and Maré, 2001). One of the most complete studies on the dynamic impact of agglomeration economies is De la Roca and Puga (2017). The authors attempt to disentangle the static urban wage premium from a dynamic urban wage premium using data for Spain and find an estimated dynamic elasticity of 5.1%, which is more than double the static elasticity of 2.2%.

economies play an important role in this regard, while urbanization economies are present but only play a marginal role. The results show that if employment in any sector in any region were to double, productivity as measured by value-added would increase by 11%. Lee and Zang (1998) report a comparable result in their study of the manufacturing industry in South Korea. The authors find that doubling the employment in a given sector and region is associated with a 7.9% increase in value-added per worker as a measurement of productivity. For Indian cities, the studies by Mills and Becker (1986), Becker et al. (1992), and Shukla (1996) show that equally significant increases in productivity are generated by urbanization.

Among the more recent studies, Da Mata et al. (2007) examine Brazilian cities and find that the urban elasticity, which measures urbanization economies as market potential, is 11%. Similarly, Combes et al. (2015), who study Chinese cities and instrument density using three variables: peripherality, the historical status of the city, and the distance to historical cities. They find that the elasticity of wages with respect to density is between 10% and 12%. For the case of India, Chauvin et al. (2014) evaluate the effect of density on individual annual earnings at the district level and find a large elasticity of approximately 9%-12%.

In the Colombian context, Duranton (2016) provides a comprehensive analysis of agglomeration effects for cities in Colombia in the period 1996-2012. In the paper, several dimensions of agglomeration economies are considered in the context of a developing country. In particular, the author analyzes the aspects of agglomeration that are associated with the complementarity between education and density as well as agglomeration effects by different types of workers, such as younger and informal workers. Duranton (2016) estimates wage equations where the city population is the measure of agglomeration economies. To correct for the reverse causation problem between wages and the agglomeration variable, the author implements a 2SLS strategy and instruments the current city population with long lags of population from 1843, 1870 and 1938. The main results show that the elasticity of wages with respect to the city population is 5%, and the result is robust to a several econometric specifications. Other interesting results include the absence of complementarity between city size and education and higher returns to city size for younger workers. The author explains these opposing findings to what would be predicted by the existence of greater learning effects in large cities by higher returns to agglomeration in the informal sector, where the workers are younger and less educated. This positive effect of agglomeration on wages in the informal sector is novel in the literature, which the author explains as possibly being related to the fact that informal workers may sell their products locally, where their incomes are more directly tied to local housing and transportation costs.

Three considerations must be taken into account in the results found by Duranton (2016) regarding the informal sector. First, the results may depend on the definition of informality used. Duranton (2016) defines informal workers as those workers who do not have a written labor contract. However, this definition overlaps with that of self-employed workers, who are not the target population of analysis. Second, having a written labor contract is positively correlated with wages and with city size, such that workers with a written labor contract may not be comparable across cities of different sizes. Third, the estimation of agglomeration effects that distinguishes between formal and informal workers is not corrected for endogeneity of the city population, such that the estimators are biased and inconsistent. Thus, the results should be treated with caution. In this study, we address these issues. First, we use alternative definitions of informality and test the sensitivity of the results to these different definitions. Second, in order to compare informal workers across cities of different sizes, we include occupation and economic sector variables in the regressions to control for current skills. Finally, all of our results are based on estimations using 2SLS techniques.

3. Theoretical framework

In this section, we present the theoretical model upon which we structure our empirical specification of wages to test the impact of agglomeration economies on local productivity as measured by nominal wages (Combes et al., 2008; Melo and Graham, 2009; Combes and Gobillon, 2015).

The profit of a competitive representative firm located in area a in year t is:

$$\pi_{at} = p_{at}Y_{at} - w_{at}L_{at} - r_{at}K_{at} \tag{1}$$

where Y_{at} is the output of the firm, which uses two inputs, labor L_{at} and other production factors K_{at}, such as land, capital, or intermediate inputs. p_{at} is the price of the good produced, w_{at} is the wage rate on the local labor market, and r_{at} is the unit cost of non-labor inputs.

Suppose that the production function of the firm is Cobb-Douglas and can be represented by the following equation:

$$Y_{at} = A_{at}(s_{at}L_{at})^{\alpha}K_{at}^{1-\alpha} \tag{2}$$

where $0 < \alpha < 1$ is a parameter, s_{at} denotes the local labor skills, and A_{at} is the local total factor productivity. In a competitive equilibrium, the first-order condition for the optimal use of labor is given by the following expression:

$$w_{at} = \alpha p_{at}A_{at}s_{at}^{\alpha}\left(\frac{K_{at}}{L_{at}}\right)^{1-\alpha} \tag{3}$$

Using the first-order condition for profit maximization with respect to the other factors, reordering in terms of K_{at}/L_{at}, and inserting it into equation (3), we obtain:

$$w_{at} = \alpha(1-\alpha)^{(1-\alpha/\alpha)}\left(p_{at}\frac{A_{at}}{r_{at}^{1-\alpha}}\right)^{1/\alpha} s_{at} = B_{at}s_{at} \tag{4}$$

From this last expression, we note that the local average nominal wage depends positively on labor skills, s_{at}, the output price, p_{at}, and the technological efficiency of the local economy, A_{at}. We can also observe that the local level of wages is negatively determined by the costs of the other non-labor input factors, r_{at}. The effects of agglomeration and dispersion forces work through these three factors. The output and input prices, p_{at} and r_{at}, capture a number of agglomeration mechanisms that operate through local markets, sometimes referred to as ‘pecuniary externalities’, while the local environmental efficiency, A_{at}, captures the effects from pure local externalities or ‘technological externalities’, that are not mediated by the market.

According to Combes and Gobillon (2015), there are two types of 'technological externalities'. On the one hand, firms and consumers are grouped together in cities where they share indivisible goods such as airports, libraries, museums, universities, and hospitals. In this situation, all the local market participants benefit from the infrastructure via a reduction in access costs to such public goods, as the costs are spread across all the beneficiaries. This generates a first type of agglomeration economy, where the local total factor productivity, A_{at}, is larger in larger cites due to the presence of local public goods, thus increasing the composite labor productivity effect, B_{at}, and therefore the local wages. The second type of 'technological externality' emerges when the spatial concentration induces local knowledge spillovers that make firms more productive. This type of mechanism again makes A_{at} large in large cities.

The previous discussion attempted to explain how city size generates agglomeration economies. However, observation additionally tells us that city size generates not only agglomeration economies but also dispersion forces. Typically, excess concentration in large cities can imply negative externalities due to congestion, such as longer commuting costs and scarce land for housing and plants. Congestion in local public goods implies a reduction of A_{at}, while scarce land can lead to costs of inputs that are not perfectly mobile, r_{at}, that are higher in large cities. These space constraints work as a dispersion force that has a negative effect on local wages (Tabuchi, 1998).

Note that the composite labor productivity effect, B_{at}, is affected by both the pure externalities, A_{at}, and the effects related to goods or input prices, $p_{at}$and r_{at}. Thus, with this approach, we can estimate the overall effect of local characteristics but not the exact channel through which agglomeration economies work. In other words, we cannot identify price and technology effects separately; we can only estimate the combined net overall effect of all three mechanisms: $\left(p_{at}\frac{A_{at}}{r_{at}^{1-\alpha}}\right)^{1/\alpha}$. Furthermore, note that the correlation between wages and density only shows the overall impact of both agglomeration economies and dispersion effects. While the net effect of spatial concentration can be identified, it is not possible to identify these effects separately (Duranton and Puga, 2004; Rosenthal and Strange, 2004; Combes et al., 2008).

4. Estimation strategy

To apply equation (4) to the data, we formulate the following wage equation that expresses the theoretical model above as the empirical model to be estimated:

$$\ln\, w_{i(t)} = \alpha_0 + \beta \ln density + X_{i(t)}\varphi + \pi_{oi(t)} + \sigma_{ei(t)} + \theta_{ai(t)} + \delta_t + \varepsilon_{i(t)} \quad (5)$$

where i identifies the worker, o refers to occupation, e refers to the economic sector, a identifies the region, and t specifies the time period. The "$i(t)$" subscripts indicate that the observations are an independent cross-sectional series where only N individuals are available in each period.

The dependent variable is the logarithm of the nominal hourly wage. The economics literature agrees that considering nominal wages is a good measure of workers' productivity, and in the case of bargaining or in the presence of externalities, for instance, workers' wages are likely to be higher in larger cities (Combes and Gobillon, 2015). Given that workers are mobile, the possible differences in real wages across cities should to some extent reflect differences in amenity value and not in productivity differentials across cities (Roback, 1982). For instance, Albouy (2008 and 2009) for the US and more recently Albouy et al. (2013) for Canadian cities find that real wages are correlated with arts and climate cities, coastal proximity, sunshine, and mild seasons. Additionally, Duranton (2015) and De la Roca and Puga (2017) argue that wages should be measured in nominal instead of real terms because the former reflect differences in productivity across places, while the latter measure differences in quality of life levels. Glaeser (2008) offers more details on this point by using the spatial equilibrium approach.

Our measure of urban agglomeration, $\ln density$, is the logarithm of the municipality's employment density, which is defined as the number of workers per square kilometer in each municipality using the average over the 2008-2014 period.[2] The basic idea behind this variable is that high density is a potential source of increasing returns resulting from stronger knowledge and technological spillovers in areas of dense economic activity.

As mentioned in the discussion of the theoretical framework, this measure of agglomeration economies can estimate various factors at the same time. First, this variable reflects the quality of urban life, which is expressed in higher urban rents. According to Roback (1982), the quality of urban life is positively correlated with higher wages in cities. Thus, the coefficient on employment density should be positive. Second, la variable can measure the level of negative amenities or disamenities in denser areas, including congestion, pollution, and noise. Disamenities would make working in a denser area unpleasant, which could be expected to be compensated by a higher wage (Borjas, 2008; Lee, 2016), causing the coefficient of employment density to be positive. Finally, a denser area can imply a plentiful labor supply that could decrease wages, which would make the coefficient negative. In this sense, Rosenthal and Strange (2008a) show that a negative sign could occur if there is a limited amount of job creation for a certain type of employment, which might mean that when there are more workers of a certain type, each worker would earn a lower wage. If the abovementioned three factors coincide, it would be difficult to determine the sign of the coefficient on employment density, indicating that the relationship should be tested empirically.

It is important to note that we use the municipality as the spatial unit of analysis. Although it is not an ideal unit, it is the best available approximation of a self-contained labor market in Colombia. Municipalities are areas where a high proportion of people who live (work) in the area also work (live). As Dominicis et al. (2007) argue, if there is evidence of a concentration of residential activities, of work activities, and of social relationships that are created within

[2] Following Combes and Gobillon (2015), we prefer to use employment rather than population because employment better reflects the magnitude of local economic activity. In addition, the results using population are generally very similar to those obtained with employment. The magnitudes of the coefficients on the logarithm of population density for an analogous estimation of Column 3 in Table 4 are as follows: for the total sample, 2.4% (statistically significant at the 5% level); for the formal sector, -0.5% (not statistically significant at the 5% level); and for the informal sector, 2.2% (statistically significant at the 5% level). To save space, full results using population variables are not presented here but are available upon request.

an area, then the area can be considered a self-contained labor market or a Local Labor System.[3] According to the 2005 Colombian census, which is the most recent census available, 89% of workers are employed in the municipality where they live, 8% work in a different municipality from where they live, 0.2% work in another country, and the remaining 2.8% do not know or did not answer.[4] Another possibility is to use metropolitan areas as the spatial unit of analysis; however, as we will see, the results are very similar when we use metropolitan areas instead of municipalities (a similar conclusion is reached by Duranton (2016)) because for many cities but particularly the smallest, the municipality and the metropolitan areas coincide, whereas only the few largest, form large metropolitan areas that comprise several municipalities.

The vector $X_{i(t)}$ contains the variables that measure a standard set of demographic attributes such as the worker's level of education, gender, age and its square, and years in the current job and its square. In addition, in our model, we included sets of dummy variables to control for several sources of heterogeneity that can lead to omitted variable bias and inconsistency in the model parameter estimates. To capture macro-level changes in wage rates that are common to all individuals, we included time dummies, δ_t. Similarly, to control for current skills, we added a set of occupation dummy variables, $\pi_{oi(t)}$. We also included a set of dummy variables to control for economic sector heterogeneity and regional characteristics; these are represented by $\sigma_{si(t)}$ and $\theta_{ai(t)}$, respectively.

To identify the agglomeration effects by job type, we split the sample into the formal sector and the informal sector, and we estimate equation (5) for each sector. We define informal workers as those workers who are not covered by health insurance and the pension system. Another aspect to be considered in the estimation is the endogeneity bias caused by the reverse causality between wages and agglomeration. Wages can increase due to higher employment density, but higher wages may also attract more people and firms to a given area. To avoid endogeneity bias, we implemented instrumental variable (IV) techniques. In the literature, long-lagged values of endogenous variables have been widely used as instruments since Ciccone and Hall's (1996) pioneering work. The basic idea behind these instrumental variables is that deep time lags of urban density can to some extent explain the distribution of present densities, but they do not explain the distribution of current urban productivity levels.

To construct our instruments for current density following Ciccone and Hall (1996), we use population data collected from the 1912, 1918 and 1928 censuses. Although national censuses had taken place in Colombia prior to these, we prefer to use censuses from the early 1900s because most of the current municipalities were created at the end of the 1800s and the beginning of the 1900s. As such, we have complete information for the past populations of 390 municipalities.

[3] As Openshaw and Taylor (1979) note, municipalities or metropolitan areas are much more related to the concept of local labor markets than the usual administrative areas, so they are a good option for overcoming the Modifiable Areal Unit Problem (MAUP).

[4] We calculated these percentages using information from the IPUMS-International database (https://international.ipums.org/international/)

We have to take into account that to yield unbiased estimates of the effect of density on wages using instrumental variables, our instruments must satisfy two conditions to be valid: relevance and exogeneity. While the first condition demands that our instruments be correlated with the contemporaneous employment density, the second requires that our instrument be uncorrelated with the error term $\varepsilon_{i(t)}$. As mentioned by Combes and Gobillon (2015), it is possible to imagine a number of possible violations caused by alternative links between past populations and current wages, such as permanent local characteristics that may have affected past location choices and continue to affect local productivity today. Such characteristics include the centrality of the location in the country, a suitable climate, or geographic features such as access to the coast or the presence of a large river. To minimize these potential problems, we control for geographic characteristics in regressions and try to preclude such correlations. The details of the relevance and exogeneity tests for the instrumental variables are presented in Section 7.

5. Data and variables

Some studies of agglomeration economies use detailed spatial data on panels of workers or firms (see, for example, Combes et al., (2010) and Glaeser and Maré (2001)), which allows greater administrative scale analysis and the ability to control for unobserved individual characteristics that may be correlated with location choices. Unfortunately, these types of data are not available in Colombia or, generally, for most developing countries. Instead, we use a cross-sectional survey, the Colombian Great Integrated Household Survey (GEIH), which is carried out by the National Administrative Statistics Department (DANE). By using cross-sectional data, it is not possible to control for all the characteristics of individuals that shape their skills that do not change over time, the effect of which can be considered constant over time (Combes and Gobillon, 2015). However, there are various measures of observed skills that can be used at the cost of not controlling for unobservable individual characteristics. For instance, Duranton and Monastiriotis (2002) and Wheaton and Lewis (2002) use measures such as diplomas or years of education. Another measure that has been used is the socio-professional category "occupation", which captures the exact job carried out by workers and part of the effects of their past career. As such, occupation can be considered a measure that should be more highly correlated with current skills than education. Given that the GEIH gathers detailed information about populations' general characteristics (gender, age, year of education, and municipality of residence) and their employment conditions (whether they work, what they do, how much they earn, number of hours worked and whether they have social security for health care), we included education and occupation as measures of workers' current skills. These types of measures are often recorded in labor force surveys, and they allow greater comparability across developing countries.[5]

We analyzed the period between 2008 and 2014. The databases for earlier years are not comparable because several methodological changes were made by DANE in 2007. After excluding individuals with no labor income, those who did not report their municipality of residence, and the 1% of workers with the lowest and highest wages each year, we had

[5] Given the confidentiality of the data at the municipal level, all the estimations in this paper were conducted following DANE's microdata-access policy, which implies working in situ under the supervision of DANE's staff and with blinded access to sensitive information.

1,920,678 observations, with a mean of 270,000 observations per year, and information for 568 municipalities.[6]

We divide the workers into formal and informal workers. As noted above, informal workers are defined as individuals who do not have access to the social security system to receive healthcare and a retirement pension. Note that this definition has been widely used in prior research, including Perry et al. (2007), Jütting and De Laiglesia (2009), and García (2017), among many others. Following this definition of informality, we can observe in Table 1 that approximately 60% of the employees in Colombia are informal workers and that informal work is a persistent phenomenon.

We also controlled for a standard set of demographic attributes in the models. These include the workers' level of education, gender, age, years in the current job, occupation (10 indicators), economic sector (8 indicators), and regional and geographic variables (five regional indicators: Central, Eastern, Western, Caribbean, and Orinoco[7]; and three geographic variables: water availability, soil erosion, and altitude and its square). We also included a measure of market access, namely, the distance in kilometers to the capital city of the department. In Tables 1 and 2, we show some descriptive statistics of these variables, which were calculated using person sampling weights from GEIH to ensure that the estimates are representative.

[6] Colombia covers an area of roughly 1,200,000 km2 and is divided into 32 administrative units called departments and a Capital District that is the country's capital, Bogotá. Departments are country subdivisions similar to US states and are granted a certain amount of autonomy. Each department is composed of municipalities, among which there is a capital city of the department. In total, Colombia has 1,119 municipalities (a more detailed characterization of Colombia can be found in Royuela and García (2015) and Nicodemo and García (2015)).

[7] Departments and municipalities are grouped into very broad regions according to their economic, demographic and social conditions as well as their location. These regions are composed of the following departments (capital city in parentheses): Central region: Bogotá, Antioquia (Medellín), Caldas (Manizales), Quindío (Armenia), Risaralda (Pereira), and Tolima (Ibagué); Eastern region: Boyacá (Tunja), Norte de Santander (Cúcuta), and Santander (Bucaramanga); Western region: Cauca (Popayán), Choco (Quibdó), Huila (Neiva), Nariño (Pasto), and Valle (Cali); Caribbean region: Atlántico (Barranquilla), Bolivar (Cartagena), Cesar (Valledupar), Córdoba (Montería), La Guajira (Riohacha), Magdalena (Santa Marta), Sucre (Sincelejo); Orinoco region: Caquetá (Florencia) and Meta (Villavicencio). A more detailed description of these regions can be found in Galvis (2001), Barón (2002), and García (2017).

Table 1. Summary statistics at individual level

	2008		2011		2014		All years	
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
Hourly wage $	1.52	2.04	1.96	2.38	1.83	2.01	1.76	2.16
Male	0.57	0.49	0.56	0.50	0.55	0.50	0.56	0.50
Age	37.99	12.81	38.31	13.18	38.54	13.43	38.29	13.15
Years of education	9.53	4.56	8.93	5.26	9.96	4.52	9.56	4.65
Education by levels:								
Primary school	0.26	0.44	0.26	0.44	0.23	0.42	0.25	0.43
Middle school	0.19	0.39	0.16	0.37	0.18	0.38	0.18	0.38
High school	0.33	0.47	0.36	0.48	0.34	0.47	0.34	0.47
Technical or technological	0.08	0.27	0.08	0.27	0.14	0.34	0.10	0.30
University	0.13	0.33	0.13	0.34	0.12	0.32	0.12	0.32
Years in the current job	6.63	8.78	6.31	8.26	6.04	8.06	6.30	8.79
Informality	0.63	0.48	0.63	0.48	0.58	0.49	0.61	0.48
Region								
Central	0.51	0.50	0.51	0.50	0.53	0.50	0.52	0.50
Eastern	0.11	0.31	0.11	0.31	0.11	0.31	0.11	0.31
Western	0.17	0.38	0.17	0.37	0.17	0.37	0.17	0.37
Caribbean	0.18	0.38	0.18	0.38	0.17	0.37	0.18	0.38
Orinoco	0.03	0.16	0.02	0.16	0.03	0.16	0.02	0.16
Occupation								
Professional	0.05	0.22	0.06	0.23	0.07	0.25	0.06	0.23
Managers	0.11	0.31	0.11	0.31	0.10	0.30	0.11	0.31
White collar	0.07	0.26	0.06	0.24	0.06	0.24	0.07	0.25
Low white collar	0.09	0.28	0.09	0.28	0.09	0.29	0.09	0.28
Sales employees	0.01	0.12	0.01	0.12	0.01	0.12	0.01	0.12
Blue collar	0.20	0.40	0.19	0.39	0.18	0.38	0.19	0.39
Low blue collar	0.06	0.24	0.07	0.25	0.07	0.26	0.07	0.25
Skilled service workers	0.08	0.27	0.09	0.28	0.10	0.30	0.09	0.28
Unskilled service workers	0.27	0.45	0.27	0.45	0.27	0.44	0.28	0.45
Agricultural workers	0.04	0.20	0.04	0.20	0.04	0.19	0.04	0.20
Sector								
Agriculture	0.05	0.22	0.05	0.22	0.04	0.20	0.05	0.22
Industry	0.16	0.37	0.15	0.36	0.14	0.35	0.15	0.36
Building	0.06	0.23	0.07	0.25	0.07	0.25	0.07	0.25
Commerce and hotel	0.29	0.45	0.29	0.46	0.30	0.45	0.29	0.45
Transport and tel	0.10	0.30	0.10	0.30	0.10	0.30	0.09	0.29
Financial	0.09	0.29	0.10	0.29	0.10	0.31	0.10	0.30
Adm. Pub	0.09	0.29	0.09	0.28	0.09	0.28	0.09	0.28
Service	0.15	0.36	0.15	0.35	0.15	0.36	0.15	0.36

Note: All data are weighted using person sampling weights from GEIH to be representative.

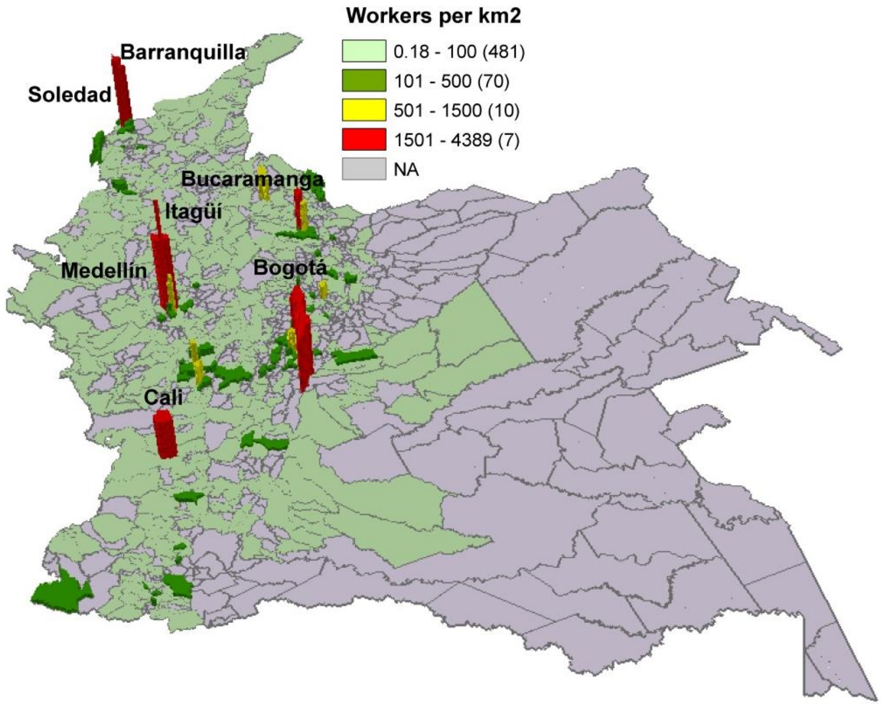
Table 2. Summary statistics at municipal level

	Municipalities	p25	Median	p75	Mean	Std. Dev	Min	Max
Number of workers	568	5,200	8,879	13,905	26,574	156,488	341	3,399,830
Municipal area (Km2)	568	161.3	358.6	842.2	765.9	1,529.6	15.4	17,641.70
Density (workers per Km2)	568	10.6	25.5	57.0	94.8	326.3	0.176	4388.98
Altitude (m.a.s.l.)	568	123.0	1,026.5	1,786.5	1,098.0	923.9	2	3,087
Distance to the capital in Km	537	45.5	87.1	137.0	104.2	81.4	0	452

Note: Number of workers using an average over the period 2008-2014. The distance to the capital city of the department in kilometers is calculated using Google Map.

For all of the models, we used the logarithm of nominal hourly wages as the dependent variable. As mentioned above, our measure of urbanization is the logarithm of the employment density of each municipality. The municipalities have an average of approximately 26,000 workers, ranging from 341 workers to over three million. Figure 1 shows the employment density by municipality, and we can see that Bogotá, Medellín, Itagüí, Cali, Bucaramanga, Barranquilla, and Soledad have the highest levels of urbanization. Itagüí is the densest city in Colombia, with just over 4,500 workers per km2.[8]

Figure 1. Employment density by municipality in Colombia



Note: Average employment density between 2008 and 2014.

6. Documenting the agglomeration-wages relationship in the presence of an informal sector

We begin with an illustration that stresses the topics included in this paper. Table 3 shows the average hourly wages earned by formal and informal employees for the three largest municipalities and for municipalities with low employment density, that is, with a density of between 70 and 100 workers per km2. We can observe that there is a clear relationship between wages and agglomeration. For formal employees, average wages are similar for the two groups of cities. In contrast, we can deduce from the data in the table that for all years, informal workers earned substantially higher wages in the larger cities. Taken as a whole, Table 3 suggests a positive relationship between agglomeration and wages for informal workers, but this relationship is weak for formal workers.

[8] These same cities also have the highest populations, and Itagüí is the most densely populated city in Colombia, with 12,114 people per km2 in 2014.

Table 3. Average wages between formal and informal employees in select municipalities

Sector	Municipalities	Hourly wages ($)			
		2008	2011	2014	All Years
Formal	Bogotá, Medellín, Cali	2.13	2.75	2.47	2.45
	Less dense cities	2.02	2.94	2.48	2.58
Informal	Bogotá, Medellín, Cali	1.49	1.82	1.63	1.63
	Less dense cities	1.08	1.32	1.21	1.18

Note: Less dense cities are those cities with an employment density between 70 and 100 workers per Km2. All data are weighted using person sampling weights from GEIH to be representative. All differences in means between groups of municipalities for the formal sector and the informal sector are significant at 1%.

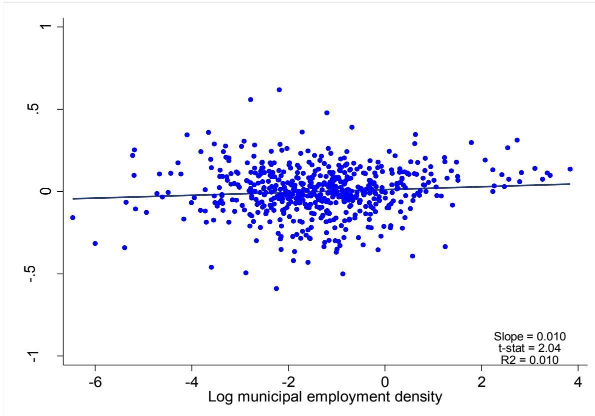
To confirm these relationships between agglomeration and wages by sector, we plotted the logarithm of wages against employment density for 568 municipalities for the formal sector and the informal sector separately. Figure 2(a) shows that for the total sample, the slope of the regression line between the logarithm of wages and the logarithm of density, which measures the density elasticity of wages, is 1%. Regarding the formal and informal sectors, Figures 2(b) and (c) show that while for the formal sector, the density elasticity of wages is -4.4%, for the informal sector, this elasticity is 1.3%. This result confirms our previous results of lower agglomeration returns for formal workers than for informal workers.

These results are somewhat surprising because formal workers are more educated than informal workers and might have a greater ability to learn from nearby human capital. Furthermore, formal workers work in medium-large enterprises that can obtain greater benefits from labor market pooling and input sharing associated with agglomeration (Rosenthal and Strange, 2008b). On the other hand, informal workers are characterized by a limited education, and they tend to work in very small enterprises (Perry et al., 2007; Jütting and De Laiglesia, 2009; García, 2017; García and Badillo, 2017), which might imply that the workers are less able to absorb new knowledge, while the activities of small enterprises tend to be geared toward small local markets more than toward generating input-output linkages (Moreno-Monroy and García, 2016).

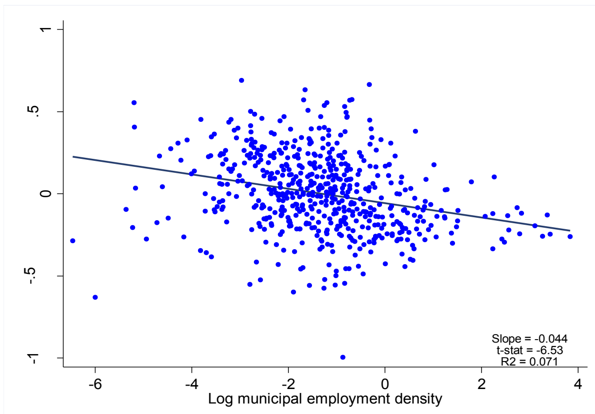
According to Duranton (2016), this greater agglomeration effect in the informal sector may be due to workers in this sector obtaining higher benefits from the local market, and therefore, their incomes are more influenced by local housing and transportation costs. Another possible explanation could be that given that the creation of formal jobs in the economy is limited, having more formal workers might tend to result in each worker earning lower wages. This type of work-spreading would imply the opposite sign on employment density (Rosenthal and Strange, 2008a). The possibility that workers might concentrate in equilibrium in this manner, is consistent with the Harris-Todaro (1970) model, which shows that in a context of industrialization in a developing country, when the urban wage is fixed above the market-clearing level, it can lead to unemployment in equilibrium, where unemployment goes undercover in the informal sector.

Figure 2. Employment density and wages in Colombia

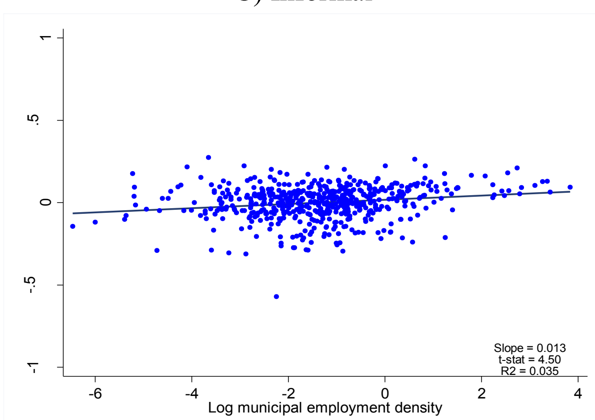
a) Total



b)Formal



C) Informal



Note: The vertical axis represents log municipal hourly wages computed using 2008-2014 wage data after controlling for years effects, individual characteristics, occupation and economic sector. The horizontal axis represents log of average employment density between 2008 and 2014. There are 568 municipalities. All variables are centered around their mean.

7. Results

In this section, we present the results of the estimation of the wage equation by 2SLS, which are reported in Table 4. Tables A1 and A2 in the Appendix report the results of the estimations by OLS for the total sample and after dividing the sample into formal and informal workers, respectively. To simplify the presentation, only the coefficients on the elasticity of wages with respect to density are provided. We begin by discussing the instrument diagnostic test reported at the bottom of the table.

Regarding the exogeneity condition of the instruments, we used Hansen's J test (1982) to test the null hypothesis of exogeneity of the long-lagged instruments. The results for instrument exogeneity for all models coincide with previous studies using similar instruments: the null hypothesis of exogeneity is not rejected at a 5 percent level of significance, suggesting that the instruments are exogenous.

With regard to the relevance of the instruments, the first stage of the regression results indicates that the instruments for city density have considerable explanatory power. The explanatory power is tested using Shea's (1997) partial R-squared, and we found values that range between 0.5 and 0.8 in our regressions. To further examine the relevance of the instruments, we carried out the Kleibergen-Paap test of under-identification, which tests whether the model is identified, where identification requires that the excluded instruments be correlated with the endogenous regressor. When the instruments are uncorrelated with the endogenous regressor, the matrix of reduced-form coefficients is not of full rank, and the model will be unidentified. Since we allow intra-group correlation, the relevant statistic in this case is the Kleibergen and Paap (2006) rank LM statistics. If we fail to reject the null hypothesis that the matrix of reduced-form coefficients is under-identified, it means that the instrument variable bias of the parameter estimates will be increased. The values presented in Table 4 for all models show that the tests reject the null hypothesis of under-identification at a 5 percent level of significance, implying that the instruments are relevant.

Nonetheless, a rejection result for the null hypothesis in the Kleibergen-Paap test should be treated with caution because weak instrument problems may still be present. Weak identification arises when the instruments are correlated with the endogenous regressor, but only weakly. As noted by Murray (2006) and Stock and Yogo (2005), when the instruments are poorly correlated with the endogenous regressors, the estimates from the instrumental variable model will be biased. In this case, and allowing intra-group correlation, the relevant test is the Kleibergen-Paap (2006) rank Wald F statistic, which involves testing the significance of the excluded instruments in the structural equation, resulting in the substitution of the reduced-form expression for the endogenous regressor in the main equation of the model (Baum et al., 2007; Davidson and MacKinnon, 2010). The critical values for this test are derived from Stock and Yogo (2005). The results reveal that the Kleibergen-Paap (2006) rank Wald F statistic is higher than the Stock and Yogo (2005) critical values, suggesting that our instruments are not weak.

Table 4. Agglomeration effects and informality (2SLS)

Dependent variable: log hourly wage

	(1) Only emp. density		(2) Individual characteristics		(3) Sector-Occupation		(4) Geographic variables		(5) Market access		(6) Non-lineal		(7) Education effects	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Log employment density	-0.038***	0.027***	0.010	0.022**	-0.004	0.021**	-0.003	0.021**	-0.005	0.020**	0.031	-0.082	-0.016	0.008
	(0.0111)	(0.0116)	(0.0080)	(0.0106)	(0.0067)	(0.0107)	(0.0061)	(0.0096)	(0.0069)	(0.0100)	(0.0678)	(0.1419)	(0.0108)	(0.0133)
Log employment density^2											0.002	-0.006		
											(0.0038)	(0.0080)		
Log dist. Km to capital city									0.008	0.022				
									(0.0366)	(0.0691)				
Log dist. Km to capital city^2									-0.003	-0.009				
									(0.0064)	(0.0117)				
Educ x Log emp density													0.001	0.001
													(0.0006)	(0.0007)
Observations	645,758	1,174,248	644,583	1,169,978	644,144	1,169,338	640,255	1,165,021	643,312	1,166,400	644,144	1,169,338	645,269	1,173,463
Municipalities	388	390	388	390	388	390	383	385	378	380	388	390	388	390
R2	0.021	0.024	0.501	0.208	0.562	0.237	0.566	0.248	0.562	0.238	0.562	0.237	0.543	0.233
Instruments exogeneity														
Hansen J statistic	4.910	1.445	3.610	1.345	1.717	0.996	6.140	5.677	2.596	0.299	6.590	8.606	1.706	0.820
Chi-sq P-val	0.086	0.486	0.164	0.511	0.424	0.608	0.051	0.059	0.273	0.861	0.159	0.072	0.426	0.663
Instruments relevance														
1. First-stage statistics														
Shea partial R2														
Log employment density	0.793	0.804	0.793	0.803	0.786	0.803	0.754	0.786	0.801	0.824	0.504	0.502	0.762	0.778
Log employment density^2											0.508	0.497		
Educ x Log emp density													0.769	0.782
2. Under-identification test														
Kleibergen-Paap rk LM stat	15.12	14.83	15.53	14.91	16.174	14.99	13.47	12.98	14.87	13.88	12.18	12.47	16.30	15.15
Chi-sq P-val	0.002	0.002	0.001	0.002	0.001	0.002	0.004	0.005	0.002	0.003	0.032	0.029	0.001	0.002
3. Weak identification test														
Kleibergen-Paap rk Wald F stat	53.62	71.34	55.76	72.13	57.21	73.78	63.70	94.71	55.73	71.71	8.767	10.92	36.73	45.57

Note: Robust standard errors clustered at the municipality level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

This table replicates Table A2 in the Appendix using 1912, 1918 and 1928 populations as instrument for contemporaneous working population to calculate the log density variable in all columns. The square of these instruments are used in column 6. In columns 7 we use the average of population in 1912, 1918 and 1928 in the calculation of the product of education and log density variable.

Consider now the estimates of the impact of agglomeration on wages. We start by analyzing the results without distinguishing between the formal and informal sector, as reported in Table A1 in the Appendix. The specification without any other control in Column 1 shows an elasticity of wages with respect to density of 4.4%. When we add the control variables to this estimation (Columns 2 and 3), there is a reduction in elasticity to a value of approximately 3.7%. In Table A3 in the Appendix, the estimations are corrected for the endogeneity problem, and the results show that the elasticity when the control variables are added is approximately 2%. These values are lower than the 5% found by Duranton (2016) for the period from 1996 to 2012 using a similar database for Colombia. This difference may be due to potential measurement errors in the labor variables calculated by Duranton (2016) when the author compared the GEIH before and after 2007, the year in which there were several methodological changes, implying that the surveys could not be compared (DANE, 2009).[9]

When we estimate the model using the separate samples of formal and informal workers separately[10], the results of the elasticity of wages with respect to density without any other control in Column 1 in Table 4 show that in the formal sector, the elasticity is -3.8% and is significant, and for the informal sector, the elasticity is 2.7% and is highly significant. When we add individual characteristics, occupation, and economic sector as control variables in Columns 3 and 4, we can observe that the elasticity of the formal sector is not statistically significantly different from zero, while the elasticity among informal workers shows a slight decrease to 2.2%, indicating that the productivity of informal workers in a city twice as dense is approximately 1.5% greater.[11] This difference between formal and informal workers echoes the summary measure in Table 3 and Figure 2 and persists throughout the paper.

So far, the results show that the density of city employment has a significantly positive effect on productivity in the informal sector but little effect in the formal sector. Comparing a small municipality with a density of 20 workers per km2 to Bogotá with a density of approximately 2000 workers per km2, the agglomeration effect in the informal sector suggests that informal workers in denser cities will earn approximately 11% more than those in less dense cities. In the case of the formal sector, the non-significance of the city density elasticity of wages indicates the low level of cluster benefits in this sector. These results suggest that although there are important agglomeration economies in the informal sector, the benefits could be curtailed due to the negative effects of having a large informal sector in the economy. Informality may generate diseconomies of agglomeration that undermine the benefits associated with the urban scale in the economy in general and for those that can reach the formal sector in particular (Harris, 2014). As noted by Annez and Buckely (2009), informality may crowd out agglomeration benefits because the informal sector is an unproductive sector that consists of small-scale producers, and it increases the costs for the

[9] One of the main methodological changes that affected the calculations of variables such as wages, employment, and labor participation was an increase in the size and structure of the sample, where the statistical representativeness for large cities changed to an increase from 13 to 23 cities in 2007.

[10] It is important to highlight that with the instruments, the number of municipalities analyzed is reduced from 568 to 390. This means that the youngest municipalities and, therefore, the smaller cities are not considered in this part of the analysis.

[11] We followed the formula developed by Combes and Gobillon (2015): $2^{\beta} - 1$, where β is the elasticity of productivity with respect to density.

formal sector through unfair competition (e.g., selling at very low prices that merely support subsistence). Also, a large informal sector may imply low institutional levels of the labor market because most jobs tend to be filled in an informal way through relatives, friends and social connections. This limits mobility across cities and discourages more productive workers from moving to better jobs, thus limiting urban scale effects on productivity. Additionally, a rationing of formal sector jobs in the economy may imply low incentives for workers to improve their skills locally, which limits the scope of agglomeration benefits (Duranton, 2015).[12]

Regarding the other columns in Table 4, we can note that including geographic controls (Column 4), such as regional indicators, water availability, soil erosion, and altitude as well as a variable for market access (Column 5), measured as the distance in kilometers to the capital city of the department, yields no change to the coefficients on city density for the formal and informal sectors in Column 3.[13] We can also observe that including geographic controls does not substantially increase the explanatory power of the regression; in fact, although these results are not reported, the coefficients on several geographic controls are not statistically significant. On the other hand, we found that wages are lower in the Caribbean region of Colombia than they are in the Central region in both the formal and informal sectors, while there are no wage differences between the rest of the regions and the Central region.

We now turn to analyze possible heterogeneities in the agglomeration effects. Column 6 attempts to detect non-linearities by adding the square of the logarithm of density as an independent variable to the specification of Column 3. The results show that the coefficients on the logarithm of density and the quadratic term are not statistically significantly different from zero for either the formal or informal sector, which suggests an absence of non-linearity in the agglomeration effects. This result is consistent with those found by Duranton (2016) for Colombia using data between 1996 and 2012.

Finally, Column 7 adds the product of the worker's number of years of education and the logarithm of density in order to determine whether there are differences in agglomeration benefits across skilled workers, that is, whether all workers benefit equally from the urban scale. The results show that the coefficients on the interaction term are not statistically significantly different from zero for either the formal or informal sector, which does not corroborate the hypothesis of higher returns to cities for more educated workers (Wheeler, 2001; Rosenthal and Strange, 2008a; Bacolod et al., 2009; Glaeser and Resseger, 2010). On the one hand, these results could be due to the better amenities in larger cities, which lead higher educated workers to locate in these cities, therefore decreasing the returns to education there. On the other hand, there is a complementarity between city density and individual skills that is an important factor in explaining the over-representation of more highly skilled workers in large cities. In large cities, there are urban amenities that are used and enjoyed

---

[12] García and Badillo (2017), using data from Colombia, confirm that formal job rationing does exist and that it affects approximately 62% of the workers who find themselves in the informal labor market.

[13] We also used time distance to the capital city of the department (calculated using Google Maps) as a measure of market access. The results are very similar to those using distance in kilometers. The results using the time distance variable are not presented here to save space but are available upon request.

more by more educated workers. Nevertheless, although this over-representation can occur in Colombia, the large cities in this country and in many large cities in developing countries have important urban disamenities, such as pollution, traffic congestion, crime, and excess garbage, to which more educated workers could be more sensitive, thus limiting the agglomeration benefits for this group. These results corroborate Duranton´s (2016) findings.

8. Robustness Checks

To assess the robustness of these results, Table 5 shows the results when we experiment with different city samples and types of workers, when we usw an alternative definition of informality, and when we employ different estimation techniques. Table 5 uses the same specification as Column 3 of Table 4.

The presence of larger cities in the sample could affect the results because there may be measurement errors at the top end of the labor income distribution, meaning that it is possible that the highest wages are not measured in the formal sector in the largest cities, which could lead to an observation of lower wages in the formal sector in the densest cities. Column 1 in Table 5 shows the results of a regression that excludes observations corresponding to Bogotá, the capital of Colombia[14], while Column 2 shows the results excluding observations from the five densest cities in Colombia, namely, Itagüí, Medellín, Barranquilla, Soledad, and Bogotá. Accordingly, the elasticities are slightly lower than when the capital city is included: for the informal sector, the elasticity is equal to 1.6%. When the densest cities are excluded, we observe that there are no significant changes in the elasticities.

To evaluate the sensitivity of the results when we consider only the main cities in Colombia, Column 3 uses the information for only the 23 main municipalities and metropolitan areas. We grouped the municipalities into metropolitan areas to determine whether there are changes when the spatial unit of analysis is modified. It is important to highlight that for small areas, the municipality and the metropolitan area coincide, whereas large metropolitan areas are formed from several municipalities. The results show that the elasticities are lower than when all municipalities are considered and they are not grouped into metropolitan areas, but the difference is not sizable: the elasticity of wages with respect to density in the informal sector is 1.7%. When we consider the all of the municipalities and group them into metropolitan areas, the results in Column 4 are very similar to those when we do not group the municipalities into metropolitan areas. These findings suggest that even in the main municipalities of Colombia and working with a spatial unit closer to urban areas, the results showing greater agglomeration benefits in the informal sector than the formal sector remain. Interestingly, the R^2 of the regression in Columns 3 and 4 is very similar to that in Column 3 in Table 4. This is perhaps an indication that there are no differences when working with municipalities or metropolitan areas in the case of Colombia. This result is consistent with the results found by Duranton (2016) for Colombia, where using metropolitan areas instead of municipalities implies similar elasticities of wages with respect to city population.

[14] According to Arango and Bonilla (2015) and García (2017), Bogotá accounts for approximately 17% of the country's total population, 25% of its total employment, and 25% of its GDP.

Table 5. Agglomeration effects and informality, robustness check

Dependent variable: log hourly wage

	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)		(9)	
	Excluding Bogotá		Excluding the densest cities		Only the 23 main municipalities and MA		MA and municipalities		Informality ILO		Excluding low-educated self-employed informal workers		Real hourly wages		GMM		LIML	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Log employment density	-0.007	0.016**	-0.009	0.023**	-0.003	0.017*	-0.003	0.019**	0.008	0.022**	-0.004	0.029**	-0.001	0.017**	-0.006	0.021*	-0.004	0.021**
	(0.0068)	(0.0078)	(0.0084)	(0.0116)	(0.097)	(0.0102)	(0.0067)	(0.0098)	(0.0081)	(0.0112)	(0.0067)	(0.0121)	(0.0105)	(0.0088)	(0.0066)	(0.0123)	(0.0067)	(0.0107)
Observations	590,396	1,111,789	516,560	1,009,116	603,364	1,059,188	644,144	1,169,338	764,444	1,049,038	644,144	446,607	603,364	1,059,188	644,144	1,169,338	644,144	1,169,338
Municipalities	387	389	383	385	40	40	374	376	390	390	388	388	40	40	388	390	388	390
R2	0.567	0.236	0.572	0.241	0.562	0.239	0.562	0.237	0.519	0.196	0.562	0.402	0.557	0.239	0.562	0.237	0.562	0.237
Instruments exogeneity																		
Hansen J statistic	0.792	1.516	0.091	1.995	3.130	0.485	1.770	1.293	0.272	0.848	1.717	1.439	4.115	1.247	1.717	0.996	1.717	0.996
Chi-sq P-val	0.673	0.468	0.955	0.369	0.209	0.785	0.413	0.524	0.873	0.654	0.424	0.487	0.128	0.536	0.424	0.608	0.424	0.608
Instruments relevance																		
1. First-stage statistics																		
Shea partial R2																		
Log employment density	0.805	0.814	0.779	0.777	0.723	0.781	0.691	0.734	0.794	0.801	0.786	0.795	0.821	0.859	0.786	0.803	0.786	0.803
2. Under-identification test																		
Kleibergen-Paap rk LM stat	14.94	14.03	13.84	13.93	13.52	12.84	14.56	12.76	15.56	15.29	16.17	14.83	13.92	12.70	16.17	14.99	16.17	14.99
Chi-sq P-val	0.002	0.003	0.003	0.003	0.004	0.005	0.002	0.005	0.001	0.002	0.001	0.002	0.003	0.005	0.001	0.002	0.001	0.002
3. Weak identification test																		
Kleibergen-Paap rk Wald F stat	59.09	83.11	71.06	82.13	49.93	65.76	47.71	72.54	58.66	74.69	57.20	62.84	54.69	74.30	57.21	73.78	57.21	73.78

Note: Robust standard errors clustered at the municipality level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

This table replicates column 3 of Table 4. Column 1 excludes observations corresponding to Bogotá and Column 2 excludes observations corresponding to the densest municipalities in terms of employment: Itagüí, Medellín, Barranquilla, Soledad and Bogotá. Column 3 aggregates municipalities into metropolitan areas (MA) and only uses information to 7 MA (Medellín-Valle de Abúrra, Cali-Yumbo, Barranquilla-Soledad, Bucaramanga-Girón-Piedecuesta-Floridablanca, Manizalez-Villa María, Pereira-Dos Quebradas-La Virginia, and Cúcuta-Villas del Rosario-Los Patios-El Zulia) and 16 municipalities, which represent the 23 main municipalities and MA in Colombia. Column 4 aggregates municipalities into MA (7) and uses the rest of municipalities. Column 5 uses the informality definition according to the International Labor Organization (ILO). Column 6 excludes low-educated (less than secondary education) self-employed informal workers. Column 7 reports the results using log real hourly wage as a dependent variable, adjusting hourly wages for price level using consumer price index (base year 2008) for each city as a deflator. 2SLS in Columns 1-7, GMM in Column 8, and LIML in Column 9.

We now test whether the results are sensitive to alternative definitions of informality. In the literature, at least two definitions are regularly used: the legalistic definition and the productivity definition. The former is the definition we follow in this paper, while the productive approach is based on the type of job and firm size. In Column 5, we use the productivity definition proposed by the International Labor Organization (ILO), which includes as informal workers all own-account workers (excluding administrative workers, professionals and technicians), unpaid family workers, and employers and employees working in establishments with fewer than five people. We observe that the results for the informal workers remain the same, with an elasticity of 2.2%, while in the formal sector, the elasticity is not statistically significantly different from zero. It is important to highlight that the productivity definition of informality underestimates the population compared to the legalistic definition because it does not include the possible presence of informal employment within large firms.[15] Hence, it is possible to state that under either a narrower (productivity) or broader (legalistic) definition of informality, the positive agglomeration in the informal sector remains and is still higher than that in the formal sector.

To consider the fact that there is a high proportion of workers in the informal sector who are self-employed[16] and therefore their wages may not reflect the marginal productivity of labor, we re-estimated the model while excluding low-educated self-employed, informal workers.[17] We verify the main results of the positive effect of agglomeration economies on wages in the informal sector and no effect in the formal sector. We note that the estimated elasticity of wages to employment density in the informal sector is higher than in Column 3 in Table 4. This result suggests that the higher agglomeration effects tend to be due to the nature of the occupation and the type of activity. By focusing on salaried informal workers, we also focus on more skilled workers, for whom the agglomeration effects should be stronger.

As mentioned above, the relevant dependent variable in the models is the logarithm of the nominal hourly wage because it is more appropriate to measure differences in productivity across cities rather than measuring the variable in real terms, which is more related to differences in the standard of living among cities. However, to check our results against differences in regional price disparities, where we expect to find a higher price level in densely populated cities, and to make the results comparable to research that uses real wages (see, for instance, Wheeler (2006), Melo and Graham (2009), and Melo et al. (2017)), we re-estimated the models using the logarithm of the real hourly wage as the dependent variable. The results are shown in Column 7. Note that when using real wages, some cities are excluded from the sample because price data were unavailable; in fact, the consumer price index (base year 2008) that was used to deflate the wages is only available for the 23 main metropolitan areas, which leaves us with the sample in Column 3. The estimated elasticity is 1.7%, which is similar to the elasticity found in Column 3 using nominal wages and the same sample of cities.

[15] We calculated the informality rate using the productivity definition and found that it is approximately six percentage points lower than the rate calculated based on the legalistic definition, which is consistent with the results of previous studies that compare different measures of informality in Colombia (see, Bernal, 2009; Galvis, 2012; García, 2017).
[16] On average, 67% of informal workers are self-employed.
[17] Low-educated workers are those workers with less than a secondary school education, that is, we excluded professionals and technicians. Low-educated self-employed informal workers represent 59% of the total informal workers in our sample.

Finally, Columns 8 and 9 estimate the regression using the Generalized Method of Moments (GMM) and with Limited Information Maximum Likelihood (LIML) instead of 2SLS. The results show that the coefficients on employment density are again similar to those found in Table 4.

In general, it appears that the instrumented coefficients on employment density by sector are generally statistically similar to our baseline 2SLS: they are not statistically significantly different from zero in the formal sector, and they are positive and significant in the informal sector. These results are robust to alternate samples of cities, type of employment, and estimation technique.

9. Conclusions

This paper sheds light on evidence regarding the relationship between agglomeration economies and productivity in a developing country, Colombia, which has a large informal sector. Among the main results, we have found that the density of local employment has a significantly positive effect on productivity in the informal sector, while there is little effect on the formal sector. The elasticity of wages with respect to employment density in the informal sector is approximately 2%, which in quantitative terms implies that moving from a city with a density of approximately 20 workers per km2 to Bogota—with an employment density of approximately 2000 workers per km2—is associated with approximately 9% higher wages in the informal sector. This paper therefore provides empirical evidence that supports the idea that there are positive agglomeration returns in the informal sector and that such returns are higher than those achieved in the formal sector.

The limited agglomeration returns in the formal sector may be due to the harmful effects of a large informal sector in the economy. On the one hand, low institutional levels in the labor market associated with a large informal sector limit labor mobility, which could discourage more productive workers and formal workers from moving to better jobs and restrict the effects of agglomeration economies. On the other hand, this sizable loss for formal workers due to urbanization may be due to the constraints in creation on formal jobs at the urban level, which may limit the incentives for workers to improve their skills locally and limit the scope of agglomeration benefits. Additionally, it is possible that workers in the informal sector may obtain higher benefits from the local market, and therefore, their incomes are more influenced by local housing and transportation costs, which could imply positive benefits associated with the urban scale.

We also found evidence of limited agglomeration returns for highly educated workers. This result contrasts with the results from the extant literature for developed countries that highlight the existence of higher agglomeration returns for more educated people. One possible explanation for this result could be the existence of important disamenities in denser cities in developing countries that are not compensated by wages, particularly for more educated workers who are more sensitive to the negative amenities that make working unpleasant, and therefore affect the benefits of agglomeration. However, it is important to note that further studies are necessary to better understand this inference.

This paper also contributes to the literature that argues that agglomeration economies encourage hard work (Rosenthal and Strange, 2008a), in this case, informal work. According to the literature on agglomeration, cities are productive places because they allow for labor pooling, the sharing of intermediate inputs, and knowledge spillover.

Informal workers also receive the benefits of these productive effects in the form of higher wages in denser cities.

There are certain limitations that are worth recognizing in this research and that, at the same time, can serve to identify areas for future research. First, in this paper, we focus on static gains from agglomeration, but more recent research recognizes that agglomeration economies can be dynamic and present a permanent effect (Combes and Gobillon, 2015; De la Roca and Puga, 2017). This limitation is due to data constraints because in Colombia, longitudinal information at the worker level is not available, which would allow researchers to take into account the dynamics of workers. However, we could perhaps use information at the firm level, which may make it easier to generate a panel structure. In the case of Colombia, it is possible to use the Colombian Annual Manufacturing Survey, in which each establishment has a unique ID that would allow us to follow them over time. Second, we consider employment density as our measure of agglomeration economies, but it is possible that there are effects of the densities of formal and informal workers on the productivity of each sector. Including these densities distinguished by type of worker could be a way to understand how the quality of agglomeration economies affects their economic productivity.

References

Albouy, D. (2008). "Are Big Cities Really Bad Places to Live? Improving Quality-of-life Estimates Across Cities". *Working Paper*, No 14472, National Bureau of Economic Research.

Albouy, D. (2009). "What are Cities Worth? Land Rents, Local Productivity, and the Capitalization of Amenity Values". *Working Paper*, No 14981, National Bureau of Economic Research.

Albouy, D., Leibovici, F., and Warman, C. (2013). "Quality of Life, Firm Productivity, and
the Value of Amenities across Canadian Cities". *Canadian Journal of Economics*, 46(2): 379–411.

Annez, P. and Buckely, R. (2009). Urbanization and Growth: Setting the Context. In M. Spence, Annez, P., and Buckley, R. (ed) *Urbanization and Growth*, the Commission Growth and Development, Washington, D.C.

Arango, L. and and Bonilla, G. (2015). "Human Capital Agglomeration and Social Returns to Education in Colombia". *Borradores de Economía* 883, Banco de la República, Bogotá-Colombia

Bacolod, M., Blum B., and Strange, W. (2009). "Skills in the City". *Journal of Urban Economics* 65(2): 136–153.

Barón, J. (2002). "Las Regiones Económicas de Colombia: Un Análisis de Clusters" *Documentos de Trabajo sobre Economía Regional*, No 23, Banco de la República, Cartagena.

Baum, C., Schaffer, M. and Stillman, S. (2007). "Enhanced Routines for Instrumental Variables/Generalized Method of Moments Estimation and Testing". *Stata Journal*, 7(4): 465-506.

Becker, C., Williamson, J. and Mills, E. (1992). *Indian Urbanization and Economic Growth since 1960*. Baltimore, MD: Johns Hopkins University Press.

Bernal, R. (2009). "The Informal Labor Market in Colombia: Identification and Characterization," *Desarrollo y Sociedad*, 63: 145–208.

Borjas, G. J. (2008). *Labor Economics*. Boston: McGraw-Hill.

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

Chauvin, J.P., Glaeser, E., and Tobio, K. (2014). *Urban Economics in the US and India*. Harvard University.

Combes, PP., Duranton, G., and Gobillon, L. (2008). "Spatial Wage Disparities: Sorting Matters!" *Journal of Urban Economics*, 63(2): 723–742.

Combes, PP., Duranton, G., Gobillon, L. and Roux, S. (2010). Estimating Agglomeration Economies with History, Geology, and Worker Effects. In E.L. Glaeser (ed), *Agglomeration Economies*, University of Chicago Press: Chicago.

Combes, PP., Démurger, S., and Li, S. (2015). "Migration Externalities in China". *European Economics Review*, 76: 152-167.

Combes, PP. and Gobillon, L. (2015). The Empirics of Aglomeration. In: G. Duranton, V. Henderson, J. Strange, (ed), *Handbook of Regional and Urban Economics*, vol. 5. North-Holland, Amsterdam.

D'Costa, S. and Overman, H. (2014). "The Urban Wage Growth Premium: Sorting or Learning?". *Regional Science and Urban Economics*, 48: 168–179.

Da Mata, D., Deichmann, U., Henderson, J., Lall, S. and Wang, H. (2007). "Determinants of City Growth in Brazil". *Journal of Urban Economics*, 62(2): 252–72.

DANE (2009). Resultados Fase 1: Empalme de las Series de Mercado Laboral, Pobreza y Desogualdad. Consulted at: https://www.dane.gov.co/files/noticias/Presentacion_pobreza_dane_DNP.pdf

Davidson, R. and MacKinnon, J. (2010). "Wild Bootstrap Tests for IV Regression". *Journal of Business and Economic Statistics*, 28(1): 128-144.

De la Roca, J. and Puga, D. (2017). "Learning by Working in Big Cities". *Review of Economic Studies*, 84(1): 106-142.

Dominicis, L., Arbia, G. and de Groot, H. (2007) "The Spatial Distribution of Economic Activities in Italy". *Tinbergen Institute Discussion Paper*. 2007-094/3.

Duranton, G. (2009). Are Cities Engines of Growth and Prosperity for Developing Countries. In M. Spence, P. Annez, and R. Buckley, (ed) *Urbanization and Growth*, the Commission Growth and Development, Washington, D.C.

Duranton, G. (2015). "Growing through Cities in Developing Countries". *World Bank Research Observer*, 30(1): 39-73.

Duranton, G. (2016). "Agglomeration Effects in Colombia". *Journal of Regional Science*, 56(2): 210-238.

Duranton, G. and Monastiriotis, V. (2002). "Mind the Gaps: The Evolution of Regional Earnings Inequalities in the UK 1982-1997". *Journal of Regional Science*, 42(2): 219–256.

Duranton, G. and Puga, D. (2004). Micro-foundations of Urban Agglomeration Economies. In: Henderson, J.V., Thisse, J.F. (ed), *Handbook of Regional and Urban Economics*, vol. 4. North-Holland, Amsterdam.

Fujita, M., Krugman, P. and Venables, A. J. (1999). *The Spatial Economy*, MIT Press, Cambridge, MA.

Galvis, L. (2001). "La Topografía Econímica de Colombia" *Documentos de Trabajo sobre Economía Regional*, No 22, Banco de la República, Cartagena.

Galvis, L. (2012). "Informalidad Laboral en las Áreas Urbanas de Colombia". *Coyuntura Económica*, 42(1): 15–51.

García, G. (2017). "Labor Informality: Choice or Sign of Segmentation? A Quantile Regression Approach at the Regional Level for Colombia". *Review of Development Economies*, 21(4): 985-1017.

García, G. and Badillo, E. (2018). "Rationing of Formal Sector Jobs and Informality: The Colombian Case". *Journal of International Development*, Forthcoming. http://onlinelibrary.wiley.com/doi/10.1002/jid.3310/full

Glaeser, E. (2008). *Cities, Agglomeration and Spatial Equilibrium*. Oxford: Oxford University Press.

Glaeser, E. and Maré, D. (2001). "Cities and Skills". *Journal of Labor Economics*, 19(2): 316-342.

Glaeser, E. and Resseger, M. (2010). "The Complementarity Between Cities and Skills". *Journal of Regional Science*, 50(1): 221–244.

Groot, S.P.T., de Groot, H.L.F., and Smit, M.J. (2014). "Regional Wage Differences in the Netherlands: Microevidence on agglomeration externalities". *Journal of Regional Science*. 54(3): 503–523.

Hansen, L. (1982). “Large Sample Properties of Generalized Method of Moments Estimators”. *Econometrica*, 50(4): 1029-1054.

Harris, J. (2014). “The Messy Reality of Agglomeration Economies in Urban Informality: Evidence from Nairobi’s Handicraft Industry”. *World Development*, 61: 102-113.

Harris, J. and Todaro, M. (1970). “Migration, Unemployment, and Development: A Two Sector Analysis”. *American Economic Review*, 60(1): 126–142.

Henderson, J. V. (1986). “Efficiency of Resource Usage and City Size”. *Journal of Urban Economics*, 19(1): 47–70.

Jütting, J. and De Laiglesia, J. (2009). “Is Informal Normal? Towards More and Better Jobs in Developing Countries”. OECD Development Centre, 163 pages.

Kleibergen, F. Paap, R. (2006). “Generalized Reduced Rank Tests using the Singular Value Decomposition”. *Journal of Econometrics*, 133(1): 97-126.

Lee Y.S. (2016). “Competition, Wage, and Agglomeration Diseconomy”. *International Regional Science Review*, 39(3): 318-349.

Lee, Y.J. and Zang, H. (1998). “Urbanization and Regional Productivity in Korean Manufacturing”. *Urban Studies*, 35(11): 2085–2099.

Melo, P. and Graham, D. (2009). “Agglomeration Economies and Labour Productivity: Evidence from Longitudinal Worker Data for GB’s Travel-to-Work Areas”. SERC Discussion Paper 31. SERC.

Melo, P., Graham, D. and Noland, R. (2009). “A Meta-Analysis of Estimates of Urban Agglomeration Economies”. *Regional Science and Urban Economics*, 39(3): 332-342.

Melo, P., Graham, D., Levinson, D., and Aarabi, S. (2017). “Agglomeration, Accessibility and Productivity: Evidence for Large Metropolitan Areas in the US”. *Urban Studies*, 54(1): 179-195.

Mills, E. and Becker, C. (1986). *Studies in Indian Urban Development*. New York: Oxford University Press.

Mion, G. and Naticchioni, P. (2009). “The Spatial Sorting and Matching of Skills and Firms”. *Canadian Journal of Economics* 42(1): 28–55.

Moreno-Monroy, A. (2012). “Informality in Space: Understanding Agglomeration Economies During Economic Development”. *Urban Studies*, 49(10): 2019-2030.

Moreno-Monroy, A. and García, G. (2016) “Intra-metropolitan Clustering of Formal and Informal Manufacturing Activity: Evidence from Cali, Colombia”. *Tijdschrift voor economische en sociale geografie*, 107(4): 389-406.

Murray, M. (2006). “Avoiding Invalid Instruments and Coping with Weak Instruments”. *The Journal of Economic Perspectives*, 20(4): 111-132.

Nicodemo, C. and García, G. (2015). “Job Search Channels, Neighborhood Effects, and Wages Inequality in Developing Countries: The Colombian Case”. *Developing Economies*, 53(2): 75-99.

Openshaw, S. and Taylor, P. (1979). A Million or so Correlation Coefficients: Three Experiments on the Modifiable Area Unit Problem, in N. Wrigley (ed), *Statistical Applications in the Spatial Sciences*, London: Pion.

Overman, H. and Venables, A. (2005). “Cities in the Developing World”. *CEP Discussion Paper*, No 695, London.

Perry, G., Maloney, W., Arias, O., Fajnzylber, P., Mason, A. and Saavedra-Chanduvi, J. (2007). *Informality: Exit and Exclusion*, The World Bank.

Roback, J. (1982). “Wages, Rents, and the Quality of Life”. *The Journal of Political Economy* 90(6): 1257–1278.

Rosenthal, S. and Strange, W. (2004). Evidence on the Nature and Sources of Agglomeration Economies. In: Henderson, V., Thisse, J.F. (ed), *Handbook of Regional and Urban Economics*, vol. 4. North-Holland, Amsterdam

Rosenthal, S. and Strange, W. (2008a). “Agglomeration and Hours Worked”. *The Review of Economics and Statistics*, 90(1): 105–118.

Rosenthal, S. and Strange, W. (2008b). “The Attenuation of Human Capital Spillovers”. *Journal of Urban Economics*, 64(2): 373–389.

Royuela, V. and García, G. A. (2015). “Economic and Social Convergence in Colombia”. *Regional Studies*, 49(2): 219-239.

Shea, J. (1997). “Instrument Relevance in Multivariate Linear Models: A Simple Measure”. *The Review of Economics and Statistics*, 79(2): 348-352.

Schneider F. and Enste, D. (2000). “Shadow Economies: Size, Causes and Consequences”. *Journal of Economic Literature*, 38(1): 77–114.

Schneider, F., Buehn, A. and Montenegro, C.E. (2010). “New Estimates for the Shadow Economies all Over the Word”. *International Economic Journal*, 24(4): 443-461.

Shukla, V. (1996). *Urbanization and Economic Growth*. Delhi: New York: Oxford University Press.

Stock, J. and Yogo, M. (2005) Testing for Weak Instruments in Linear IV Regression. In D. Stock, K. A. A. J. H. (ed) *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press.

Tabuchi, T. (1998). “Urban Agglomeration and Dispersion: A Synthesis of Alonso and Krugman”. *Journal of Urban Economics*, 44(3): 333-351.

Wheaton, W. and Lewis, M. (2002). “Urban Wages and Labor Market Agglomeration”. *Journal of Urban Economics*, 51(3): 542–562.

Wheeler, C. (2001). “Search, Sorting, and Urban Agglomeration”. *Journal of Labor Economics*, 19(4): 879–99.

Appendix

Table A1. Agglomeration effects, baseline model without informality (OLS)

Dependent variable: log hourly wage

	(1) Only emp. density	(2) Individual characteristics	(3) Sector-Occupation	(4) Geographic variables	(5) Market access	(6) Non-lineal	(7) Education effects 1	(8) Education effects 2
Log employment density	0.044***	0.036***	0.038***	0.036***	0.035***	0.062	0.040***	0.040***
	(0.0102)	(0.0102)	(0.0096)	(0.0067)	(0.0106)	(0.0803)	(0.0116)	(0.0075)
Log employment density^2						0.001		
						(0.0043)		
Log dist. Km to capital city					0.016			-0.026
					(0.0524)			(0.0309)
Log dist. Km to capital city^2					-0.005			-0.003
					(0.0087)			(0.0048)
Educ x Log emp density							-0.001	-0.001
							(0.0002)	(0.0006)
Observations	1,920,678	1,914,957	1,913,815	1,901,513	1,905,428	1,913,815	1,919,330	1,907,001
Municipalities	568	568	568	548	537	568	568	548
R2	0.024	0.380	0.418	0.426	0.418	0.418	0.407	0.417

Note: Robust standard errors clustered at the municipality level in parentheses. *** p<0.01, ** p<0.05, * p<0.1
All models include year dummy variables. In Columns 2 to 6 individual characteristics included are: education indicators (primary, basic school, high school, technical or technological education, and university), gender, age and its squared, years in the current job and its squared. In Columns 3 to 8 all models include occupation (10) and economic sector (8) indicators. Geographical characteristics in Column 4 include five regional indicators (Central, Eastern, Western, Caribbean and Orinoco), a water availability index, a soil erosion index and the log of altitude and its square. Column 5 uses the distance in Km to the capital city of the department as a measure of market access. Column 7 and 8 include years of educations and the interaction between years of education and log employment density (Educ x Log emp density). Column 8 replicates Column 7 but adds geographical characteristics and the distance in Km to the capital city of the department as controls.

Table A2. Agglomeration effects, baseline model with informality (OLS)

Dependent variable: log hourly wage

	(1) Only emp. density		(2) Individual characteristics		(3) Sector-Occupation		(4) Geographic variables		(5) Market access		(6) Non-lineal		(7) Education effects	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Log employment density	-0.037***	0.044***	-0.003	0.039***	0.004	0.038***	0.008	0.038***	0.006	0.035***	0.092**	0.052	-0.001	0.026**
	(0.009)	(0.0101)	(0.0071)	(0.0109)	(0.09593)	(0.0107)	(0.0049)	(0.0071)	(0.0061)	(0.0119)	(0.0375)	(0.0926)	(0.0069)	(0.0102)
Log employment density^2											0.0051**	0.001		
											(0.0020)	(0.0049)		
Log dist Km to capital city									-0.026	-0.002				
									(0.0309)	(0.0566)				
Log dist Km to capital city^2									0.004	-0.002				
									(0.0054)	(0.0093)				
Educ x Log emp density													0.001	0.001*
													(0.0005)	(0.0006)
Observations	673,939	1,246,739	672,709	1,242,248	672,247	1,241568	667,387	1,234,126	670,577	1,234,851	672,247	1,241,568	673,426	1,245,904
Municipalities	564	568	564	568	564	568	544	548	533	537	564	568	564	568
R2	0.022	0.025	0.501	0.207	0.562	0.236	0.564	0.247	0.562	0.237	0.562	0.236	0.542	0.232

Note: Robust standard errors clustered at the municipality level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

All models include year dummy variables. In columns 2 to 6 individual characteristics included are: education indicators (primary, basic school, high school, technical or technological education, and university), gender, age and its squared, years in the current job and its squared. In columns 3 to 7 all models include occupation (10) and economic sector (8) indicators. Geographical characteristics in column 4 include five regional indicators (Central, Eastern, Western, Caribbean and Orinoco), a water availability index, a soil erosion index and the log of altitude and its square. Column 5 uses the distance in Km to the capital city of the department as a measure of market access. Column 7 includes years of educations and the interaction between years of education and log employment density (Educ x Log emp density).

Table A3. Agglomeration effects, baseline model without informality (2SLS)

Dependent variable: log hourly wage

	(1) Only emp. density	(2) Individual characteristics	(3) Sector-Occupation	(4) Geographic variables	(5) Market access	(6) Non-lineal	(7) Education effects 1	(8) Education effects 2
Log employment density	0.029**	0.021*	0.022*	0.016*	0.012	-0.069	0.020	0.019
	(0.0116)	(0.0126)	(0.0119)	(0.0079)	(0.0090)	(0.1407)	(0.0154)	(0.0115)
Log employment density^2						-0.006		
						(0.0079)		
Log dist. Km to capital city								
Log dist. Km to capital city^2								
Educ x Log emp density							0.001	-0.001
							(0.0007)	(0.0007)
Observations	1,820,006	1,814,561	1,813,482	1,813,482	1,813,482	1,813,482	1,818,732	1,818,732
Municipalities	390	390	390	390	390	390	390	390
R2	0.022	0.381	0.419	0.425	0.424	0.419	0.408	0.416
Instruments exogeneity								
Hansen J statistic	0.064	0.291	0.328	3.413	3.400	7.334	0.223	2.708
Chi-sq P-val	0.968	0.864	0.849	0.182	0.183	0.119	0.895	0.258
Instruments relevance								
1. First-stage statistics								
Shea partial R2								
Log employment density	0.801	0.801	0.798	0.774	0.811	0.498	0.777	0.781
Log employment density^2						0.497		
Educ x Log emp density							0.780	0.781
2. Under-identification test								
Kleibergen-Paap rk LM stat	15.00	15.03	15.28	9.46	14.07	12.46	15.32	15.67
Chi-sq P-val	0.002	0.002	0.002	0.024	0.003	0.029	0.002	0.001
3. Weak identification test								
Kleibergen-Paap rk Wald F stat	13.91	64.68	66.05	75.84	62.00	9.57	42.34	36.29

Note: Robust standard errors clustered at the municipality level in parentheses. *** p<0.01, ** p<0.05, * p<0.1
All models include year dummy variables. In Columns 2 to 6 individual characteristics included are: education indicators (primary, basic school, high school, technical or technological education, and university), gender, age and its squared, years in the current job and its squared. In Columns 3 to 8 all models include occupation (10) and economic sector (8) indicators. Geographical characteristics in Column 4 include five regional indicators (Central, Eastern, Western, Caribbean and Orinoco), a water availability index, a soil erosion index and the log of altitude and its square. Column 5 uses the distance in Km to the capital city of the department as a measure of market access. Column 7 and 8 include years of educations and the interaction between years of education and log employment density (Educ x Log emp density). Column 8 replicates Column 7 but adds geographical characteristics and the distance in Km to the capital city of the department as controls.