

AGGLOMERATION ECONOMIES IN THE PRESENCE OF AN INFORMAL SECTOR: THE COLOMBIAN CASE

[Gustavo A. García](#)

Armand Colin | « [Revue d'Économie Régionale & Urbaine](#) »

2019/2 Février | pages 355 à 388

ISSN 0180-7307

ISBN 9782200932275

DOI 10.3917/reru.192.0355

Article disponible en ligne à l'adresse :

<https://www.cairn.info/revue-d-economie-regionale-et-urbaine-2019-2-page-355.htm>

Distribution électronique Cairn.info pour Armand Colin.

© Armand Colin. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

Agglomeration economies in the presence of an informal sector: the Colombian case

Économies d'agglomération en présence de l'informalité : le cas de la Colombie

Gustavo A. GARCÍA

Research in Spatial Economics (RISE-group)
Department of Economics, Universidad EAFIT
Medellin, Colombia
ggarci24@eafit.edu.co

Keywords : agglomeration economies, informal sector

Mots-clés : économies d'agglomération, informalité

JEL classification: R12, J46, R23, J31

Abstract

This paper analyzes the relationship between agglomeration economies and productivity in the context of a developing country in presence of an informal sector. Using data from Colombia, 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 have a significantly positive effect on productivity in the informal sector: informal workers in denser areas will earn approximately 11% more than those in less dense areas.

Résumé

Cette étude analyse la relation entre les économies d'agglomération et la productivité dans le cas d'un pays en voie de développement en présence d'un secteur informel. En utilisant des données de Colombie, on examine si le secteur informel bénéficie des économies d'agglomération et, surtout, ses différences par rapport au secteur formel en termes de rentabilité des économies d'échelle. Les résultats démontrent que les économies d'agglomération ont un impact très important sur la productivité du secteur informel : les travailleurs du secteur informel dans les endroits les plus densément occupés gagnent environ 11 % de plus de ceux qui résident ailleurs.

- 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 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 the GDP, SCHNEIDER *et al.* (2010) show that the shadow economy in developing countries accounts for approximately 40% of the 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 sectors achieve greater benefits from the diversity of activities and the spillovers associated with urbanization economies. For the analysis, we used worker data for Colombia for the period 2008-2014. 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. Section 3 introduces the empirical model. In Section 4, we present the data sources used in the analysis. Section 5 statistically documents the relationship between agglomeration and wages while taking into account the existence of the informal sector. Sections 6 and 7 discuss the results, and Section 8 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 existant 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. 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 two-stage least square (2SLS) strategy and instruments the current city population with long lags of population. 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 -

Estimation strategy

We formulate the following wage equation to be estimated:

$$\ln w_{i(t)} = \alpha_0 + \beta \ln \text{density} + X_{i(t)}\varphi + \pi_{oi(t)} + \sigma_{ei(t)} + \theta_{ai(t)} + \delta_t + \varepsilon_{i(t)} \quad (1)$$

where i identifies the worker, o refers to the 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, $\ln \text{density}$, 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, the 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 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 (1) 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 (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.

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: relevance and exogeneity. While the first condition demands our instruments to be correlated with the contemporaneous employment density, the second requires our instrument to 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 6.

- 4 -

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; 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, years 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 the 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 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, 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 the GEIH to ensure that the estimates are representative.

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 Bogota, Medellin, Itagüi, Cali, Bucaramanga, Barranquilla, and Soledad have the highest levels of urbanization. Itagüi is the densest city in Colombia, with just over 4,500 workers per km².⁸

- 5 -

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

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 (in \$)	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

Table 1 – (suite)

	Region							
	0.63	0.48	0.63	0.48	0.58	0.49	0.61	0.48
Informality								
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

Table 1 – (suite)

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

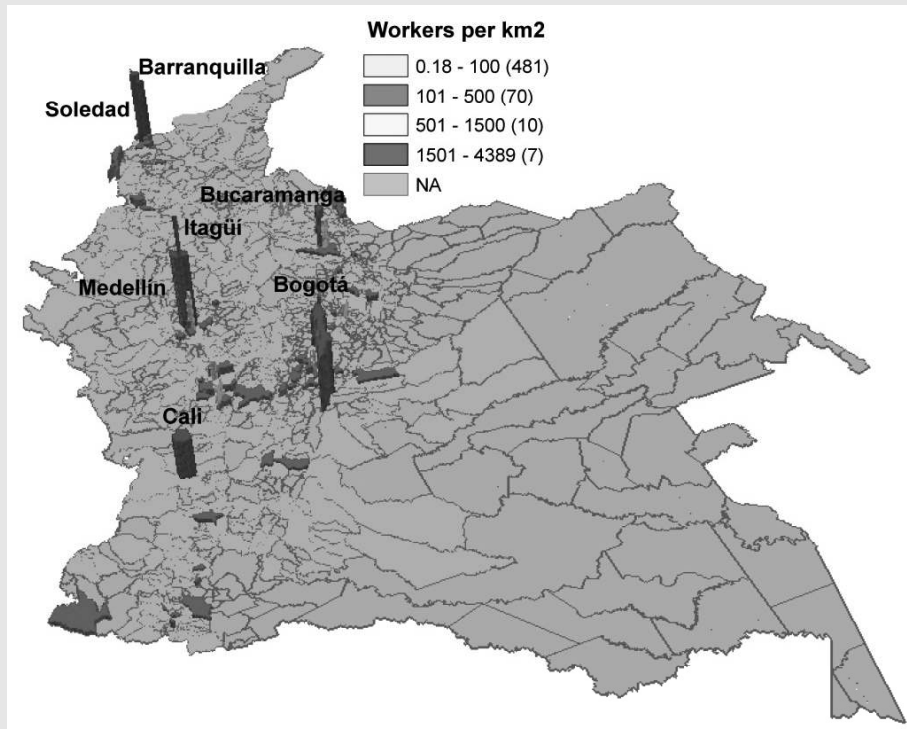
Note: All data are weighted using person sampling weights from GEIH to be representative.

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 (km ²)	568	161.3	358.6	842.2	765.9	1,529.6	15.4	17,641.70
Density (workers per km ²)	568	10.6	25.5	57.0	94.8	326.3	0.176	4388.98
Altitude (MASL)	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.

Figure 1 – Employment density by municipality in Colombia



Note: Average employment density between 2008 and 2014.

three largest municipalities and for municipalities with low employment density, that is, with a density 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.

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 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 3 and 4 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.

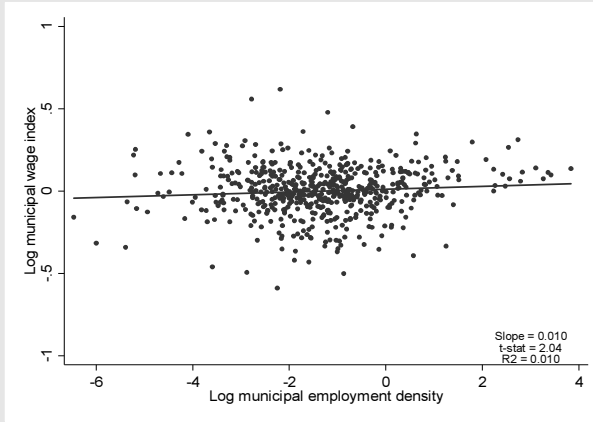
Table 3 – Average wages between formal and informal employees in select municipalities

Sector	Municipalities	Hourly wages (\$)			
		2008	2011	2014	All Years
Formal	Bogota, Medellin, Cali	2.13	2.75	2.47	2.45
	Less dense cities	2.02	2.94	2.48	2.58
Informal	Bogota, Medellin, 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%.

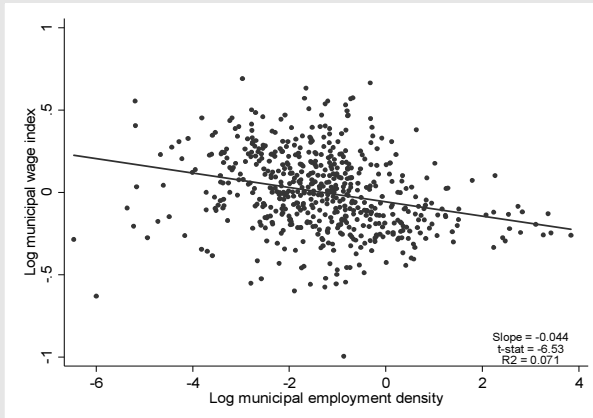
These results are somewhat surprising because formal workers are more educated than informal workers and might have a greater ability to learn from nearby human capital. Furthermore, formal workers work in medium-large enterprises that can obtain greater benefits from labor market pooling and input sharing 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, 2018), 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).

Figure 2 – Employment density and wages in Colombia - Total

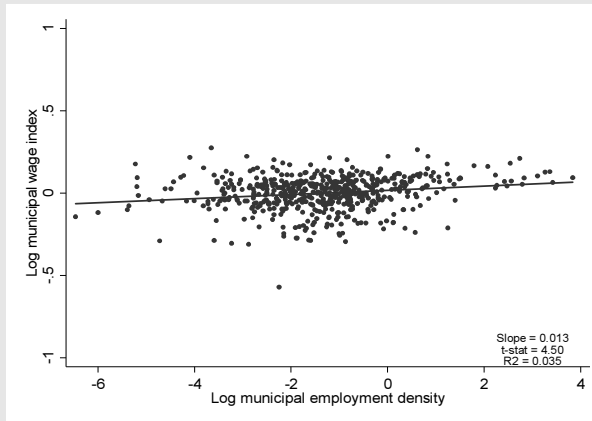


Note: The vertical axis represents log municipal hourly wages computed using 2008-2014 wage data after controlling for years effects, individual characteristics, occupation and economic sector. The horizontal axis represents log of average employment density between 2008 and 2014. There are 568 municipalities.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

Figure 3 – Employment density and wages in Colombia - Formal



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.

Figure 4 – Employment density and wages in Colombia - 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.

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 a lower wage. 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.

- 6 - Results

In this section, we present the results of the estimation of the wage equation by 2SLS, which are reported in Tables 4 and 5.⁹ 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 tables.

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 the excluded instruments to 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 Lagrange-Multiplier (LM) statistics. If we fail to reject the null hypothesis that the matrix of reduced-form coefficients is under-identified, it means that the instrumental variable bias of the parameter estimates will be increased. The values presented in Tables 4 and 5 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 rank WALD F statistic is higher than the STOCK and YOGO (2005) critical values, suggesting that our instruments are not weak.

When we estimate the model using the separate samples of formal and informal workers separately^{10,11}, the results of the elasticity of wages with respect to density without any other control in Column 1 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.¹²This difference between formal and informal workers echoes the summary measure in Table 3 and Figure 2 and persists throughout the paper.

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	
	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 **
Log employment density ²	(0.0111)	(0.0116)	(0.0080)	(0.0106)	(0.0067)	(0.0107)	(0.0061)	(0.0096)
Log distance (in km) to capital city								
Log distance (in km) to capital city ²								
Educ x Log emp density								
Observations	645,758	1,174,248	644,583	1,169,978	644,144	1,169,338	640,255	1,165,021
Municipalities	388	390	388	390	388	390	383	385
R ²	0.021	0.024	0.501	0.208	0.562	0.237	0.566	0.248
Instruments exogeneity								

Table 4 – (suite)

HANSEN J statistic	4.910	1.445	3.610	1.345	1.717	0.996	6.140	5.677
Chi-sq P-val	0.086	0.486	0.164	0.511	0.424	0.608	0.051	0.059
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
Log employment density ²								
Educ x Log emp density								
2. Under-identification test								
Kleibergen-Paap rank LM stat	15.12	14.83	15.53	14.91	16.174	14.99	13.47	12.98
Chi-sq P-val	0.002	0.002	0.001	0.002	0.001	0.002	0.004	0.005
3. Weak identification test								
Kleibergen-Paap rank Wald F stat	53.62	71.34	55.76	72.13	57.21	73.78	63.70	94.71

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.

Table 5 – Agglomeration effects and informality (2SLS) (cont.)
Dependent variable: log hourly wage

	(5) Market access		(6) Non-linear		(7) Education effects	
	Formal	Informal	Formal	Informal	Formal	Informal
Log employment density	-0.005	0.020 **	0.031	-0.082	-0.016	0.008
	(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 distance (in km) to capital city	0.008	0.022				
	(0.0366)	(0.0691)				
Log distance (in 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	643,312	1,166,400	644,144	1,169,338	645,269	1,173,463
Municipalities	378	380	388	390	388	390
R ²	0.562	0.238	0.562	0.237	0.543	0.233
Instruments exogeneity						
HANSEN J statistic	2.596	0.299	6.590	8.606	1.706	0.820
Chi-sq P-val	0.273	0.861	0.159	0.072	0.426	0.663
Instruments relevance						
1. First-stage statistics						
Shea partial R²						
Log employment density	0.801	0.824	0.504	0.502	0.762	0.778
Log employment density ²			0.508	0.497		

Table 5 – (suite)

Educ x Log emp density					0.769	0.782
2. Under-identification test						
Kleibergen-Paap rank LM stat	14.87	13.88	12.18	12.47	16.30	15.15
Chi-sq P-val	0.002	0.003	0.032	0.029	0.001	0.002
3. Weak identification test						
Kleibergen-Paap rank Wald F stat	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.

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 Bogota 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 which consists of small-scale producers, and it increases the costs for the 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).¹³

Regarding the other columns in Tables 4 and 5, 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 3.¹⁴ 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 of 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 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.

- 7 -

Robustness checks

To assess the robustness of these results, Tables 6 and 7 show the results when we experiment with different city samples and types of workers, when we use an alternative definition of informality, and when we employ different estimation techniques. Tables 6 and 7 use 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 6 shows the results of a regression that excludes observations corresponding to Bogota, the capital of Colombia¹⁵, while Column 2 shows the results excluding observations from the five densest cities in Colombia, namely, Itagüi, Medellín, Barranquilla, Soledad, and Bogota. 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 all 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 in 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.

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

Table 6 – Agglomeration effects and informality, robustness check
Dependent variable: log hourly wage

	(1) Excluding Bogota		(2) Excluding the densest cities		(3) Only the 23 main municipalities and MA		(4) MA and Municipalities		(5) Informality ILO	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Log employment density	-0.007 (0.0068)	0.016 ***	-0.009 (0.0084)	0.023 **	-0.003 (0.097)	0.017 *	-0.003 (0.0067)	0.019 **	0.008 (0.0081)	0.022 **
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
Municipalities	387	389	383	385	40	40	374	376	390	390
R ²	0.567	0.236	0.572	0.241	0.562	0.239	0.562	0.237	0.519	0.196
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
Chi-sq P-val	0.673	0.468	0.955	0.369	0.209	0.785	0.413	0.524	0.873	0.654
Instruments relevance										
1. First-stage statistics										

Table 6 – (suite)

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
2. Under-identification test										
Kleibergen-Paap rank LM stat	14.94	14.03	13.84	13.93	13.52	12.84	14.56	12.76	15.56	15.29
Chi-sq P-val	0.002	0.003	0.003	0.003	0.004	0.005	0.002	0.005	0.001	0.002
3. Weak identification test										
Kleibergen-Paap rank Wald F stat	59.09	83.11	71.06	82.13	49.93	65.76	47.71	72.54	58.66	74.69

Table 7 – Agglomeration effects and informality, robustness check (cont.)
Dependent variable: log hourly wage

	(6) Excluding low-educated self-employed informal workers		(7) Real hourly wages		(8) GMM		(9) LIML	
	Formal	Informal	Formal	Informal	Formal	Informal	Formal	Informal
Log employment density	-0.004	0.029 **	-0.001	0.017 **	-0.006	0.021 *	-0.004	0.021 **
Observations	644,144	446,607	603,364	1,059,188	644,144	1,169,338	644,144	1,169,338
Municipalities	388	388	40	40	388	390	388	390
R ²	0.562	0.402	0.557	0.239	0.562	0.237	0.562	0.237
Instruments exogeneity								
HANSEN J statistic	1.717	1.439	4.115	1.247	1.717	0.996	1.717	0.996
Chi-sq P-val	0.424	0.487	0.128	0.536	0.424	0.608	0.424	0.608
Instruments relevance								
1. First-stage statistics								
Shea partial R ²								
Log employment density	0.786	0.795	0.821	0.859	0.786	0.803	0.786	0.803
2. Under-identification test								
Kleibergen-Paap rank	16.17	14.83	13.92	12.70	16.17	14.99	16.17	14.99
LM stat								
Chi-sq P-val	0.001	0.002	0.003	0.005	0.001	0.002	0.001	0.002
3. Weak identification test								
Kleibergen-Paap rank	57.20	62.84	54.69	74.30	57.21	73.78	57.21	73.78
Wald F stat								

Table 7 – (suite)

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 Bogota and Column 2 excludes observations corresponding to the densest municipalities in terms of employment: Itagüi, Medellín, Barranquilla, Soledad and Bogota. Column 3 aggregates municipalities into metropolitan areas (MA) and only uses information to 7 MA (Medellin-Valle de Aburra, Cali-Yumbo, Barranquilla-Soledad, Bucaramanga-Girón-Piedecuesta-Floridablanca, Manizales-Villa Maria, Pereira-Dos Quebradas-La Virginia, and Cucuta-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.

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.¹⁶ 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-employed¹⁷ and therefore their wages may not reflect the marginal productivity of labor, we re-estimated the model while excluding low-educated self-employed, informal workers¹⁸. 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 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, 2001; 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.

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 and type of employment.

- 8 -

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 which 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.

Aknowledgements

This paper has benefited from comments from participants at the 62nd Annual North American Meetings of the Regional Science Association International (NARSC-2015), and CEDE seminar at Universidad de Los Andes. I am grateful to Harris SELOD, Rosella NICOLINI, José Luis ROIG and three anonymous referees for their helpful suggestions on an early version of the paper. Financial support from Universidad de Los Andes and Colciencias program “Es Tiempo de Volver” is gratefully acknowledged. Any remaining error is my own responsibility.

References

- ALBOUY D (2008) Are big cities really bad places to live? Improving quality-of-life estimates across cities. *Working Paper* 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* 14981. National Bureau of Economic Research.
- ALBOUY D, LEIBOVICI F, 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-C, BUCKLEY R-M (2009) Urbanization and growth: setting the context. In: SPENCE M, ANNEZ P-C, BUCKLEY R-M (eds) *Urbanization and Growth*. Commission Growth and Development. World Bank Publications, Washington, DC.
- ARANGO L-E, BONILLA G (2015) Human capital agglomeration and social returns to education in Colombia. Banco de la República n° 012788, Bogota.
- BACOLOD M, BLUM B, STRANGE W (2009) Skills in the city. *Journal of Urban Economics* 65(2): 136-153.
- BAUM C-F, SCHAFER M-E, STILLMAN S (2007) Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *Stata Journal* 7(4): 465-506.
- BERNAL S-R (2009) The informal labor market in Colombia: identification and characterization. *Revista Desarrollo y Sociedad* 63: 145-208.
- BORJAS G-J (2008) *Labor economics*. McGraw-Hill, Boston.

Agglomeration economies in the presence of an informal sector

- CICCONE A, HALL R-E (1996) Productivity and the density of economic activity. *American Economic Review* 86(1): 54-70.
- CHAUVIN J-P, GLAESER E, TOBIO K (2014) *Urban economics in the US and India*. Harvard University.
- COMBES P-P, DURANTON G, GOBILLON L, ROUX S (2010) Estimating agglomeration economies with history, geology, and worker effects. In: GLAESER E-L (ed) *Agglomeration Economics*. University of Chicago Press.
- COMBES P-P, DÉMURGER S, LI S (2015) Migration Externalities in China. *European Economics Review* 76: 152-167.
- COMBES P-P, GOBILLON L (2015) The empirics of agglomeration. In: DURANTON G, HENDERSON J-V, STRANGE W-C (eds) *Handbook of Regional and Urban Economics* (vol. 5). Amsterdam.
- DA MATA D, DEICHMANN U, HENDERSON J-V, LALL S-V, WANG H-G (2007) Determinants of city growth in Brazil. *Journal of Urban Economics* 62(2): 252-272.
- DAVIDSON R, MACKINNON J-G (2010) Wild bootstrap tests for IV regression. *Journal of Business and Economic Statistics* 28(1): 128-144.
- DE LA ROCA J, PUGA D (2017) Learning by working in big cities. *Review of Economic Studies* 84(1): 106-142.
- DOMINICIS L, ARBIA G, DE GROOT H-L-F (2007) The spatial distribution of economic activities in Italy. *Tinbergen Institute Discussion Paper* N° TI 2007-094/3.
- DURANTON G (2009) Are cities engines of growth and prosperity for developing countries. In: SPENCE M, ANNEZ P-C, BUCKLEY R-M (eds) *Urbanization and Growth*. Commission Growth and Development. World Bank Publications, Washington, DC.
- 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, 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.
- FUJITA M, KRUGMAN P-R, VENABLES A-J (1999) *The Spatial Economy*. MIT Press, Cambridge, MA.
- GALVIS L-A (2012) Informalidad laboral en las áreas urbanas de Colombia. *Coyuntura Económica*, 42(1): 15-51.
- GARCÍA G-A (2017) Labor informality: choice or sign of segmentation? A quantile regression approach at the regional level for Colombia. *Review of Development Economics* 21(4): 985-1017.
- GARCÍA G-A, BADILLO E-R (2018) Rationing of formal sector jobs and informality: The colombian case. *Journal of International Development* 30(5): 760-789.
- GARCÍA G-A (2018) Agglomeration economies in the presence of an informal sector. The colombian case. *Documentos de Trabajo* No. 18-012018, Universidad EAFIT.
- GLAESER E-L (2008) *Cities, agglomeration and spatial equilibrium*. Oxford University Press.
- GLAESER E-L, MARÉ D-C (2001) Cities and skills. *Journal of Labor Economics* 19(2): 316-342.
- GLAESER E-L, RESSEGER M-G (2010) The complementarity between cities and skills. *Journal of Regional Science* 50(1): 221-244.
- GRAHAM D-J, MELO P-C (2009) Agglomeration economies and labour productivity: evidence from longitudinal worker data for GB's travel-to-work areas. SERC Discussion Paper 31. SERC.
- HANSEN L-P (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50(4): 1029-1054.
- HARRIS J-R (2014) The messy reality of agglomeration economies in urban informality: evidence from Nairobi's handicraft industry. *World Development* 61: 102-113.
- HARRIS J-R, TODARO M-P (1970) Migration, unemployment, and development: a two sector analysis. *American Economic Review* 60(1): 126-142.
- JÜTTING J, DE LAIGLESIA J (2009) Is informal normal? Towards more and better jobs in developing countries. OECD Development Centre.

- 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.
- MELO P-C, GRAHAM D-J, NOLAND R-B (2009) A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics* 39(3): 332-342.
- MELO P-C, GRAHAM D-J, LEVINSON D, AARABI S (2017) Agglomeration, accessibility and productivity: evidence for large metropolitan areas in the US. *Urban Studies* 54(1): 179-195.
- MORENO-MONROY A-I (2012) Informality in space: understanding agglomeration economies during economic development. *Urban Studies* 49(10): 2019-2030.
- MORENO-MONROY A-I, GARCÍA G-A (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-P (2006) Avoiding invalid instruments and coping with weak instruments. *The Journal of Economic Perspectives* 20(4): 111-132.
- NICODEMO C, GARCÍA G-A (2015) Job search channels, neighborhood effects, and wages inequality in developing countries: the colombian case. *Developing Economies* 53(2): 75-99.
- OPENSHAW S, TAYLOR P (1979) A million or so correlation coefficients: three experiments on the Modifiable Area Unit Problem. In: WRIGLEY N (ed) *Statistical Applications in the Spatial Sciences*. Pion, London.
- OVERMAN H, VENABLES A (2005) Cities in the developing world. *CEP Discussion Paper*, No 695, London.
- PERRY G-E, MALONEY W-F, ARIAS O, FAJNZYLBER P, MASON A, SAAVEDRA-CHANDUVI J (2007) *Informality: exit and exclusion*. The World Bank, Washington, DC.
- ROBACK J (1982) Wages, rents, and the quality of life. *The Journal of Political Economy* 90(6): 1257-1278.
- ROSENTHAL S-S, STRANGE W-C (2008a) Agglomeration and hours worked. *The Review of Economics and Statistics*, 90(1): 105-118.
- ROSENTHAL S-S, STRANGE W-C (2008b) The attenuation of human capital spillovers. *Journal of Urban Economics* 64(2): 373-389.
- ROYUELA V, 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, ENSTE D-H (2000) Shadow economies: size, causes and consequences. *Journal of Economic Literature* 38(1): 77-114.
- SCHNEIDER F, BUEHN A, MONTENEGRO C-E (2010) New estimates for the shadow economies all over the world. *International Economic Journal* 24(4): 443-461.
- STOCK J-H, YOGO M (2005) Testing for weak instruments in linear IV regression. In: STOCK J-H, YOGO M, ANDREWS D-W (eds) *Identification and inference for econometric models: essays in honor of Thomas ROTHENBERG*. Cambridge University Press.
- WHEATON W-C, LEWIS M-J (2002) Urban wages and labor market agglomeration. *Journal of Urban Economics* 51(3): 542-562.
- WHEELER C-H (2001) Search, sorting, and urban agglomeration. *Journal of Labor Economics* 19(4): 879-99.

Notes

- 1 - A detailed theoretical framework is described in Section 3 of the working paper version (GARCÍA, 2018).
- 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.

- 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/>)
- 5 - Given the confidentiality of the data at the municipal level, all the estimations in this paper were conducted following the DANE's microdata-access policy, which implies working in situ under the supervision of the DANE's staff and with blinded access to sensitive information.
- 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, Bogota. 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 - A more detailed description of these regions can be found in the working paper version (GARCÍA, 2018).
- 8 - These same cities also have the highest populations, and Itagüi is the most densely populated city in Colombia, with 12,114 people per km² in 2014.
- 9 - The estimations by OLS are provided in the working paper version (GARCÍA, 2018).
- 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 - The estimations without distinguishing between the formal and informal sector are reported in the working paper version (GARCÍA, 2018).
- 12 - We followed the formula developed by COMBES and GOBILLON(2015): $2^\beta - 1$, where β is the elasticity of productivity with respect to density.
- 13 - GARCÍA and BADILLO (2018), 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.
- 14 - 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.
- 15 - According to ARANGO and BONILLA (2015) and GARCÍA (2017), Bogota accounts for approximately 17% of the country's total population, 25% of its total employment, and 25% of its GDP.
- 16 - 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 (BERNAL, 2009; GALVIS, 2012; GARCÍA, 2017).
- 17 - On average, 67% of informal workers are self-employed.
- 18 - 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.