# Agglomeration effects and informal firms in the internal structure of cities

Andres Dominguez

*Department of Economics, Universidad Sergio Arboleda, Bogota, Colombia*

## Abstract

**Purpose** – This paper aims to estimate the effect of agglomeration on the probability of being an informal firm in Cali, Colombia. Informal firms produce legal goods but do not comply with official regulations. This issue is relevant because, similar to other developing countries, the informal sector in Colombia employs more than 50 per cent of the workforce. The results of this study demonstrate that one standard deviation increase in agglomeration reduces by 52 per cent the probability of being informal. Results are consistent with the idea that informal firms benefit less from agglomeration because of legal restrictions that block the relationship with formal firms.

**Design/methodology/approach** – The objective of the present paper is to estimate the effect of agglomeration on the probability that a firm – given a location – chooses to be informal. The authors deal with endogeneity issues by using soil information related to earthquake risk, which reduces the height of buildings and therefore increases the cost of agglomeration. The analysis focuses on Cali, Colombia, where the informal sector employs 60 per cent of the workforce. The registration of economic activities is used as a criterion to identify informal firms, in such a way that the percentage of informal firms is 42 per cent.

**Findings** – The authors find that the effect of agglomeration is strongly negative. The probability of being informal diminishes by 52 per cent when agglomeration increases by one standard deviation. Results in this paper shed light on how formal firms tend to be localized in high-density commercial and industrial areas, while informal firms are localized in low-density and peripheral areas where the land for production is cheaper and where they can avoid the control of authorities.

**Originality/value** – Theory argues that spatial production externalities and commuting costs are among the main forces that shape the city's internal structure. Externalities include effects that increase firms' production, and therefore workers' income, when the size of the local economy grows. The authors now have strong evidence that firms' productivity is positively related with the volume of nearby employment. Most of the empirical findings concern firms in the formal sector and, accordingly, the literature says little about the effect of agglomeration on informal firms' location. However, this effect is crucial for developing countries where informal work is the main option for less-educated workers facing unemployment.

**Keywords** Informality, Agglomeration effects, Firm location

**Paper type** Research paper

## 1. Introduction

Cities can be studied as market responses to production and income opportunities (Mills, 1967). Accordingly, the size and growth of urban areas can be interpreted as

responses to these opportunities[1]. Indeed, there is a strong positive relationship between productivity and economic density for different industries and levels of aggregation (Ciccone and Hall, 1996; Ciccone, 2002; Brulhart and Mathys, 2008; Melo and Noland, 2009; Combes *et al.*, 2010; Morikawa, 2011).

Theory argues that spatial production externalities and commuting costs are among the main forces that shape the city's internal structure. Externalities include effects that increase firms' production, and therefore workers' income, when the size of the local economy grows[2]. We now have strong evidence that firms' productivity is positively related with the volume of nearby employment (Rosenthal and Strange, 2003; Combes and Gobillon, 2015). Most of the empirical findings concern firms in the formal sector and accordingly, the literature says little about the effect of agglomeration on informal firms' location (Duranton, 2009). However, this effect is crucial for developing countries where informal work is the main option for less-educated workers facing unemployment. Indeed, in several countries, more than 50 per cent of employment is in the informal sector (Maloney, 2004; Perry *et al.*, 2007).

One of the main factors that explain informality is a significant cost related to formalization. A representative entrepreneur evaluates costs and benefits and could find that it is efficient to choose informality. Indeed, the literature points out that high taxes and social security contributions are costly regulations that lead entrepreneurs to not set up formal business and hence to not register their firms (de Soto, 2000; Maloney, 2004)[3]. Nevertheless, there is a threshold in the level of the economic activity when the incentives of formalization become more important.

The informal sector comprises a heterogeneous mixture of self-employed entrepreneurs, small and short-life firms that, although they produce legal goods, do not comply with legal regulations. When firms choose to be informal they do not have access to all markets where property rights are secure and well-defined. When a firm is informal, the owner faces restrictions in the financial system: they cannot make long-time capital investments and cannot use their property as collateral to secure loans (Feige, 1990; de Soto, 2000; Sindzingre, 2006). As a result, informal firms have lower productivity levels, lower fixed assets per worker and less access to government services than formal firms (Cárdenas and Mejia, 2007; Santa Maria and Rozo, 2009). In addition, they do not comply with labour regulations, may practice smuggling and, frequently, do not carry accounting[4].

We have evidence that formal and informal firms (of similar size and belonging to the same economic sector) display different locational patterns within an urban area (Moreno-Monroy and García, 2015). Furthermore, there is evidence (for São Paulo) that informality rates decrease on average 15 per cent faster in areas with new transport infrastructure (Moreno-Monroy and Roman, 2015) in line with the idea that informality is a choice based on a cost-benefit calculation of the entrepreneur.

The objective of the present paper is to estimate the effect of agglomeration on the probability that a firm – given a location – chooses to be informal. We deal with endogeneity issues by using soil information related to earthquake risk, which reduces the height of buildings and therefore increases the cost of agglomeration. The analysis focuses on Cali, in the west of Colombia, where the informal sector employs 60 per cent of the workforce. Using the registration of economic activities as a criterion to identify formal and informal firms (Schneider, 2005), we identify informal firms when they are not registered in the Chamber of Commerce, in such a way that the percentage of informal firms is 42 per cent.

We find that the effect of agglomeration is strongly negative. The probability of being informal diminishes by 52 per cent when agglomeration increases by one standard deviation. Results in this paper shed light on how formal firms tend to be localized in high-

density commercial and industrial areas, while informal firms are localized in low-density and peripheral areas where the land for production is cheaper and where they can avoid the control of authorities.

This paper proceeds as follows: Section 1 contains the theoretical framework, Section 2 presents the data and a discussion about the results and Section 3 concludes.

## 2. Theoretical framework
A linear city model where employment clustering is determined by an agglomeration externality was introduced by Fujita and Ogawa (1982)[5]. An improvement was made by Lucas and Rossi-Hansberg (2002) where, in a circular city model, firms and workers compete for land at different locations, and the external agglomeration effects lead firms to outbid residential use for land near production centres. The interaction between agglomeration effects and commuting costs is then the main determinant of urban structure: firms have an incentive to be close to each other to obtain benefits from agglomeration, whereas workers prefer proximity to the workplace to minimize commuting costs. Then, market prices (land rents and wages) give firms and households incentives for making land use decisions.

Workers consume residential land and a good which is produced using labour and land. If productivity increases with employment levels in neighbouring locations, firm production per unit of land at location $s$, $x(s)$, is expressed as $x(s) = A z_s^{\gamma} n^{\alpha}$, where $z_s$ represents the agglomeration effect on production at location $s$, $A$ is a productivity constant, $n$ is the number of employees and $\alpha < 1$[6]. The profit per unit of land at location $s$ is represented as $q(s) = A z_s^{\gamma} n^{\alpha} - w(s)n.$, where $w(s)$ is the wage rate. Firms choose employment $n$ to maximize profits. From the first-order condition we obtain $n = \hat{n}(w, z)$ and $q = \hat{q}(w, z)$. Therefore, given $w$ and $z$, the business bid rent is determined[7]. The model implies that land use depends on the difference between bids made by households and firms.

For simplicity, let us assume that informal and formal firms are identical except that $\gamma = 0$ for informal firms, because they are not allowed to have formal business contracts. When $\gamma = 0$ for informal firms, we assume that informal firms cannot benefit from agglomeration effects. These informal firms have the following production function:

$$x(s) = A n^{\alpha}$$

The model predicts then that these firms will be less productive and smaller. Now let us suppose that firms can choose to be formal or informal at a given cost $c > 0$. Given this assumption, firms will be less likely to choose being informal given higher levels of agglomeration. This happens because informal firms benefit less from positive technological and pecuniary externalities that arise in close spatial proximity due to legal restrictions that block the relationship with formal firms. Formal property allows assets to be identified and linked to other assets in the economy. Then, to benefit from external effects, the owner has incentives to formalize the business. As a result, we will find formal firms located in high density areas meanwhile informal firms will be located in low density areas. Accordingly, we aim to estimate the causal effect of agglomeration on the probability of being informal (controlling for the firm size and the economic sector).

## 3. Empirical results
### 3.1 Data
We focus on Cali, which is the third city in terms of population in Colombia. It was founded on 25 July 1536 and is located in the west of Colombia, in the Cauca Valley. The metropolitan area of Cali has a population of about 2,200,000 and a density of 21,295 persons per square

kilometre (in 2005). The city has two natural limits: the limit to the east is the Cauca River and the limit to the west is the Western Mountain Range (Cordillera Occidental)[8]. The urban area is divided into 338 administrative neighbourhoods with an average of 0.36 km².

Our main source of information is the Economic Census carried out by the National Institute of Statistics (*Departamento Administrativo Nacional de Estadística*, DANE). The database contains population data as well as establishment-level information, including employment, economic sector (two-digit level), geographical location and compliance with legal requirements. From now on, we will refer to establishments as firms.

Information about the compliance of legal requirements enables us to identify informal firms which, although they produce legal goods and services (e.g. bread, shoe repair, groceries shop), do not fully comply with legal regulations[9]. This means that informal firms are known, but not registered in the Chamber of Commerce (*Camara de Comercio de Cali*), which certifies firm ownership. These firms evade taxes, have less rigorous bookkeeping, do not contribute to social security and face restrictions to formal financial credit (Cárdenas and Rozo, 2009)[10]. Furthermore, informal firms are not allowed to have commercial or financial relations with formal or public firms. In Cali, there are 22,208 informal firms, which is about 43 per cent of the 51,457 firms.

We measure agglomeration, $z_s$, as a weighted average of the number of jobs at locations $h$, with weights that are a decreasing function of distance between $s$ and $h$ (see also Koster and Rouwendal, 2013)[11]. To be specific, $z_s$ is defined as follows:

$$z_s = \int J_h e^{-\delta d_{sh}} \, dh \tag{1}$$

where $J_h$ denotes the number of jobs at location $h$, $d_{sh}$ denotes the distance between locations $s$ and $h$, and $\delta > 0$ is a given decay parameter. The external effect is more localized for higher values of $\delta$, which implies that the value to a firm of locating near other producers is also higher. Figure 1 shows the weights as a decreasing function of distance, and it depends on the decay parameter $\delta$. The horizontal axis measures the distance between locations. For instance, if $\delta = 5$, $z_s$ will be similar at the number of jobs in the location $s$; if $\delta = 0,5$ then the surrounding areas are very important.
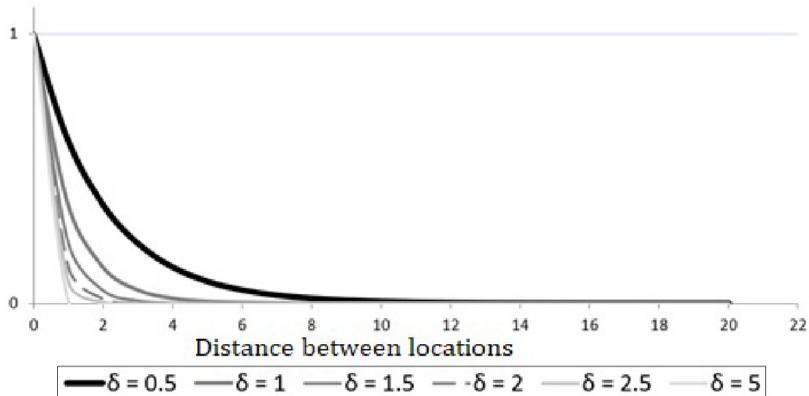


**Figure 1.**
Weights as a function
of the decay
parameter $\delta$

**Source:** Own elaboration

Panel (a) in Figure 2 shows the agglomeration variable for $\delta = 2$. We can observe that employment density is high near the city centre; as we move away from the centre we find areas of residential use. Panel (b) shows that the proportion of informal firms is higher in areas away from the city centre; and Panel (c) shows that the east of the city is affected by a liquefaction risk, which is present in areas where the soil is saturated with water and then acts like a liquid when shaken by an earthquake[12]. From among 338 neighbourhoods, 158 are affected by liquefaction risk in the east of the city, which represents 47 per cent of the neighbourhoods.

Table I presents the descriptive statistics of the main variables by neighbourhood. On average, the share of informal firms in a neighbourhood is 0.42, with a standard deviation of
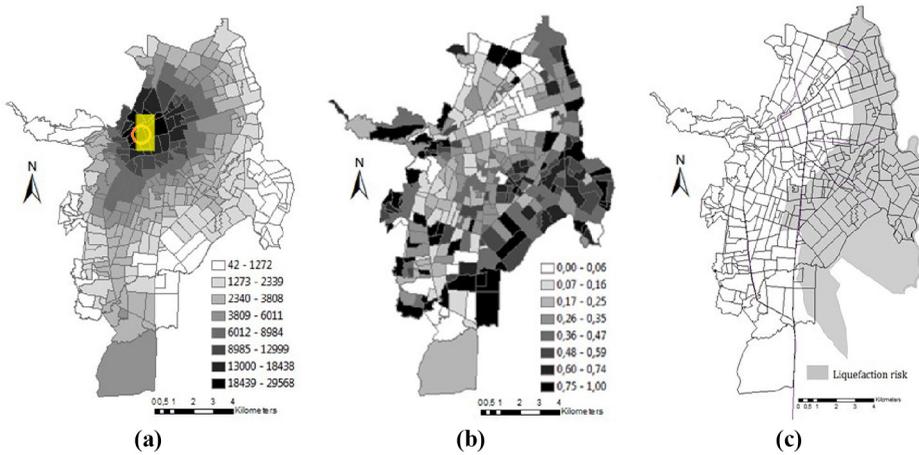
**Source:** Own elaboration

**Figure 2.**
(a) Agglomeration, (b) informality share and (c) liquefaction risk

| Variable | Mean | SD | Minimum | Maximum |
| --- | --- | --- | --- | --- |
| Local share of informal firms | 0.42 | 0.31 | 0 | 1 |
| ln (Agglomeration, $\delta = 2$) | 7.81 | 0.87 | 4.35 | 10.30 |
| Agglomeration, $\delta = 2$ | 3,648 | 4,020 | 77 | 29,804 |
| ln (Formal agglomeration, $\delta = 2$) | 7.40 | 1.03 | 3.35 | 10.07 |
| Formal agglomeration, $\delta = 2$ | 2,749 | 3,343 | 29 | 23,709 |
| Firm's size | 5 | 35 | 1 | 2,415 |
| Jobs | 727 | 1,417 | 0 | 16,087 |
| Formal jobs | 551 | 1,191 | 0 | 13,172 |
| Informal jobs | 176 | 324 | 0 | 3,319 |
| Job density (per km$^2$) | 3,109 | 5,794 | 0 | 43,435 |
| Population | 5,972 | 5,763 | 0 | 45,702 |
| Population density (per km$^2$) | 21,296 | 13,022 | 0 | 56,716 |
| Neighbourhood Area (km$^2$) | 0.36 | 0.50 | 0.02 | 7.84 |
| Distance from CBD (km) | 3.79 | 2.47 | 0 | 16.43 |
| Distance from main corridors (km) | 0.56 | 0.62 | 0 | 6.55 |

**Notes:** 338 neighbourhoods. In levels, the agglomeration variable has a mean equal to 3,648, and the standard deviation is 4,020. We use these data to calculate the effect of a one standard deviation change

**Table I.**
Descriptive statistics

0.31. On average, a firm has five workers and the city has around 727 jobs per neighbourhood, of which 551 are formal jobs and 176 are informal. The area of a neighbourhood is, on average, 360 m². The Central Business District (CBD) is localized in the city centre of Cali, around the *Plaza de Caicedo* (it is identified by the orange circle in the first map of Figure 1. It is the neighbourhood with the highest employment density). The average distance between a neighbourhood's centroid and the CBD is approximately 3.8 km, and the average distance between a centroid and the main corridors of the city is 560 m. Next section describes the methodology.

### 3.2 Methodology
We estimate a probit model where the dependent variable, $y_{is}$, is equal to 1 for an informal firm $i$ in a neighbourhood $s$, and equal to 0 otherwise. We are interested in the effect of agglomeration $z_s$, hence we assume:

$$Prob(y_{is} = 1 | x) = \Phi(X\beta), \tag{2}$$

where $X$ contains the effect of agglomeration $z_s$ and variables which represent firms' characteristics such as industrial sector (two-digit-level ISIC) and the number of workers. Moreover, it includes spatial variables such as distance from the city centre (in discrete categories to capture non-monocentric effects), distance from the main corridors (also in discrete categories), and X and Y coordinates, which are included to control for unobserved factors that smoothly vary over space.

The agglomeration variable, $z_s$, depends on the value of $\delta$, which is unknown[13]. We have estimated the model for different values of $\delta$ (from 0.1 to 5). The maximum fit occurs when $\delta$ is 2 (see Figure A5 in the Appendix: $R^2$ and log-pseudo likelihood using different values of $\delta$). We report results using this value, which implies that the agglomeration effect disappears within 2 km [Panel (a) in Figure 2 shows the geographical distribution of agglomeration and Table I shows that the logarithm of this variable has a mean of 7.81 with a standard deviation of 0.87; we use values in levels to interpret the effect of the variable on the probability of being informal][14].

### 3.3 Main results
The estimated marginal effects using information from 51,454 firms are presented in Table II (standard errors are clustered by neighbourhood). We control for size and industrial

| Explanatory variable | Probit | IV-probit |
| --- | --- | --- |
| ln (Agglomeration) | −0.1051 (0.023)*** | −0.6953 (0.107)*** |
| Size of the firm | (–) | (–) |
| Industrial sectors (47) | Yes | Yes |
| Distance from CBD | Yes | Yes |
| Distance from main corridors | Yes | Yes |
| Number of observations | 51,454 | 51,454 |
| F-statistic (weak instrument) | | 18.81[a] |

**Table II.**
Marginal effects on probability of being informal using total agglomeration as an explanation variable

**Notes:** Marginal effects at the means of the independent variables. Clustered standard errors by neighbourhood in parenthesis; *** $p < 0.01$; [a]The $F$-statistic for weak instruments is higher than 10, which implies that instrument has a strong effect on agglomeration (Stock and Yogo, 2005)

sector of the firm (to control the productivity), distance from CBD and distance from main corridors. In the probit estimation the marginal effect of agglomeration is negative (−0.1051), which means that one standard deviation increase in agglomeration reduces the probability of being informal by 7.8 per cent[15].

The estimated effects of the Probit model are likely to be biased because of the presence of omitted variables which are correlated with agglomeration and the probability of informality (Ellison and Glaeser, 1999; Bayer and Timmins, 2007). These omitted variables may be related to the educational level of employees and fixed assets of the firm.

Using an instrumental variable that is correlated with agglomeration but uncorrelated with any unobserved locational advantage may correct the bias. Geological variables such as soil composition, rock depth, water capacity, soil erodibility and seismic and landslide hazard have been used for coping with endogeneity (Rosenthal and Strange, 2008; Combes et al., 2010; Combes and Gobillon, 2015). Characteristics of soil were important to localize original settlements, and agglomeration processes have then developed in those areas. In that case, the instrument is relevant. Our instrument is based on liquefaction risk, which refers to the strength and stiffness of the soil when it is affected in the case of earthquakes. Liquefaction risk is present in areas where the soil is saturated with water and then acts like a liquid when shaken by an earthquake. Earthquake waves cause water pressure to increase in the sediment, so sand grains lose contact with each other and the soil loses its ability to support high buildings. It can be argued that the instrument is exogenous because there is no reason to say that the probability of being informal affects the liquefaction risk of one area.

Panel (c) in Figure 2 shows areas with liquefaction risk in Cali. This risk is present in the east of the city because of the Cauca River[16]. For most neighbourhoods the risk is either zero or one. When a proportion of the neighbourhood is affected we use the share of the area where the risk is present. Clearly, the instrument does not vary randomly over space[17]. As shown in the estimation IV-probit we also control for firm size and the individual industrial sector and the reduction of the probability of being informal is 52 per cent[18]. Table AI-col 1 in the Appendix shows the results of the first-stage estimation. The F-statistic for weak instruments is higher than 10, which implies that the instrument has a strong negative effect on agglomeration. We confirm that firm size is negatively related to the probability of being informal, as mentioned in the literature. In our results we obtain that the IV estimated coefficient is more negative than the OLS estimated coefficient. Hence, solving the endogeneity problem, allows to estimate a stronger relationship between agglomeration and informality.

Table III shows that when formal agglomeration increases by one standard deviation the probability of being informal diminishes by 9.3 per cent[19], according to the probit model and 27.4 per cent[20] according to the IV-probit model. Table AI-col 2 in the Appendix shows the results of the first-stage estimation. These results allow us to conclude that the urban structure is determined by formal agglomeration and informal firms will occupy spaces that formal firms do not occupy. This means that informal firms have to make decisions in a different set of constraints, including those that link them to the formal sector, and supports the hypothesis that informal firms face restrictions that do not allow them to benefit from agglomeration externalities. These firms are marginalized from accessing the same set of external effects or participating in the same economic transactions as their formal counterparts (Moreno-Monroy, 2012). In short, formalization is fundamental to reap all the benefits associated with property rights.

*3.4 Sensitivity analysis*

We re-estimate the effect of agglomeration on the probability of being informal using different specifications. It is important to bear in mind that we have 338 neighbourhoods. There is extreme variation in the level of agglomeration (in log from 4.3 to 10.3). The result may be sensitive to extreme "outliers". We have estimations of the model excluding 20 neighbourhoods with the lower and higher agglomeration levels (the variation in log is now from 6.4 to 9.2). The results are similar: when the agglomeration variable increases by one standard deviation the probability of being informal diminishes by 7 per cent[21] in the probit model and 39 per cent[22] in the IV specification. Table IV shows descriptive statistics for the agglomeration variable without outliers, and Table V reports the model estimations for probit and IV-probit specification (first stage estimation of the IV model is reported in the Appendix: Table AI-col. 3).

The validity of our instrument can be questioned as it is non-random over space. We therefore re-estimate the model for observations within 1 *km* from the liquefaction limit. The panel (c) in Figure 2 shows the liquefaction limit as the division between areas affected and not affected for the liquefaction risk. The control of 1 km is in order to include in the regression analysis similar neighbourhoods where the most important difference is the condition of liquefaction risk. The results are shown in Table VI. The estimated marginal effect from the probit model is −0.1103 and −0.4197 for the IV model. These results are in the same direction compared to marginal effects in Table II, −0.1051 and −0.6953, respectively (first-stage estimation of the IV model is reported in the Appendix: Table AI-col. 4.

## 4. Conclusion

Literature provides evidence about the relationship between spatial density and aggregate increasing returns. Moreover, the structure of a city is determined by a production

| Explanatory variable | Probit | IV-probit |
|---|---|---|
| ln (formal agglomeration) | −0.1174 (0.0185)*** | −0.344 (0.041)*** |
| Size of the firm | (–) | (–) |
| Industrial sectors (47) | Yes | Yes |
| Distance from CBD | Yes | Yes |
| Distance from main corridors | Yes | Yes |
| Number of observations | 51,454 | 51,454 |
| *F*-statistic (weak instrument) | | 27.37 |

**Table III.**
Marginal effects on probability of being informal using formal agglomeration as an explanation variable

**Notes:** Marginal effects at the means of the independent variables. Clustered standard errors by neighbourhood in parenthesis; *** $p < 0.01$

| Variable | Mean | SD | Minimum | Maximum |
|---|---|---|---|---|
| ln (Agglomeration, $\delta = 2$) | 7.76 | 0.64 | 6.41 | 9.23 |
| Agglomeration, $\delta = 2$ | 2,892 | 1,966 | 608 | 10,175 |

**Table IV.**
Descriptive statistics without outliers

**Source:** Own elaboration

externality under which employment at any site is more productive when the employment at neighbouring sites is higher. As profit increases with productivity, firms should locate where their expected profit is highest. Nevertheless, when the property-rights system fails, the market cannot work efficiently, because the owner has *de facto* rights to their property but does not have a legal enforceable title.

The literature does not say too much about the structure of cities where the percentage of the informal sector is significant. In particular, we analyse the case of a Colombian city where 42 per cent of firms are informal. We shed light on how formal firms tend to be localized in high density commercial areas, while informal firms are localized in low density and peripheral areas where the land is cheaper and where they can avoid the control of authorities.

We aim to estimate the effect of formal agglomeration on the probability of being informal, which could be interpreted as a local share of informal firms. The main result is that when agglomeration increases by one standard deviation the probability of being informal diminishes around 52 per cent. This may happen because informal firms have less opportunity to benefit from agglomeration effects because of legal restrictions that block their relationship with formal firms. The result explains why formal and informal firms display different locational patterns in the urban structure. We conduct an IV analysis to tackle the potential endogeneity problem.

| Explanatory variable | Probit | IV-probit |
| --- | --- | --- |
| ln (Agglomeration) | −0.1361 (0.028)*** | −0.7564 (0.117)*** |
| Size of the firm | (−) | (−) |
| Industrial sectors (47) | Yes | Yes |
| Distance from CBD | Yes | Yes |
| Distance from main corridors | Yes | Yes |
| Number of observations | 36,914 | 36,914 |
| $F$-statistic (weak instrument) | | 16.52 |

**Notes:** Marginal effects at the means of the independent variables. Clustered standard errors by neighbourhood in parenthesis; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Table V.
Marginal effects excluding outliers

| Explanatory variable | Probit | IV-probit |
| --- | --- | --- |
| ln (Agglomeration) | −0.1103 (0.035)*** | −0.4197 (0.1419)*** |
| Size of the firm | (−) | (−) |
| Industrial sectors (47) | Yes | Yes |
| Distance from CBD | Yes | Yes |
| Distance from main corridors | Yes | Yes |
| Number of observations | 12,850 | 12,850 |
| $F$-statistic (weak instrument) | | 4.48 |

**Notes:** Marginal effects at the means of the independent variables. Clustered standard errors by neighbourhood in parenthesis; *** $p < 0.01$; ** $p < 0.05$; * $p < 0.10$

Table VI.
Marginal effects controlling the distance 1 km from liquefaction limit

## Notes

1. In developing countries, however, migration from rural to urban areas has been associated with push rather than pull factors, because the population is expelled from rural areas rather than attracted to urban areas by the prospects of better living standards (Bairoch, 1988; Barrios *et al.*, 2006).

2. As is well known, the literature focuses on technological spillovers, labour pooling and intermediate input linkages (Marshall, 1890; Ellison *et al.*, 2010), and sharing, matching and learning effects, as in Duranton and Puga (2004).

3. There are different perspectives to analysing informality. Dualism assumes that informal firms do not have linkages with formal firms. Structuralism assumes that formal and informal firms are intrinsically linked (formal firms aim to reduce their input costs by promoting informal activities). Legalism focuses on the regulatory environment of the relationship between formal and informal firms (Chen, 2006; Perry *et al.*, 2007).

4. For example, around 57 per cent of informal firms in Cali do not carry accounting (Santa Maria and Rozo, 2009).

5. The Alonso–Mills–Muth model has served as the most important base to analyse urban spatial structures but assumes employment clustering in a city centre (Alonso, 1969; Mills, 1967; Muth, 1969).

6. The agglomeration effect, $z_s$, is calculated using the employment at neighbouring locations. The function is assumed linear and decays exponentially at a rate $\delta$ with the distance from s.

7. The bid rent is defined as the rent per unit of land that a firm will be willing to pay.

8. Figure A1 in the Appendix shows the geographical location of the city.

9. At the moment of collecting information, business owners are informed that if they declare that the business fails to comply with legal regulations, they would not experience any negative legal consequences (the information provided is confidential). Informality is not persecuted, which leads business owners to provide truthful information.

10. Business informality is closely related to labour informality. Informal workers are characterized by lower levels of education and wages.

11. Fujita and Ogawa (1982) define the externality effect as a potential of employment.

12. Figure A2 in the Appendix shows the map for job and population density (per $km^2$) and Figure A3 shows cumulative population and employment (formal and informal) as a function of distance from the city centre in kilometres.

13. The parameter $\delta$ can be estimated with non-linear regression estimation procedures, but this is cumbersome.

14. Figure A4 in the Appendix shows that the logarithm of agglomeration is approximately normally distributed.

15. $-0.1051$ [ln(3648 + 4020) − ln(3648)] = $-0.078$. We multiply the change of one standard deviation on agglomeration variable by the estimated marginal effect.

16. It is one of the main rivers of Colombia measuring some 1,350 km in length.

17. A sensitivity analysis about this fact is performed in section 2.4.

18. $-0.6953$ [ln(3648 + 4020) − ln(3648)] = $-0.52$. We multiply the change of one standard deviation on agglomeration variable by the estimated marginal effect.

19. $-0.1174$ [ln(2749 + 3343) − ln(2749)] = $-0.093$.

20. $-0.344$ [ln(2749 + 3343) − ln(2749)] = $-0.274$.

21. The estimated coefficient is $-0.1361$. Then, $-0.1361\,[\ln(2892 + 1966)\,-\ln(2892)] = -0.07$.
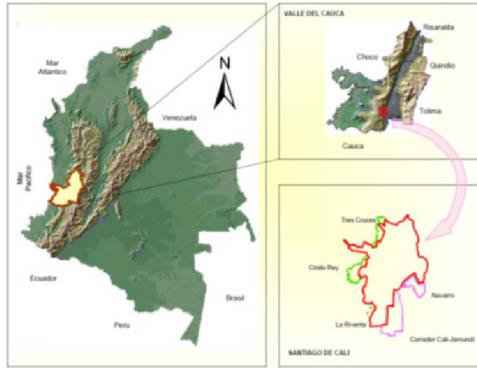22. The estimated coefficient is $-0.7564$. Then, $-0.7564\,[\ln(2892 + 1966)\,-\ln(2892)] = -0.39$.

## References

Alonso, W. (1969), *Location and Land Use: Towards a General Theory of Land Rent*, Harvard University Press, Cambridge, MA, available at: www.hup.harvard.edu/catalog.php?isbn=9780674730854

Bairoch, P. (1988), *Cities and Economic Development: From the Dawn of History to the Present*, The University of Chicago Press, Chicago, available at: www.press.uchicago.edu/ucp/books/book/chicago/C/bo5971175.html

Barrios, S., Bertinelli, L. and Strobl, E. (2006), "Climate change and rural-urban migration: the case of sub-Saharan Africa", *Journal of Urban Economics*, Vol. 60 No. 3, pp. 357-371.

Bayer, P. and Timmins, C. (2007), "Estimating equilibrium models of sorting across locations", *The Economic Journal*, Vol. 117 No. 518, pp. 353-374.

Brulhart, M. and Mathys, N. (2008), "Sectoral agglomeration economies in a panel of European regions", *Regional Science and Urban Economics*, Vol. 38, pp. 348-362.

Chen, M. (2006), "Rethinking the informal economy: linkages with the formal economy and the formal regulatory environment", in Basudeb, Ravi and Ostrom (Eds), *Linking the Formal and Informal Economy*, Oxford University Press.

Ciccone, A. (2002), "Agglomeration effects in Europe", *European Economic Review*, Vol. 46 No. 2, pp. 213-227.

Ciccone, A. and Hall, R. (1996), "Productivity and the density of economic activity", *American Economic Review*, Vol. 86 No. 1, pp. 54-70.

Combes, P.P., Duranton, G., Gobillon, L. and Roux, S. (2010), "Estimating agglomeration effects with history, geology and worker fixed effects", in Glaeser, E. (Ed.), *Agglomeration Economies. Chicago*, University of Chicago Press, IL.

Combes, P.P. and Gobillon, L. (2015), "The empirics of agglomeration economies", in Duranton, Henderson and Strange (Eds), *Handbook of Urban and Regional Economics*, Vol. 5, Elsevier.

deSoto, H. (2000), *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*, Basic Books, New York, NY, available at: www.amazon.es/Mystery-Capital-Capitalism-Triumphs-Everywhere/dp/0465016154

Duranton, G. (2009), "Are cities engines of growth and prosperity for developing countries?", in Spence, A. and Buckely (Eds), *Urbanization and Growth*, The Commission on Growth and Development, Washington, DC.

Duranton, G. and Puga, D. (2004), "Micro-foundations of urban agglomeration economies", in Henderson and Thisse (Eds), *Handbook of Urban and Regional Economics*, Vol. 4.

Ellison, G. and Glaeser, E. (1999), "The geographic concentration of industry: does natural advantage explain agglomeration?", *American Economic Review*, Vol. 89 No. 2, pp. 311-316.

Ellison, G., Glaeser, E.L. and Kerr, W.R. (2010), "What causes industry agglomeration? Evidence from Coagglomeration patterns", *American Economic Review*, Vol. 100 No. 3, pp. 1195-1213.

Feige, E. (1990), "Defining and estimating underground and informal economies: the new institutional economics approach", *World Development*, Vol. 18 No. 7, pp. 989-1002.

Fujita, M. and Ogawa, H. (1982), "Multiple equilibria and structural transition of non-monocentric urban configurations", *Regional Science and Urban Economics*, Vol. 12 No. 2, pp. 161-196.

Koster, H. and Rouwendal, J. (2013), "Agglomeration, commuting costs, and the internal structure of cities", *Regional Science and Urban Economics*, Vol. 43 No. 2, pp. 352-366.

Lucas, R. and Rossi-Hansberg, E. (2002), "On the internal structure of cities", *Econometrica*, Vol. 70 No. 4, pp. 1445-1476.

Maloney, W. (2004), "Informality revisited", *World Development*, Vol. 32 No. 7, pp. 1159-1178.

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

Melo, G. and Noland, R.B. (2009), "A meta-analysis of estimates of urban agglomeration economies", *Regional Science and Urban Economics*, Vol. 39 No. 3, pp. 332-342.

Mills, E. (1967), "An aggregative model of resource allocation in a metropolitan area", *The American Economic Review*, Vol. 57 No. 2, pp. 197-210.

Moreno-Monroy, A. (2012), "Critical commentary. Informality in space: understanding agglomeration economies during economic development", *Urban Studies*, Vol. 49 No. 10, pp. 2019-2030.

Moreno-Monroy, A. and Roman, F. (2015), "The impact of public transport expansions on informality: the case of the São Paulo metropolitan region", *ERSA conference papers*, European Regional Science Association.

Moreno-Monroy, A. and García, G. (2015), "Intra-metropolitan agglomeration of formal and informal manufacturing activity: evidence from Cali, Colombia", Working Papers, Universitat Rovira i Virgili, Department of Economics.

Morikawa, M. (2011), "Economies of density and productivity in service industries: an analysis of personal service industries based on establishment-level data", *Review of Economics and Statistics*, Vol. 93 No. 1, pp. 179-192.

Muth, R. (1969), *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*, University of Chicago Press, Chicago and London.

Perry, G.E., Maloney, W.F., Arias, O.S., Fajnzylber, P., Mason, A.D. and Saavedra-Chanduvi, J. (2007), *Informality: Exit and Exclusion*, World Bank, Latin American and Caribbean Studies, Washington, DC, available at: https://openknowledge.worldbank.org/handle/10986/6730

Rosenthal, S. and Strange, W. (2003), "Geography. Industrial organization and agglomeration", *The Review of Economics and Statistics*, Vol. 85 No. 2, pp. 377-393.

Rosenthal, S. and Strange, W. (2008), "The attenuation of human capital spillovers", *Journal of Urban Economics*, Vol. 64 No. 2, pp. 373-389.

Schneider, F. (2005), "Shadow economies around the world: what do we really know?", *European Journal of Political Economy*, Vol. 21 No. 3, pp. 598-642.

Sindzingre, A. (2006), "The relevance of the concepts of formality and informality: a theoretical appraisal", in Basudeb, Ravi and Ostrom (Eds), *Linking the Formal and Informal Economy*, Oxford University Press, available at: www.oxfordscholarship.com/view/10.1093/0199204764.001.0001/acprof-9780199204762

Sock Yogo. (2005), "Testing for weak instruments in linear IV regressions", in Andrews and Stock (Eds), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge University Press, pp. 80-108, available at: www.cambridge.org/core/books/identification-and-inference-for-econometric-models/76B7B5986247FC8CC13C9D8C7E02DCBF
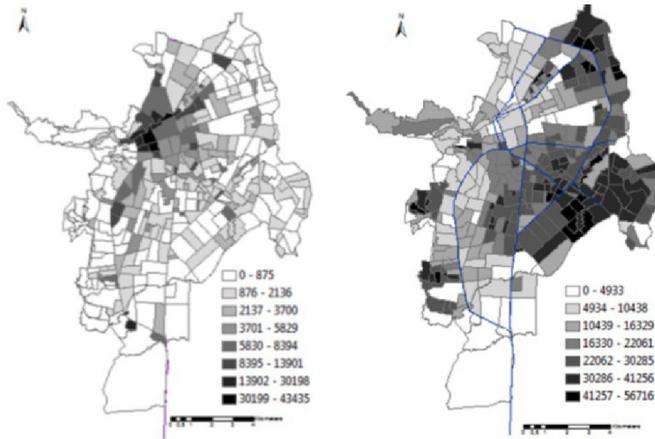
## Further reading

Cárdenas, M. and Mejía, C. (2007), "Informalidad en Colombia: Nueva evidencia", Fedesarrollo. Working Papers Series. No. 35.

Posada, E. and Mejia, D. (2007), "Informalidad: Teoría e implicaciones de política", *Borradores de Economía*, Vol. 455, Banco de la República, Bogotá – Colombia.

Santa María, M. and Rozo, S. (2009), "Qualitative and quantitative analysis of firm informality in Colombia", *Desarrollo y Sociedad*, pp. 269-296, available at: www.researchgate.net/publication/262449637_Qualitative_and_quantitative_analysis_of_firm_informality_in_Colombia
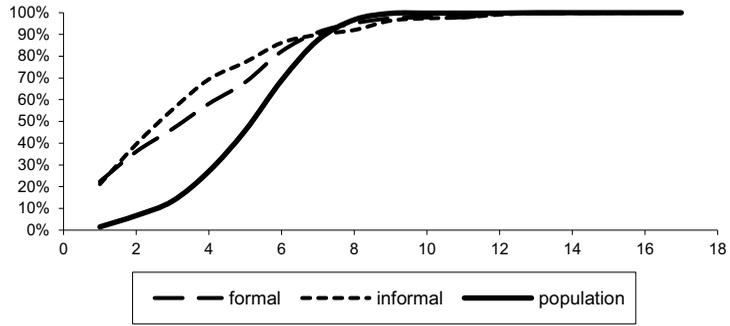
**Appendix**



**Source:** Colombian geological survey

| 0 - 875 |
| 876 - 2136 |
| 2137 - 3700 |
| 3701 - 5829 |
| 5830 - 8394 |
| 8395 - 13901 |
| 13902 - 30198 |
| 30199 - 43435 |

| 0 - 4933 |
| 4934 - 10438 |
| 10439 - 16329 |
| 16330 - 22061 |
| 22062 - 30285 |
| 30286 - 41256 |
| 41257 - 56716 |

**Source:** Census 2005, DANE and own elaboration

**Figure A2.**
Jobs and population
density per square
kilometre by
neighbourhood
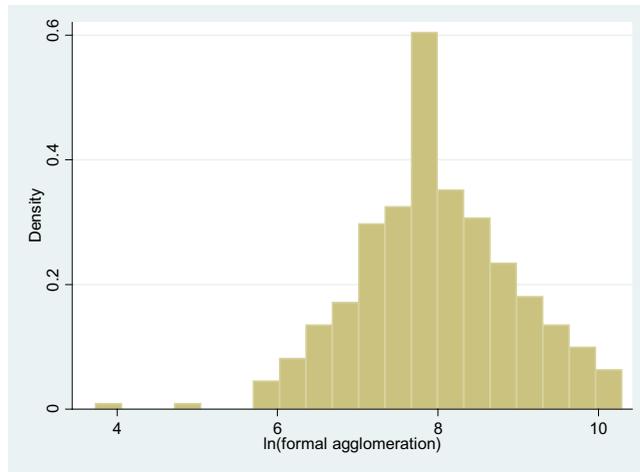
**Source:** Own elaboration

**Source:** Own elaboration
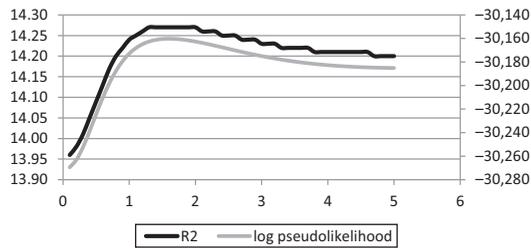
**Source:** Own elaboration

| | (1) Total Agglomeration | (2) Formal Agglomeration | (3) Total Aggl. No outliers | (4) Total Aggl. Control (+) |
|---|---|---|---|---|
| Liquefaction | −0.2269 (0.0573)*** | −0.5192 (0.0672)*** | −0.2144 (0.056)*** | −0.2968 (0.066)*** |
| Size of the firm | (+) | (+) | (+) | (+) |
| Industrial sectors (47) | Yes | Yes | Yes | Yes |
| Distance form CBD | Yes | Yes | Yes | Yes |
| Distance from main corridors | Yes | Yes | Yes | Yes |
| Adj. $R^2$ | 0.91 | 0.90 | 0.77 | 0.77 |
| $F$-statistic (weak instrument) | 18.81 | 27.37 | 16.52 | 4.48 |
| Number of observations | 51,454 | 51,454 | 36,921 | 12,850 |

**Notes:** Standard errors clustered by neighbourhood in parenthesis; *** $p < 0.01$. (+) controlling the distance 1 km from liquefaction limit

**Table AI.**
First-stage
estimations

**Corresponding author**
Andres Dominguez can be contacted at: jandresdominguezm@gmail.com