URBAN EPIDEMIOLOGY OF LABOR INFORMALITY IN BOGOTÁ: A NEIGHBORHOOD INTERACTIONS APPROACH[1]

Alberto Montoya Cruz

Advisor: Dario Maldonado

Abstract

Using geo-referenced data on labor outcomes for the city of Bogotá, we develop a strategy to identify the effect of the neighborhood informality rate on the individual probability of getting an informal job. We found evidence of the existence of such neighborhood effects. Those effects operate differently for salaried informality than for self-employed informality.

Keywords: labor informality, social interactions, neighborhood effects

1 Introduction

Labor informality is widely recognized as a major concern of most Latin American economies. And as many of contemporary social concerns, our understanding of its mechanisms is still precarious. However, economic research has made significant progress explaining its causes and consequences: we are closer than ever before to identify the features of the economic organism causing this alteration in the labor relations. Like physicians treating patients, empirical and theoretical literature has focused on the study of the body functioning. Our questions are mainly about the features of firms and workers that make them informal. But, so far, we are putting aside the lesson that doctors learned during the 19th century, when epidemiology appeared: find the environmental sources!

That will be our task in this paper, to assess the social channels that make possible the transmission of informality in the city, and construct a first approximation to a social urban epidemiology of labor informality. Disregarding the environmental influences affecting labor informality can make us consider labor informality as an immutable trait of certain firms or certain workers. On the contrary, we must be willing to accept that there may be a path

[1] The research leading to this thesis report was made possible through the continuous support of CEIBA (Centro de Estudios Interdisciplinarios Básicos y Avanzados). I would like to thank professors Dario Maldonado, Juan Miguel Gallego and Darwin Cortés for their decisive support through this process; professors Javier Pérez and Juan Carlos Guataquí for their insightful comments; the group of research fellows in Economics School of Universidad del Rosario, particularly Juan Sebastián Ordóñez, Jacobo Rozo and Jeisson Cárdenas; and my parents, Nancy and René, and my grandparents, Sagrario and Gumercindo. All errors remain mine.

leading them to this pattern of labor relations. And one of the possibilities is that this path is the job search process.

The characteristics of the jobs that people have depend on the characteristics of the jobs that people *can get*. Then, we have to take a closer look at the way people seek jobs. There is evidence to conclude that an important share of the active population seek jobs through non-formal channels, like asking for help from relatives, friends and neighbors. So, the features of the jobs that this social network can offer will influence the employment status of the job seeker. This is the source of infection, the social interaction with people who has already some contact with the informal sector.

A growing body of literature have pointed that social distance is correlated with geographical distance, in the extent that a person is more likely to interact with someone nearby that with somebody far away. Then, we can analyze the role of the neighborhood in the searching process; this is one of the environments in which the worker can find a job. This approach brings us two advantages: first, the availability of data about this particular social context for worker in the city of Bogotá; and second, the possibility of putting our urban epidemiology analysis on a spatial basis.

In fact, the possibility of a neighborhood component in the likelihood of being an informal worker adds a new dimension to this social problem: the segregation. The place of residency, in this way, can affect the hire prospects, causing a sort of poverty trap. Those who live in neighborhoods with more contact with the informal sector will face slightly higher chances to find an informal job and slightly higher difficulties to get a formal job. The results of this process are the appearance of neighborhoods with bigger prevalence of informal jobs, perpetuated in time, and some spatial concentration of informality across the city.

These emergent results are common in situations of 'social interactions'. Segregation, social multipliers and spatial correlations are some of the empirical regularities that allow us to identify this kind of phenomena. We will take advantage of this to find a suitable methodology to isolate the effect of social interactions (at the neighborhood level) on the job search process, and particularly in the interaction of the search process and the entry in the informal sector. There is not an established and widely recognized method for this analysis, so one of the main purposes of this paper will be the comparison of different methodologies of the 'social interactions' literature in the labor market context.

The remainder of the paper is organized as follows. Section 2 reviews some related work. In section 3 we describe the dataset used, the adequate definition of informality for our purposes and the constructions of the variables characterizing the neighborhoods. Section 4 describes labor informality in Bogotá in terms of job-search channels and spatial location. Section 5 presents some of the difficulties for the estimation of neighborhood effects and their potential solutions. Section 6 presents the estimations of neighborhood effects. Section 7 concludes.

2 Labor informality, job search and neighborhoods effects

The idea that labor relations are undertaken in an 'informal sector' arises from the study of urban labor markets in Africa by Keith Hart (1970; 1973) and ILO (1972). They observed that an important portion of the labor is not traded in formal markets, and that many people are self-employed and work mainly in subsistence activities. It was necessary to explain the existence and operation of what had been so far a dark side of the work relations, and this explorations result in the appearing of the dualist labor market approach (Lewis, 1954; Harris & Todaro, 1970; Fields, 1975; Piore, 1983): the economy of these developing countries is segmented into a modern sector, satisfying the established labor regulations, and a pre-modern sector. Processes such as migration from the countryside to the city increase the labor supply, and the modern sector of the economy does not expand fast enough to incorporate them all. As a result, some workers are excluded, having to work in the pre-modern sector; that is, the informal sector. Tokman (1992) shows that this sector is characterized by lower levels of productivity and lower capital accumulation, which in turn will translate into lower income levels.

These explanations have in common a negative view of informality (Portes & Haller, 2004); it is seen only as a forced alternative to the narrowness of the modern segment of the labor market – and therefore as something involuntary (Atuesta, 2010). Other approaches suggest that informality might be a choice for small business owners who find it as a profitable strategy. Perhaps the first to explore this path was Hirschman (1970), who argues that, given the state's impossibility to provide appropriate enforcement, small firms are more likely to evade labor regulations. De Soto (1989) asserts that state norms generate privileges and immobilize the economy, and labor informality is just a response to this, the return of the real economy, when the individual expects bigger benefits than costs from avoiding state's regulations. Further studies (Feige, 1990; Portes, Castells, & Benton, 1989; Pratap & Quintin, 2006; Quintin, 2008) have contributed to understand informality as a matter of institutions. Maloney (2004), comparing several Latin-American countries, has suggested presence of large informal sectors in countries with more flexible labor markets. This suggests that informality could be a matter of unregulated micro-entrepreneurship activities.

A synthesis of these two perspectives can be found in a document by the World Bank (Perry, 2007), which recognizes the heterogeneity of the informal sector. There are disadvantaged workers who are excluded into informality; they would prefer to work in the formal sector but can't because of the rationing of jobs. But there are also workers who voluntarily evade the formal sector; they found better incomes and better working conditions (especially flexibility) in the informal sector.

Despite the efforts to understand why some workers enter the informal sector, there is still much to explore on this topic. Maybe one of the disregarded aspects in informality research is the consequences of social contacts on the job matching process. A growing body of literature has sought to demonstrate that references play an important role, either because they transmit a signal to potential employers or because they provide access to information on vacancies. Montgomery (1991) shows that references are mechanisms to classify applicants according to their productivity – as long as employers think that highly productive workers refer other highly productive workers. Pellizarri (2010) shows that firms invest more resources in recruiting programs when the vacancy is for a high productivity job, and they are more likely to use references if the vacancy is for a low productivity job. The, the job search channel is related to productivity. Calvó-Armengol and Jackson (2004) show that job search through social contacts generate path-dependence and segmentation in the labor market. That is, two neighborhoods with similar characteristics but different in their level of employment will tend to maintain their inequality – the group with higher levels of employment improves their member's employment prospects, which in turn reinforces the group's advantage in employment levels.

Recent studies (Uribe & Gómez, 2005; Uribe, Viáfara, & Oviedo, 2009) have used data from the Quality of Life Survey 2003 in Colombia to estimate the frequency and effectiveness of use of formal and informal search channels[2]. Almost 75% of the employed population got his job with help from a relative or a friend, and 50% of the unemployed uses these contacts as their principal mean to obtain a job. Quiñonez (2011) also find that informal search channel are less effective and they are related with the rigidities of the Colombian labor market.

The lesson of this literature is that the social structure is important when analyzing the labor market structure. There are social interactions that occur outside markets, which partly explain the outcomes we observe. What have referred back as the effect of references is very similar to what in other literature is known as peer effect: in addition to individual characteristics, the characteristics of the social network or reference group are crucial; in two ways, in one hand, they determine the context in which the individual makes decisions (contextual effects), but, on the other hand, there could be a direct social interaction that makes my peers outcomes affect my outcomes (endogenous effect).

Nevertheless, the analysis of job search channel implications in regard to the informal sector is a less discussed topic. Most of the work in this subject comes from sociology. Portes and Haller (2004) states that informal economy, instead of being the 'real market economy' working, is actually a system of informal exchanges made possible by specific social ties. This results in the problem of embeddedness (Granovetter, 1985), in which the pertinence to the same social structure or social network generates the necessary trust to the functioning of the informal labor relation.

We will propose that the contact network of the individual is important to his probabilities of establishing an informal labor relation. The amount and quality of the vacancies to which a person can apply will depend, if using this search channel, on the size and characteristics of his network: if his contacts are informed of vacancies for only a certain branch of activity, or only at certain salary levels, or only from small firms that evade the regulation, then these are the positions to which he can access. So it would be reasonable to expect the existence of peer

[2] Informal job search channels are those which use personal contacts or talking with the employer. Its definition has nothing to do with the definition of labor informality.

effects in informality. If one belongs to a social network in which informal employment predominates, then most of the information one would get about vacancies would be in the informal sector.

Sometimes we cannot observe the whole social network of individuals, but just a fraction. The neighborhood is one of these cases. Labor market survey usually allows to observe at least a fraction of the vicinity of each respondent, and there is evidence that neighborhoods are relevant in the social networks. Wellman (1996) finds, for the city of Toronto, that 42% of the yearly social contacts with one's social networks happen with people living less than a mile away. Ontari (1999) finds, for the United States, that nearly 20% of the components of the social network are physical neighbors. Lee and Campbell (1999) look at "micro-neighborhoods" for the city of Nashville, defined as groups of 10 adjacent housing units on the same block[3]; they find that almost one third of the neighbors are considered close, and about 13% of the people that a person can contact to look for a job live in these micro-neighborhoods.

Some papers have taken advantage of the geographical distance and neighborhoods structure as a way to make observable at least fraction of the social network in which the individual is immersed (See Durlauf (2004) for a review on this). Topa (2001) and Conley and Topa (2002) analyze the geographic patterns of unemployment, and found spatial dependence. This is, the employment rate in certain place shows a close correlation with employment rates in surrounding places, and clustering conditions exist: there are places with concentration of high employment rate, and places of concentration of low employment rates. This is consistent with models of information exchange through local networks. Bayer, Ross and Topa (2008) extend this analysis, using the workplace in addition to place of residence, to find more robust results. Ioannides and Loury (2004) extend the neighborhood effect analysis to other labor outcomes.

There are two pending task with which this paper will deal. First, the analysis of the potential effects of job search channels on the size and distribution of the informal sector is still to be done. And second, the analysis of the consequences in the labor market of urban neighborhood interactions is a topic that hasn't been studied in Colombia, and the conclusions for the labor market in a Latin American country could be different from what has been found in studies done, most of the times, for the United States.

3 Labor Informality in Bogotá

3.1 Definition of informality

The purpose of this paper is to find out what is the effect of the characteristics of the neighborhood on the characteristics of the jobs that people have. Then, we should use a definition of labor informality that focuses on the quality of the job. Also, we have data about

[3] As we will see, this definition of micro-neighborhood is very similar to the DANE's definition of segment, so there is an analogy between the spatial framework of Lee and Campbell and ours.

the labor supply, so it does make sense to use a definition of informality designed for labor demand (firm's information). The usual choice in this case is a definition of informality relying on social security protection. Guataquí, García, & Rodríguez (2011) propose a 'strong' definition of informality, which recognizes the difference between the institutional environment for wage-earners and independent workers, and for the former includes additional institutional features, as the existence of a written labor contract and the compliance of minimum wage.

A wage-earner or a domestic worker is considered a formal worker if he satisfices the following conditions: a) he is contributing to health insurance, b) he is contributing to a pension scheme or is a pensioner, c) he has a written labor contract, and d) he earns more than 95% of minimum a wage per hour. An independent worker is considered formal if he satisfices the following conditions: a) he is contributing to health insurance, b) he is contributing to a pension scheme or is a pensioner. An individual is considered informal worker if he has a job but is not a formal worker.

We also show, in the appendix, the estimations with two other definitions of labor informality: the 'weak' definition of labor informality (by Guataquí, García and Rodríguez), which points for the minimum level of acceptable worker's protection in a labor relation; and the PREALC-DANE definition, in which the main characteristic of the informality is the size of the firm.

Table 1: Informality rate

	Rate	**SD**
Strong definition	0.61	0.49
Weak definition	0.22	0.42
PREALC-DANE definition	0.46	0.50

Source: Own calculations, with information from DANE

3.2 Spatial patterns in labor informality in Bogotá

Labor informality is a relevant concern from an academic and form a normative point of view. It is a normative issue (on a welfare perspective) because a significant portion of the urban labor force works in worse conditions than those that are considered by the Government as the minimal requirements of the labor relations. Formal regulations do not cover these jobs, bringing on situations of vulnerability, like the lack of insurance against the risks of unemployment, health or aging. Additional to these drawbacks, informal hobs are also correlated with lower productivity, lower incomes and tax evasion. And labor informality is an academic issue to the extent that many of his mechanism and causes remain unexplored.

About 61% of the workers in Bogotá are in the informal sector, using the 'strong definition' of informality as proposed by Guataquí, García and Rodríguez (2011). This implies that only 39% of the labor relations satisfy the fundamental conditions of social protection of the worker. The weak definitions point out that 22% of the labor relations doesn't even the minimum acceptable level of protection.

But informal and formal workers are not only different in the conditions of their jobs, but also in the way they search for them. Table 2 shows the job search channels for formal and informal wage-earners[4].

Table 2: Search channel

	Formal	Informal	Total
Family and acquaintances	57.9	87.6	69.2
Sent CV to firm	20.6	6.7	15.4
Employment agency	7.1	1.2	4.9
Job ads	3.5	1.5	2.8
Job calls	7.9	0.8	5.2
SENA employment agency	1.2	1.1	1.2
Others	1.7	1.0	1.5

Source: Own calculations, with information from DANE

This shows the importance of taking into account the social interactions effects in the labor market. Almost 70% of the employees found his current work by appealing to a relative, a friend or an acquaintance. Then, worker's social group is probably important in the definition of what kind of job he can obtain. Even though relatives and acquaintances are prevalent in the search process for both formal and informal workers, the channel composition is not the same. Almost 60% of those who got a formal job were helped by his social group, and one in five got it sending his resume to the employer or an intermediary. For those who got an informal job, almost 90% were helped by the social group in the search process. All the other search methods seem to be nearly irrelevant for getting an informal job.

People get informal jobs through social contacts. Someone in one's the social network makes a referral or gives information about job openings and in this way he helps to match employer and employee. From the employer's point of view this way of looking for workers can be useful to avoid government surveillance. But that may also be the result of the strategic behavior of a firm. Pellizzari (2010) shows that a firm prefers to invest in a recruiting program when the vacancy is for a high productivity task, and high productivity employees tend to be in the formal sector. In the other hand, the firm prefers to use references if the vacancy is for a low productivity job, and low productivity is associated with informality.

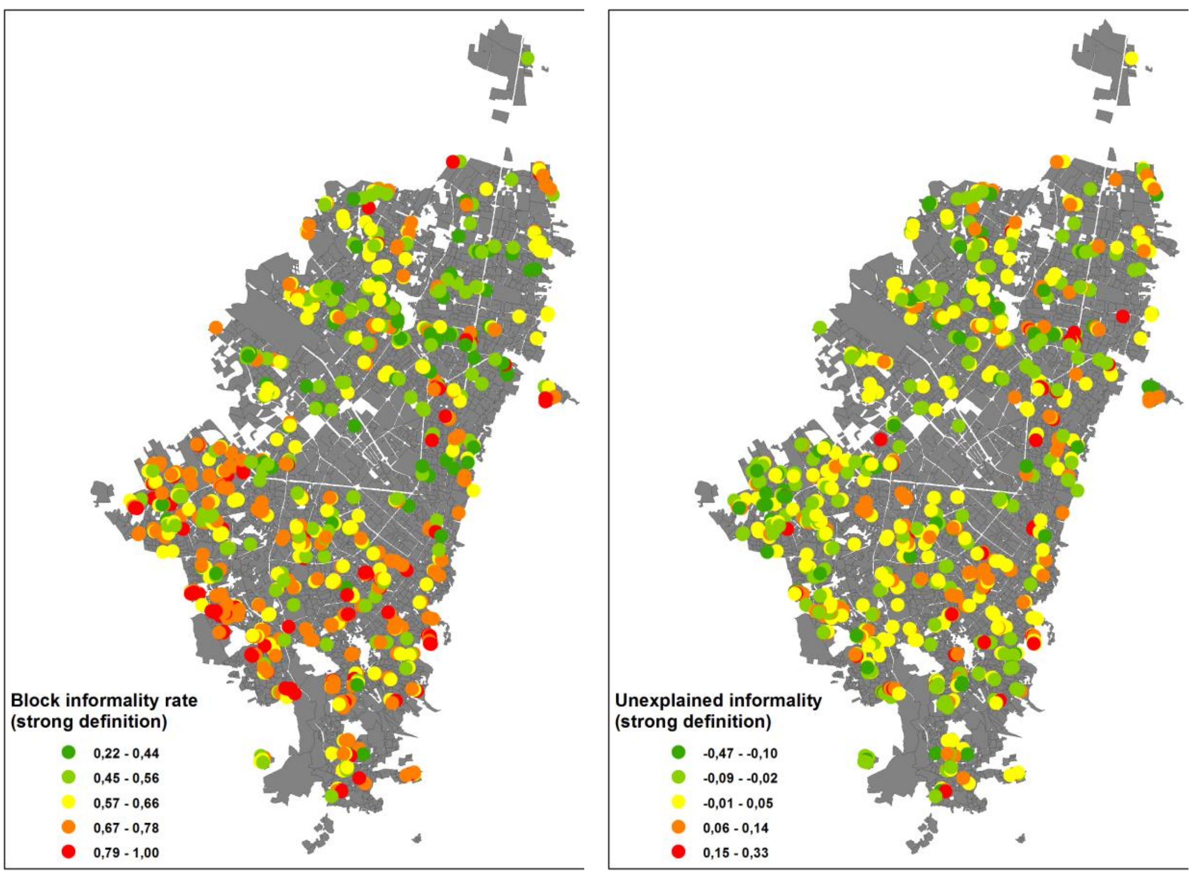
Then, the chances of getting an informal job would depend on the information sets of the people in the social group. And their information sets are likely to be associated with their labor context: if most of them work in informal activities, they are most likely to hold information about informal jobs. If this is so, a person with a lot of relatives and friends in the

[4] The exclusion of the independents is a big deal, as they are important in the composition of informality, as we just see. But the information about the search channel is available only for employees.

informal sector will find easier to obtain an informal job. We will observe an association between the individual probability of entering the informal sector and the labor informality rate in his social context. The nature of the job he is likely to find would depend on the people he meets, and the nature of their employments.

The main problem with this type of hypothesis is that in most of the cases the social group is not observable. But there is a social context in which valuable exchange of information can be done: the neighborhood. If we accept that some of the individual's social group can be found in his vicinity, then we can hypothesize that a neighborhood with higher prevalence of informality will increase the probability of his unemployed workers of finding an informal job. And this will generate a segregation process: neighborhoods with higher informality rates will tend to remain informal, and neighborhoods with lower informality rates will tend to remain formal. This is one of the features of the labor market in Bogota: informality is geographically concentrated, and apparently there is some kind of job segregation process. The left panel in Figure 1 shows the spatial distribution of informality in the city. It is remarkable that there are high concentrations of formal employment in the north of the city and accumulation of informal employment south and southeast of the city.

Figure 1: Informality distribution in Bogotá



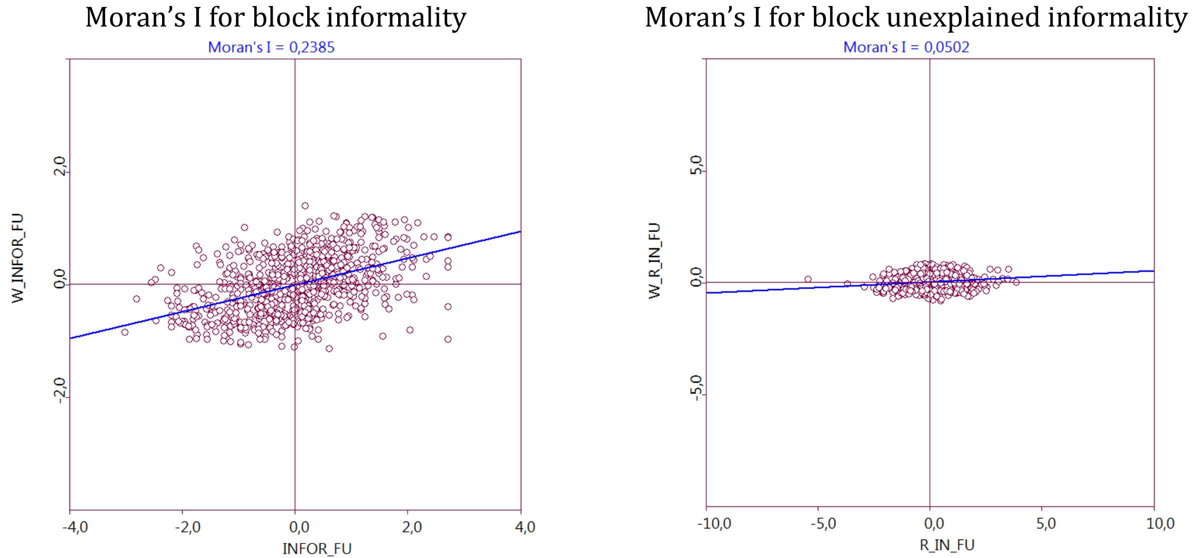
The map in the left depicts the informality rate by block for the respondents of the GEIH between 2009 and 2011. The map in the right depicts the unexplained informality rate as the residuals from a linear regression of block informality rate and certain socio-economics characteristics of the block (age, gender composition, education, family structure, occupation distribution and economic activity composition).
Source: Own calculations, with data form IDECA and GEIH (DANE).

But this agglomeration of labor informality can be caused, in part, by other kinds of segregations: maybe people with lower productivity (due to age, education, etc.) tend to live in the same places, or maybe some occupations or economic activities with higher propensity to informality tend to locate together. Since some of these neighborhood variables are observable, we can try to depict the neighborhood's informality which is not explained in this way. This is what the second panel of figure 1 shows. Most of the segregation seems to have vanished. Nevertheless, this is not a strong argument against the existence of social interactions effects at neighborhood level, but a consequence of one of the difficulties in the estimation of such effects: the reflection problem – it is hard to isolate the effect of the informality rate (the endogenous variable) from the effect of other characteristic of the context (exogenous variables).

We can get an insight of the agglomeration of informality using a spatial auto-correlation statistic, specifically the Moran's I. The closer it is to 1, the higher is the correlation between one block's informality rate and the adjacent blocks' rates[5]; if the statistic is around 0 then there is no correlation between a block's informality rate and its surroundings', the informality among the city is randomly distributed; and if the statistic is near -1, there is a negative association, blocks with high informality rates tend to be surrounded by block with low informality rates. Figure 2 show an important spatial correlation, about 0.24. It indicates that neighborhoods with high informality tend to be surrounded by other neighborhood with high informality. That is consistent with a patron of urban segregation of informality. But, when we make the same analysis with the unexplained informality the correlation is positive but much lower. Again, because of the reflection problem, we cannot rule out the special association. The next section presents some methodologies to overcome these problems.

[5] We use blocks because it is the more precise location information available in our data.

Figure 2: Spatial Auto-correlations – Moran's I



The spatial weights matrix is constructed using the 20 nearest neighborhoods of each spatial unit.
Source: Calculations in GeoDa, with data form IDECA and GEIH (DANE).

The neighborhood social interaction regarding labor informality outcomes also would have influence in terms of variance; it would look as a conformity effect: people tend to get similar jobs as their neighborhood, inducing homogeneity between inhabitant of the same vicinity, and heterogeneity between distinct vicinities. Table 3 shows the variance decompositions for different definitions of neighborhood. It shows the standard deviation of informality status (0 formal, 1 informal), the standard deviation across neighborhoods (SD between) and the standard deviations inside the neighborhood at the individual level (SD within). There is an important dispersion across neighborhoods, in part unexplained in terms of individual characteristics. We observe greater *between* dispersion with the smallest definition of neighborhood – a standard deviation of 13.6 points in the informality rate across neighborhoods. The city is more homogenous in terms of the weak definition of labor informality, and the heterogeneity in term of the strong definition of informality or the firm-size definition is about the same.

Table 3: Variance decomposition of the informality rate

Informality definition	SD	SD between	SD within
Strong	0.488	0.136	0.472
Weak	0.416	0.081	0.411
PREALC-DANE	0.498	0.141	0.482

Source: Own calculations, with data form IDECA and GEIH (DANE).

The evidence reviewed so far indicates the existence of a relation between the location of workers among the city and their likelihood to be employed in the informal sector. However, this is not enough to establish the existence of a neighborhood effect of the informality rate. Geographical variance could be explained by the variance in neighborhood characteristics, or by the non-random sorting of households (the choice of neighborhoods by families is correlated with unobserved characteristics)[6]. We will show that, even after controlling for these two sources of bias, the neighborhood-level peer effect is important.

4 Empirical methodology for the study of social interactions

We have suggested before that the probability of getting certain type of job could be affected by the prevalence that type of job among the reference group of the individual. This is consistent with a common observation in sociology, in the sense that the reference group to which a person belongs can affect the decisions that he takes, or the outcomes that he gets. This is an interaction outside the labor market, which we can study in the framework of *social interactions*.

> "Social interactions arise when individuals (or house-holds) affect each other's decisions, preferences, information sets, and outcomes, directly rather than indirectly through markets. These interpersonal effects are known as endogenous social effects when own decisions and those of others in the same social milieu are inter-dependent" (Ioannides & Topa, 2010)

These endogenous effects induce some empirical regularities that give us clues about their presence. First, the interdependence between individual outcomes and group outcomes can cause a positive correlation: the environmental influence make the individuals more likely to be similar to their group. At the end, we will observe homogeneity between individuals of the same group, but heterogeneity between different groups. In a spatial framework, this would mean identifiable patterns of segregation.

Additionally, endogenous social effects can induce *social multipliers*. A change in a relevant characteristic has a direct effect on individual outcomes, but also an indirect effect on the peers through the social interaction effect; and the shift in the peer's outcome generates in turn a feedback. In this way the direct effect is amplified; this is the social multiplier, and it will cause a variance not explained by individual characteristics (Ioannides & Topa, 2010). It is also possible for endogenous effects, under certain circumstances, to generate multiple equilibriums, which is also reflected in the variance not explained by individual characteristics (Durlauf, Neighborhood effects, 2004).

The former empirical regularity –the individual-group correlation– has allowed the development of methodologies of linear regression analysis to assess the existence and magnitude of endogenous social effects[7]. The latter empirical regularity –the social multiplier– has allowed the development of methodologies of variance contrast (Glaeser, Sacerdote, &

[6] Families are not distributed randomly in the city, the residence place is an economic decision which is influenced by several factors, and what we identify as a neighborhood effect may simply reflect that people tend to live around similar people.

[7] In an urban framework, like the one we will use, a common strategy is to use spatial correlation analysis. Another possibility, which suitability we will discuss in the next section, is to use group effects models.

Scheinkman, 1996; Glaeser, Sacerdote, & Scheinkman, 2003; Graham, 2008). These methods employ different informational assumptions, the first using exclusion restrictions and the second using covariance restrictions on the error term, but are at the end two sides of the same coin (Durlauf & Tanaka, 2008). The decision of use one or the other depends on the other features of the identification strategy. In particular, we will see that the solutions to the most prevalent problems in the identification of social interactions effects have been built on the regression analysis framework.

However, the identification of social interactions effects has additional complexities. In particular, two main issues arise in the estimation: the reflection problem (the confusion between endogenous effects and other group effects) and the selection bias (the effect of variables associated with the neighborhood selection). To describe those problems and show the solutions offered by the literature, we will built on the linear-in-means model (Moffitt & others, 2001).

4.1 Estimation issues

Part of the identification problem of social interactions is that there are several reasons explaining why individual in the same reference group (i.e. they have the same peers, or belong to the same neighborhood) tend to behave in similar ways. Manski (1993) classify them in three categories:

- The individual's outcome is influenced by the group's outcome (observed or expected). This is called the *endogenous effect* – the endogenous variable for the individual is affected by the endogenous variable at the group level.
- The individual's outcome is influenced by the group's characteristics. This is called the contextual effect, because the conditions of the reference group are the context in which the individual makes choices or gets outcomes. It is also called the *exogenous effect* – the endogenous variable is affected by the exogenous variables at the group level.
- The individual's outcome is influenced by attributes in which he is similar to his reference group, or he faces a similar institutional environment to the rest of the group. This is called the *correlated effect*.
  Sorting is a common type of correlated effect, where the group composition is correlated with certain attribute which also affects the outcomes of individuals, generating a correlation between individual and group outcomes. This is not a form of group-influence, but might be mistakenly seen as such if the sorting attribute is not observed. Correlated effects can also be caused by common shocks; in this case the correlation between the individual outcome and the group outcome is produced by an external influence shared by the reference group members.

The linear-in-means model is the simplest way to formalize those effects. A naïve model, assuming the lack of social interactions spillovers, would describe individual's outcome in terms of individual characteristics:

$$y_{ir} = \alpha + X_{ir}\eta + u_{ir}$$

Where y_{ir} is the outcome of the *i-th* individual, who is member of the *r-th* group and X_{ir} in a vector of individual characteristics. This is the approach of the papers explaining the probability of being informal with individual variables like education, experience, etc. A first modification social interaction analysis would propose to this model is the inclusion of contextual variables, to assess the effect of the environment on the individual's outcome. This model would take the form:

$$y_{ir} = \alpha + X_{ir}\eta + E_r(Z)\gamma + u_{ir}$$

Where Z_{ir} is a vector of individual characteristic which are relevant in the characterization of the group. $E_r(Z)$ is the vector of these characteristics aggregated at the level of group, so γ tell us about the impact of these reference group's qualities on individual's outcome. Some of the variables that compose X_{ir} may be as well in Z, but both sets of variables doesn´t have to be identical.

Although this model includes a first form of neighborhoods effects, it is still not taking into account the endogenous effect. In terms of informality, it is telling us the impact of the quality of the neighborhood on the probability of working in the informal sector, but it is not telling us the principal relation we are looking for, the impact of the informality of the neighborhood. The simplest way to include this in our model would be:

$$y_{ir} = \alpha + X_{ir}\eta + E_r(Z)\gamma + E_r(y)\beta + u_{ir}$$

Where $E_r(y)$ is the group average outcome (i.e., the informality rate). It is possible, however, to use other measures of the group behavior than group average. The parameter β is telling us about the presence and magnitude of the endogenous effect.

One of the problems of this kind of estimation is that the previous expression cannot be estimated if $E_r(y_{ir})$ is just the mean for the whole neighborhood, because this would mechanically make $\beta = 1$. But this depends on the definition of the peer group. A usual solution to this problem in the social interactions literature is to use as definition of the reference group of an individual *every other* individual his group. Our approach will be slightly different: for the purposes of our analysis the relevant reference group for a person looking for a job is every individual who already has a job.

Two main problems in the estimation of social interactions effects are still subject of discussion in the literature: i) the reflection problem and ii) the endogeneity bias, due to simultaneity or unobserved correlated effects.

The reflection problem assess that the group outcome $E_r(y)$ is, in equilibrium, a reflection of the group conditions $E_r(Z)$, and therefore is not possible to differentiate the endogenous effects and the exogenous and contextual effects. In the estimation, this mean lack of identification –there will be infinite possible solutions to β and γ (Manski, 1993). We can see this by obtaining the group mean outcome as the aggregation of the individual outcomes as described by the model.

Assume $X = Z$ (we will use this assumption in the remainder of the paper), that is to say, there is no variables generating individual effects and no group effects[8]. The result is a linear relation of the group mean outcomes and the group characteristics.

$$E_r(y|X) = \frac{\alpha}{1-\beta} + E_r(X)\frac{\gamma + \eta}{1-\beta}$$

Replacing the mean group outcome in the equation for individual outcome:

$$y_{ir} = \frac{\alpha}{1-\beta} + X_{ir}\eta + E_r(X)\frac{\gamma + \beta\eta}{1-\beta} + u_{ir}$$

Then, the group parameters (γ and β) are not identified. However, it is possible to establish the existence of group effects. If the term $\frac{\gamma+\beta\eta}{1-\beta}$ is statistically different from 0, this means there is either an exogenous effect or an endogenous effect, or both. Recall that this is a sufficient condition, and not a necessary one: significance suggest the existence of social interaction effects, but lack of significance does not implies their inexistence.

Brock and Durlauf (2001; 2007) show that a non-linear model can overcome the reflection problem present in the linear-in-means model. It is assumed that the individual's decision rule can be approached from a non-observable utility. The functional form and distribution of the errors assumed by the authors result in a logistic probability model, but can be generalized to a more general model.

$$\Pr(y_{ir} = 1|X) = \Lambda[\alpha + X_{ir}\eta + E_r(Z)\gamma + E_r(y)\beta + u_{ir}]$$

The reflection will be less severe in these models as long as it meets certain assumptions, which can be condensed into: first, individual and contextual characteristics must be not linearly dependent (to distinguish the effect of individual characteristics from the effect of the group characteristics); this is granted as long as the individual characteristic have within-group variation. And second, the relation between X_{ir} and $E_r(Y)$must be nonlinear; this is granted in a binary outcome model, like probit or logit. In a multinomial-choice model, like a multinomial logit, the nonlinearity makes less severe the identification problem due to reflection (Brock & Durlauf, 2002).

The second main problem in the estimation of the effects of social interactions is presence of endogeneity, due to simultaneity or unobserved correlated effects –particularly those related with residence selection–. The simultaneity problem arises from the possibility that the individual's outcome affects the group's outcome. If the individual have an informal job and he can help their neighbors get informal jobs, then y_{ir} can affect $E_r(y)$. In this way individual error terms will be correlated to $E_r(y)$, causing endogeneity.

---

[8] The existence of a variable in X_{ir} and not in Z_{ir} help to overcome the reflection problem.

The unobserved correlated effects are especially relevant in urban studies, like ours for informality, because the selection of neighborhoods by the families generates non-random sorting. The selection bias appears when the conformation of the reference group depends on non-observable variables that also affect the outcome. Then, the members of the group will tend to behave similarly, but not because of a mutual interdependence, but because similar people choose similar groups. A model with this kind of sorting would be like the following:

$$\Pr(y = 1|X) = \Lambda[\alpha + X_{ir}\eta + E_r(Z)\gamma + E_r(y)\beta + u_{ir}], \qquad where\ u_{ir} = E_r(D)\delta + v_r + \epsilon_{ir}$$

The term u_{ir} includes the correlated effects variables. D is a vector of characteristics associated to group selection and relevant in the determination of the outcome; v_r is a group disturbance; and ϵ_{ir} is the individual disturbance term. If D is observable, then there is no bias. But it is not, then there will be an omitted variable bias. Since $\delta \neq 0$, $E_{ir}(y)$ and D will be correlated, and we will mistakenly take a sorting effect as a endogenous effect.

In brief, empirical studies of social interactions face two main challenges: the reflection problem and the endogeneity bias. We will proceed to propose a solution to these problems taking advantage of the nature of our study and the possibilities with the data for Colombia.

4.2 A fixed effects approach using observation of neighborhoods across time

The first step to finding a solution to the three problems described above –reflection, simultaneity and unobserved correlated effects– is to determine the adequate variables for the individual outcome and the vicinities outcome. The easiest solution is to say: the individual outcome is informality status (0 formal, 1 informal), and the neighborhood outcome is the average, this is, the informality rate. But this ignores the following question: are we interested in the probability of *being* an informal worker, o in the probability of *getting* and informal job? Our goal is to determine whether an informal worker affects the odds of getting an informal job for his neighbors (i.e. if informality is contagious). Then, we should focus our attention on individuals who just got his job; and the relevant peer group for our analysis are the workers who already has a job. This, additionally, help us to overcome some of the estimation difficulties.

Reflection. Blume & Durlauf (2006) have shown that if the neighborhood average variable is lagged –i.e. today's y_{ir} does not depend on today's $E_r(y)$ but on yesterday's–, then the reflection is broken. In this case $E_r(y)$ is not any more a reflection of $E_r(X)$; it depends now on a combination of previous $E_r(X)$. This is what happens when we are comparing people getting a job with people who already have a job: the kind of job (formal or informal) that one will find depends on the people in the neighborhood who already have an informal job; but this means that it depends on the kind of job they found in the past. We are introducing an implicit lagged structure

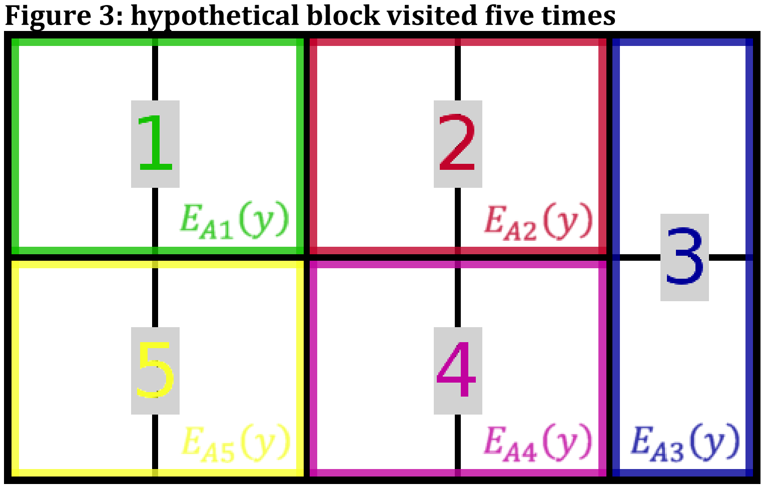
Simultaneity. The individuals we are studying –people that got his job during the last month– are not included in the reference group we are taking into account –people that got his job more than a month ago–. Then we can discard the direct effect of individuals over group averages. And any other reversal causality is unlikely to have effect, since we are comparing the informal status that a worker adopts today, with the informal status that his neighbors adopted some time before. Since today realization cannot have effect on previous realizations, simultaneity is discarded.

Selection. Non-random sorting and sharing a common institutional environment are difficulties for our estimation as long as they are not observed. A natural way to control for this kind of unobserved group effects is to use fixed effects. In a cross-section study we would use a nested model. But this is not possible in this context, because the endogenous effect variable ($E_r(y)$) and the contextual effects variables ($E_r(X)$) does not vary within the group – i.e. everyone in the same neighborhood have the same neighbors. A solution to this problem is to add the time dimension to the analysis: if one has observations at various moments, and the group outcome changes over time, then multicollinearity is broken.

We do not have georeferenced panel data in which we can observe individuals over time. As we will explain below, our dataset is composed of several cross-sections, and the survey is applied to different households each time. *But we can observe neighborhoods over time*. The GEIH survey goes to the same blocks during several quarters, but in each occasion they select different segments of houses. They poll different households and different workers, but from the same place. Then, we can control for the neighborhood, at the block level, and still have variance in the endogenous effects variable – the informality rate.

Let's see the example in figure 1. Block *A* have been divided in five segments, and each segment have been surveyed in a different quarter. All the surveyed people live close to each other, and one can consider them to be in the same neighborhood. There can be non-random sorting in the selection of the residency; however, there is no reason to think that one side of the block will be preferred by radically different people than the other. Bayer, Ross, & Topa (2008) use measures of sorting on observable attributes to show that people choose the area of the city that they prefer, but not the specific block –or the side of the block–. So the fixed effect at the block level should account for the self-selection possibility. Nevertheless, the informality rate for each period of observation will vary, and identification will be possible.

Figure 3: hypothetical block visited five times



Our nonlinear model with will become:

$$P(y_{irt} = 1|X) = \Lambda[\alpha + X_{irt}\eta + E_{rt}(X)\gamma + E_{rt}(y)\beta + \tau_t + v_r + \epsilon_{ir}]$$

$E_{rt}(X)$ are the exogenous characteristics in the neighborhood r, at time t; $E_{rt}(y)$ is the informality rate (endogenous outcome) in the neighborhood r at time t; and τ_t are time fixed effects, controlling for the city-wide common shocks; and v_r are the neighborhood (block) fixed effects.

This will overcome the problems of correlated effects and selection bias as long as the source of unobserved process in time invariant. This is a reasonable assumption to the extent that the social composition of the neighborhoods changes slowly, and our sample has only three years (2009, 2010 and 2011); moreover, about 64% of the blocks are observed during six quarter or less, thus there is not a lot of time for composition changes. The usage of fixed effects, instead of random effects, allows us to avoid the assumption of no correlation with observable characteristics. Correlated effects can be common institutional environments, even informal institutions like tolerance with others breaking the law, or common shocks, like the existence of a local source of informal employment. These things are probably related with observable characteristics of individual and neighborhoods, like education or type of occupation. Thus the orthogonality assumption is not tenable, and fixed effects are better suited.

This methodology exploits the peculiarities of the available data for Colombia. Even without a panel dataset, we can take time into account, by taking advantage of the data collection process.

5 Data

5.1 Dataset

We use repeated cross-sections from a household survey, the *Gran Encuesta Integrada de Hogares* (GEIH), for the period 2009-2011. The survey is conducted on a monthly basis and asks for demographic and socio-economic information, like gender, age, education attainment and family structure; and it asks about labor market characteristics, such as employment status, income, occupation, economic activity, job type, social security affiliation, firm size, time in the same job and unemployed duration, among other things.

The survey makes a multi-step stratified sample by geographical conglomerates (DANE, Metodología Gran Encuesta Integrada de Hogares, 2013). Therefore, the data collection has a spatial framework we can exploit. The spatial conglomerate they use to make the sampling, in each stage, is based in geographic contiguity. Using cartographic information, they divide the population in several ‘primary units of sampling’, each of this is divided in several ‘secondary units of sampling' and so on. The final unit of sampling is a 'segment', a set of about 10 houses. So, the survey inquires a worker and his neighbors.

The geographical units that organize our dataset are 'urban sectors'[9], 'urban sections'[10], and 'blocks'[11] (those are not equivalent to the units of sampling). The city is composed by 606 urban sectors, each one with 5.866 households and 19.292 residents. Each urban sector is divided in urban sections (there is about 4.5 sections per sector), with an average of 912 households and 2893 residents. Finally, the urban section is divided in blocks (about 17 blocks per section), each one with 59 households and 181 residents.

Our data has 59526 respondents in the labor force, but only 53,147 have a job at the moment of the survey. They live in 229 urban sectors, of the 606 sectors of the city. They are located in 329 different urban sections, and in 821 different blocks. The dispersion of the sample is higher in terms of urban sectors than in terms of blocks. This is a consequence of the sampling process and the fieldwork design[12].

Table 1: Size of the vicinities in the sample

	Number of geographical units		Residents per geographical unit (a)		Households per geographical unit (a)	
	(In city)	In Sample	(In city)	(In Sample)	(In city)	(In Sample)
Urban sector	606	229	19292	587	5866	170
Urban section	2718	329	2893	414	912	121
Block	45680	821	181	164	59	48

(a) Information about number of residents in the city and the number of households in the city per geographical unit is from the Colombian general census of 2005.

Source: IDECA, GEIH (DANE) and General Census (DANE)

The survey is conducted on the same places for several quarters; not to the same houses, but in the same neighborhoods. This is a consequence of the fieldwork design, which establishes a segment rotation for the application the survey. Table 5 shows that for about 94% of the workers in the sample we have several observation of the neighborhood across time. In 11% of the cases we have observation for the whole period, one each three months.

[9] The urban sector is designed to be similar to the basic unit of urban territorial organization in Colombia, the 'barrio' (DANE, 2011).
[10] The urban section is designed to consist of about 20 blocks (DANE, 2011).
[11] The cartographic definition of 'block' used in the survey is almost the generic definition, except that a few blocks with very small population are grouped together.
[12] It is easier to go to adjacent places to apply the survey. Then it's better to go to several Blocks within the same Section, instead to go to several Sections.

Table 5: Visits of the survey to the same block

Number of visits per block	Number of workers in the block	Percentage of the total
1	3,706	6.23
2	7,569	12.72
3	8,827	14.83
4	8,068	13.55
5	6,428	10.80
6	7,127	11.97
7	5,909	9.93
8	2,761	4.64
9	1,090	1.83
10	566	0.95
11	700	1.18
12	6,771	11.38

Source: Own calculations, with information from DANE

5.2 Neighborhood variables

As usual in the neighborhoods effects literature, we use group averages as our neighborhood variables. For a neighborhood r we will define two groups: the people that got his job during the month of the application of the survey, $\mathcal{I} = \{1, \dots, I\}$; and the people that got his job more than a month before the application of the survey, $\mathcal{J} = \{1, \dots, J\}$. When looking for a job during that month, an individual i can receive help from one of his employed neighbors j. Therefore, we can see through the labor informality status of individuals i how many workers were *getting* an informal job; and we can see through the characteristics of their neighbors j what were the social and economic features of the employed workers in the vicinity at that time (like the informality rate). If we find an association between these two variables that would be an evidence of neighborhood effects. The neighborhood averages are defined as follows:

$$E_r(Z) = \frac{1}{J} \sum_{j \in \mathcal{J}} Z_{ir}$$

It is worth noting that our working hypothesis is that a fraction of the social network of the individual is found in the vicinity. It does not mean that the neighborhood is the only social context relevant in terms of labor outcomes, nor even the most important.

We neither assert that the entire neighborhood is part of an individual's social network. One appealing criticism to our spatial approach for social interactions is the potential inclusion of many 'irrelevant' individuals in the reference group, because some people living close may have little or none at all interaction with each other. Nevertheless, this will not mean a bias in our estimations. This is just the way local interactions works: probably a person interacts with some of his neighbors, but not with all of them; most likely only a small percentage of the people in the surroundings have an active social tie. This is a sufficient condition to the existence of an association between the labor outcomes and the place in which people live. Of course if everyone in the neighborhood had an active social tie, then the association would be stronger. But the association do not have to be that strong to be relevant for labor market analysis.

Table 3: Descriptive statistics - neighborhood variables

	Mean	SD
Women	0.46	0.11
Age	37.95	3.73
Years of education	9.86	2.54
Unemployed duration	5.49	3.73
Participation in communal activities	0.01	0.02
Unemployment rate	0.12	0.08
Informality rate (Strong)	0.62	0.17
Informality rate (Weak)	0.21	0.11
Informality rate (DANE)	0.48	0.17
Income per household	2,361,493	2,096,175
Income per capita	837,094	683,090

Source: Own calculations, with information from DANE

Geographical distance is one of the relevant dimensions of the social networks. The creation and maintenance of social ties is costly, and a greater spatial distance will increase this cost. Then, an individual is more likely to interact with persons living close than with persons far away (Conley & Topa, 2002). But proximity is not the only feature in determining the probability of social contact. Individuals are likely to have social contact with similar people in certain socio-economic attributes that we can observe in our data. Akerlof (1997) argues that the intensity of the interaction between two persons is positively related to its 'social distance'. We will define social distance between the individual and the vicinity for variables as age, gender and years of education, in the following way:

$$D_{ir} = |Z_{ir} - E_r(Z)|$$

We also define distances in occupation and economic activity. Conley and Topa (2002) justify the relevance of the occupational distance in terms of informational content: only some contacts are useful in terms of getting information about a job opening or generating good referrals. A neighbor with a different occupation or working a different economic activity is less likely to transmit relevant information.

For an individual i, living in the neighborhood r and working in the occupation c, the occupational distance will be defined as:

$$OD_{ir} = (1 - O_{cr}) \quad for\ i \in c$$

Where O_{cr} is the percentage of employed workers in r with occupation c. The definition for economic activity is analogous.

Table 3: Descriptive statistics - social distances

	Block	
	Mean	SD
By gender	0.51	0.12
By age	12.67	7.33
By years of education	2.75	2.12
By log. Income	3.55	1.16
By economic activity	0.21	0.15
By occupation	0.26	0.16

Source: Own calculations, with information from DANE

6 Results

Table 8 shows the estimations of endogenous effect for the linear model, the linear model with fixed effects, the nonlinear model with a logistic distribution, the nonlinear model with random effects –the estimation suggested by Brock and Durlauf (2001)–, the nonlinear model with fixed effects using the unconditional estimator, and the nonlinear model with fixed effects using the conditional estimator. In all the estimations, except the one with random effects, we allow within-block correlation of the error term to correct the effect of selection of neighborhoods and unobserved variables on variances (Soetevent & Kooreman, 2007). The reported parameters are the marginal effects, the logarithm of the odds ratio and the odds ratio. The latter is included to assess the comparability of the nonlinear models, to the extent that the marginal effects cannot be computed for the conditional fixed effects estimation.

The comparison between the linear model and the nonlinear model gives an insight about the magnitude of the reflection problem, in so far as the latter is expected to have smaller biases because of this. Whilst the marginal effect for the former is slightly bigger, there is not a significant difference (the relevant comparisons are between columns 1 and 3, and between columns 2 and 5). Therefore, it is probable that the reflection problem do not generate large biases.

The inclusion of the random and fixed effects accounts for the existence of time-invariant correlated effects, especially those arousing from sorting and the unobserved component of the common local environment. We first try the random effects specification (shown in column 4), upheld by Brock and Durlauf (2001). The result is essentially the same as in the basic nonlinear model. But this is not a consequence of the inexistence of confounding factors, but a consequence of an inadequate assumption. As we have mentioned before, this approach requires no correlation between the unobserved confounding factor and the individual or neighborhood characteristics. This, in terms of self-selection in neighborhoods, implies that unobserved variables affecting the neighborhood selection is not related to education, occupation or family structure. On the other hand, in terms of common institutional environment, this implies that the neighborhood tolerance to job conditions is not related with education; or that the impact of a local job opportunity (like a local source of informal job) is not related to the workers occupation. These are not tenable assumptions.

Fixed effects model relaxes the assumption of no correlation between the covariates and the neighborhood non-observables, and it can be estimated by maximum likelihood or conditional maximum likelihood. On the one hand the maximum likelihood method allows the computation of marginal effects, but suffers from the incidental parameters problem (Neyman & Scott, 1948), causing inconsistent estimations. On the other hand the conditional maximum likelihood estimator is consistent (Chamberlain, 1980), but it does not allow the computation of marginal effects[13]. The magnitude of the bias caused by the incidental parameters problem is subject of discussion, but with a small amount of observations per group –in our case, about 5 workers per block got his job a month before the survey application– there is some evidence that the bias could be positive and large (Greene, 2004; Katz, 2001; Heckman, 1981).

Columns 5 and 6 of table 8 show the estimations by unconditional and conditional maximum likelihood respectively. As expected, the unconditional parameter is slightly larger, suggesting that the estimation could suffer from the incidental parameters problem. Then, the conditional fixed effects model is the best suited for our analysis. Nevertheless, the difference in the estimated parameters is small and not significant, so the use of the unconditional estimator should not generate big problems and the results will be about the same.

The interpretation of the odds ratio is not as straightforward as the interpretations of the marginal effects. Commonly we present probabilities as numbers between 0 and 1. Using the unconditional estimation we could say the probability of getting an informal job is 88% for workers in neighborhood in which everyone has an informal job, and 15% less for a worker living in a neighborhood in which everyone has a formal job (everything else constant). Another possibility is to present these probabilities as odds, this is, as $P/(1-P)$. Then we would say that the odds of getting and informal job are 1 in 7.58 in a fully informal neighborhood and 1 in 2.71 in a fully formal one. The odds ratio is the ratio of these odds, in this case, it is 2.79. This means that the odds of being informal are roughly three times bigger in a neighborhood completely informal than in one completely formal.

[13] The marginal effect for the nonlinear model with logistic distribution is defined as $ME = \Lambda(X\beta)\big(1-\Lambda(X\beta)\big)$, being X the vector with the independent variables and the fixed effects. But the fixed effects are not computed by the conditional or the random effects estimators, and therefore it is not possible to compute the marginal effect. It is possible to compute the marginal probabilities assuming all fixed effects to be zero, but these wouldn't be comparable with the marginal probabilities from the other estimations.

Table 8: Effect of vicinity's informality rate on the probability of getting an informal job

Models	(1) Linear	(2) Linear (FE)	(3) Nonlinear	(4) Nonlinear (RE)	(5) Nonlinear (UC-FE)	(6) Nonlinear (C-FE)
Marginal Effect	0.131***	0.123*	0.123***		0.142*	
	(0.044)	(0.064)	(0.041)		(0.077)	
Log. Odds Ratio			1.000***	1.000***	1.027*	0.907**
			(0.331)	(0.311)	(0.558)	(0.462)
Odds Ratio			2.717***	2.719***	2.794*	2.476**
			(0.900)	(0.845)	(1.558)	(1.145)
Individual Controls	✓	✓	✓	✓	✓	✓
Family Controls	✓	✓	✓	✓	✓	✓
Neighborhood Controls	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	
Block FE		✓			✓	✓
Number of obs.	4,385	4,385	4,380	4,385	3,335	3,340
Number of blocks	761	761	761	761	433	433
R2	0.203	0.174	0.204	0.189	0.294	0.229

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Notes: Fixed Effects (FE), Random Effects (RE), Unconditional estimator of Fixed Effects (UC-FE), Conditional estimator of Fixed Effects (C-FE). All standard errors, except those from the Nonlinear RE estimation, are estimated with block clusters.

It makes no sense, however, to compare neighborhoods with all workers in the formal sector versus neighborhoods with all workers in the informal sector. These are not the geographical patrons we have identified in the city of Bogotá. Table 9 presents the odds ratio for the conditional estimator (as we have pointed before, the best one) comparing a neighborhood with an informality rate 10% higher than the city average, versus a neighborhood with a rate 10% lower; the table also presents the same comparison for an inter-decile range: the 10% neighborhoods with highest informality rates versus the 10% neighborhoods with lowest rates.

Table 9: 20% and inter-decile changes

20% change odds ratio	1.199
	(0.111)
Inter-decile odds ratio	1.508
	(0.316)

Robust standard errors in parentheses (clustered in blocks).

Let's take two workers who were looking for a job, and during the last month got it; one live in a neighborhood in the top 10% in term of block's informality rate, and the other in the top 10% in terms of formality. The odds of being informal for the first one are 50% bigger than for the other. And a 20% difference in the block' informality rate would imply that the odds of getting an informal job are 20% bigger. The effect of the block's informality rate on the individual's probability of getting an informal job is small but significant. We can conclude that the existence of neighborhood effects is proven, at least at the level of blocks, since the estimation is expected to have no bias due to reflection problems nor to time-invariant correlated effects. Location affects employment; and more important, neighbors affect employment.

The magnitude of the effects needs some interpretation. Because we are not taking into account all the social network of the individual, only the fraction that is localized in the vicinity. And even so, we found a strong effect. The impact of the rest of the social contacts could be bigger.

As we have pointed out before, geographical proximity is not the only determinant of the existence and use of social ties. We will show that people with certain attributes is more likely to be affected by the informality rate in the vicinity. And we will highlight the importance of social distance: when a worker is more alike his neighbors, is more probable that they exert some influence on his labor outcomes. In other words, neighborhood effects on informality are heterogeneous.

6.1 The wage-earner and the self-employed

So far we have treated labor informality as if it were a homogeneous phenomena in terms of type of occupation. But it is not. Informality is a situation of precarious working conditions, so the role of the employer should not be discarded. Especially if we are analyzing the contagion of informality in the neighborhood: it is not the same to get an informal job from a firm avoiding regulations, than making your informal job position by yourself.

Most of the neighborhood effects literature in labor market topics has focused on salaried workers, and therefore most of the interpretations of the subjacent social interactions are associated to a labor relationship employee-employer. The employed workers can influence the labor prospects of the ones looking for job through information flow about vacancies or through referrals. But these two explanations are singular to the salaried case: either you hear about a vacancy in the firm in which you are working, and tell your neighbor; or your employer request you to refer a candidate for the job position. The interpretation of the endogenous effect should be different for the self-employed workers. There is not a firm offering a job position for self-employment. The contagion channel should be, therefore, a different one.

Young (2001) classifies endogenous effects in three categories: pure conformity, instrumental conformity and informational conformity. The first is mimic behavior because the utility of the imitation. This kind of behavior could be the case of the role model effect: a successful micro-business experience, in informal conditions, can motivate other workers in the same direction. The second, instrumental conformity, refers to the benefits that the worker receives from behaving in the same way as the other members of the reference group. This could be the case of the social support networks studied by Portes & Haller (2004): a stronger network operates with more ease. And it can be the situation of the tolerance environment for unprotected work, in the sense that a neighborhood with more informal worker is less likely to exert social control over the creation of new informal jobs. Finally, informational conformity refer to the transmission of information. It can work in a similar way to the vacancy case for salaried workers, if there are working opportunities for self-employment which other independent workers know. Or it is an observations process, in which workers realize the opportunities by looking at his neighbors. Or maybe it is a learning process, in which self-employed workers learns from others the skills needed for the job.

Aught, the social interactions effect could be a lot different for someone looking for a salaried job than for someone looking for a self-employment opportunity. To make an approximation to this dichotomy, we make our estimations with two different endogenous effects variables: salaried informal workers as percentage of total workers and independent informal workers as percentage of total workers. The results of this estimation is shown in table 10.

Table 10: Salaried and self-employed informality

Models	Marginal effect (UC-FE)	Odds ratio (C-FE)	Inter-decile Odds ratio (C-FE)
Salaried informality	-0.006	0.865	0.951
(as % of block employment)	(0.099)	(0.503)	(0.351)
Independent informality	0.243***	4.904***	1.805***
(as % of block employment)	(0.089)	(2.681)	(0.390)
Individual Controls	✓	✓	✓
Family Controls	✓	✓	✓
Neighborhood Controls	✓	✓	✓
Time FE	✓	✓	✓
Block FE	✓	✓	✓
Number of obs.	3,335	3,340	3,340
Number of blocks	433	433	433
R2	0.294	0.229	0.229

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

We found no evidence of endogenous effects for the neighborhood salaried informality. On the other hand, the influence of self-employed informality is big. The odds of getting and informal job are 80% bigger for an individual living in a neighborhood in the top 10% in terms of informal independent workers as percentage of total employed workers, when compared with an individual in a neighborhood in the bottom 10%. It seems that the neighborhood effects operate through independent labor.

6.2 Unemployment in the social interactions framework

So far, we have focused our attention on the occupied population. We have studied the probability of getting an informal job, taking formality as the only alternative. But in reality there is another possibility: to remain unemployed. A multinomial logistic model describing this situation can be stated as follows:

$$\Pr(y = j|X) = \Lambda[\alpha + X_{ir}\eta + E_r(X)\gamma + E_r(y)\beta + \epsilon_{ir}]$$

$$j = \{unemployed, formal\ employed, informal\ employed\}$$

We use unemployment as the base category. We choose the multinomial logistic model instead of a nested model (something like unemployed vs. employed, and formal and informal as subcategories of employed), because the later assumes an inexistent dichotomy between unemployed and employed. An unemployed can get a formal or an informal job, as well as an informal worker can get a formal job, or a formal worker can change his job for an informal one. There is not an a priori nesting structure to assume.

Table 11: Unemployment, formality and informality as individuals outcomes

Models	ME Unemployed	ME Formal	ME Informal	ME Unemployed	ME Formal	ME Informal
Salaried informality	-0.120**	0.001	0.119**	-0.028	0.002	0.026
(as % of block employment)	(0.051)	(0.007)	(0.048)	(0.089)	(0.013)	(0.082)
Independent informality	-0.115***	-0.009	0.123***	-0.104	-0.016	0.120*
(as % of block employment)	(0.042)	(0.006)	(0.040)	(0.067)	(0.010)	(0.062)
Individual Controls	✓	✓	✓	✓	✓	✓
Family Controls	✓	✓	✓	✓	✓	✓
Neighborhood Controls	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Block FE				✓	✓	✓
Number of obs	10,761	10,761	10,761	7,892	7,892	7,892
Number of blocks	808	808	808	430	430	430
R2	0.575	0.575	0.575	0.600	0.600	0.600

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The marginal effects of the informality rates are quite similar to what has been shown in table 8, so the consideration of the role of unemployment do not change the general results of our analysis. But note that a neighborhood with more informal workers implies higher probability of finding an informal job, but it does not implies lower probabilities of getting a formal job; this is, if you have greater ease to get an informal job, that does not mean you will necessarily stop looking for a formal job. It just decreases your probabilities of being unemployed. Informality could be a settler against unemployment, more than a substitute of formal employment.

7 Conclusions

This is, as far as we know, the first attempt to assess the neighborhood effects on labor informality using the social interactions framework. The study of neighborhood effects in developed countries has been centered in other labor outcomes, like unemployment. In developing countries this literature is starting to grow, but the availability of georeferenced data has been a major complication, especially for the most robust empirical designs that requires panel data. Our work overcomes some of these problems, with a creative design that benefits form the available Colombian data, and set up a methodology that can be used in future research on neighborhood effects in the country.

Almost 9 in 10 informal salaried workers found his job with help from relatives, friends and acquaintances. This suggest the importance of the social ties in the job matching process. And this have implications in terms of informality: if most of my social contacts have informal jobs, it raises my probabilities of getting an informal job. In other words, labor informality is contagious.

The evidence for the city of Bogotá suggest the importance of the neighborhood in the job search process, and the spatial correlations are consistent with the existence of social interactions at the neighborhood level, affecting the entry probability of informality. We found a robust relation between the vicinity's rate of informality and the individual probability of getting an informal job. Even after adequate corrections for the reflection, simultaneity and selection bias, there is evidence about the existence of this kind of peer effects.

Salaried and independent workers are different in several ways, and the presence of neighborhoods effects is one of them. The neighborhood affects the probability of getting and informal job as a wage-earner and also as an independent worker. An analysis assuming homogeneity between these different kinds of working relations will lead us to wrong interpretations of the labor market behavior and the peer effects. Further work should review the social interactions influences for independent workers – the transmission channel is subject to further research.

Finally, we would want to use these results to contradict the metaphor we portray in the introduction. Maybe informality is contagious. But maybe it is not a disease at all, and it is just an underlying social process –at least self-employed informality–: people learn from each other how work relations should be; and different social contexts generate different informal institutions. Maybe public policy against informal jobs should not focus only on incentives, but in the process leading worker to precarious job conditions. Maybe labor informality is not in the labor market, but in the streets.

8 References

Akerlof, G. A. (1997). Social Distance and Social Decisions. *Econometrica, 65*(5), 1005-1028.

Atuesta, B. (2010). Modelando informalidad voluntaria e involuntaria.

Bayer, P., Ross, S. L., & Topa, G. (2008). Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes. *Journal of Political Economy, 116*(6), 1150-1196.

Blume, L. E., & Durlauf, S. N. (2006). Identifying social interactions: A review. *Methods in social epidemiology*, 287-315.

Brock, W. A., & Durlauf, S. N. (2001). Discrete choice with social interactions. *The Review of Economic Studies, 68*(2), 235-260.

Brock, W. A., & Durlauf, S. N. (2002). A Multinomial-Choice Model of Neighborhood Effects. *American Economic Review, 92*(2), 298-303.

Brock, W. A., & Durlauf, S. N. (2007). Identification of binary choice models with social interactions. *Journal of Econometrics, 140*(1), 52-75.

Calvo-Armengol, A., & Jackson, M. O. (2004). The effects of social networks on employment and inequality. *The American Economic Review, 94*(3), 426-454.

Chamberlain, G. (1980). Analysis of Covariance with Qualitative Data. *The Review of Economic Studies, 47*(1), 225-238.

Conley, T. G., & Topa, G. (2002). Socio-economic distance and spatial patterns in unemployment. *Journal of Applied Econometrics, 17*(4), 303-327.

DANE. (2011). *Manual de recuento Gran encuesta integrada de hogares GEIH.* DANE (Departamento Nacional de Estadística).

DANE. (2013). *Metodología Gran Encuesta Integrada de Hogares.* DANE (Departamento Nacional de Estadística).

De Soto, H. (1989). The other path: The informal revolution. *Ncw York: Harper and Row*.

Durlauf, S. N. (2004). Neighborhood effects. *Handbook of regional and urban economics, 4*, 2173-2242.

Durlauf, S. N., & Tanaka, H. (2008). Understanding regression versus variance tests for social interactions. *Economic Inquiry, 46*(1), 25-28.

Feige, E. L. (1990). Defining and estimating underground and informal economies: The new institutional economics approach. *World development, 18*(7), 989-1002.

Fields, G. S. (1975). Rural-urban migration, urban unemployment and underemployment, and job-search activity in LDCs. *Journal of development economics, 2*(2), 165-187.

Glaeser, E. L., Sacerdote, B. I., & Scheinkman, J. A. (2003). The social multiplier. *Journal of the European Economic Association, 1*(2-3), 345-353.

Glaeser, E. L., Sacerdote, B., & Scheinkman, J. A. (1996). Crime and social interactions. *The Quarterly Journal of Economics, 111*(2), 507-548.

Graham, B. S. (2008). Identifying social interactions through conditional variance restrictions. *Econometrica, 76*(3), 643-660.

Granovetter, M. (1985). Economic Action and Social Structure: The Problem of Embeddedness1. *American Journal of Sociology, 91*(3), 481-510.

Greene, W. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal, 7*(1), 98-119.

Guataquí, J. C., García, A. F., & Rodríguez, M. (2011). El perfil de la informalidad laboral en Colombia. *Perfil de coyuntura económica, 16*(16), 91-115.

Harris, J., & Todaro, M. (1970). Migration unemployment and development: a two-sector analysis. *American Economic Review, 60*(1), 126-42.

Hart, K. (1970). Small-scale entrepreneurs in Ghana and development planning. *The Journal of Development Studies, 6*(4), 104-120.

Hart, K. (1973). Informal income opportunities and urban employment in Ghana. *Journal of modern african studies, 11*(1), 61-89.

Heckman, J. J. (1981). Structural Analysis of Discrete Panel Data with Econometric Applications. En C. F. Manski, & D. McFadden (Edits.). MIT Press.

Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states* (Vol. 25). Harvard university press.

ILO. (1972). *Employment, income and equality: A strategy for increasing productivity in Kenya.* Tech. rep., International Labor Office, Geneva.

Ioannides, Y. M., & Loury, L. D. (2004). Job Information Networks, Neighborhood Effects, and Inequality. *Journal of Economic Literature, 42*(4), 1056-1093.

Ioannides, Y. M., & Topa, G. (2010). Neighborhood effects: accomplishments and looking beyond them. *Journal of Regional Science, 50*(1), 343-362.

Katz, E. (2001). Bias in conditional and unconditional fixed effects logit estimation. *Political Analysis, 9*(4), 379-384.

Lee, B. A., & Campbell, K. E. (1999). Neighbor networks of black and white Americans. *Networks in the global village: Life in contemporary communities*, 119-146.

Lewis, W. A. (1954). Economic development with unlimited supplies of labour. *The manchester school, 22*(2), 139-191.

Maloney, W. F. (2004). Informality revisited. *World development, 32*(7), 1159-1178.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies, 60*(3), 531-542.

Moffitt, R. A., & others. (2001). Policy interventions, low-level equilibria, and social interactions. *Social dynamics, 4*, 45-82.

Montgomery, J. D. (1991). Social Networks and Labor-Market Outcomes: Toward an Economic Analysis. *American Economic Review, 81*(5), 1407-1418.

Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society, 16*(1), 1-32.

Otani, S. (1999). Personal community networks in contemporary Japan. *Networks in the global village: Life in contemporary communities*, 279-98.

Pellizzari, M. (2010). Do Friends and Relatives Really Help in Getting a Good Job? *Industrial and Labor Relations Review, 63*(3), 494-510.

Perry, G. (2007). *Informality: Exit and exclusion.* World Bank Publications.

Piore, M. J. (1983). Labor market segmentation: to what paradigm does it belong? *American economic review, 73*(2), 249-53.

Portes, A., & Haller, W. (2004). *La economía informal.* Cepal.

Portes, A., Castells, M., & Benton, L. A. (1989). *The informal economy: Studies in advanced and less developed countries.* Johns Hopkins University Press Baltimore.

Pratap, S., & Quintin, E. (2006). *The informal sector in developing countries: Output, assets and employment.* Research Paper, UNU-WIDER, United Nations University (UNU).

Quintin, E. (2008). Contract enforcement and the size of the informal economy. *Economic Theory, 37*(3), 395-416.

Quiñonez, D. (2011). Canales de búsqueda de empleo y duración del desempleo en Colombia. *Perfil de coyuntura económica, 16*, 133-154.

Rauch, J. E. (1991). *Productivity gains from geographic concentration of human capital: evidence from the cities.* Tech. rep., National Bureau of Economic Research.

Shang, Q., & Lee, L.-f. (2011). Two-step estimation of endogenous and exogenous group effects. *Econometric Reviews, 30*(2), 173-207.

Soetevent, A. R., & Kooreman, P. (2007). A discrete-choice model with social interactions: with an application to high school teen behavior. *Journal of Applied Econometrics, 22*(3), 599-624.

Tokman, V. E. (1992). *Beyond regulation: The informal economy in Latin America.* Lynne Rienner.

Topa, G. (2001). Social interactions, local spillovers and unemployment. *The Review of Economic Studies, 68*(2), 261-295.

Uribe, J. I., & Gómez, L. (2005). Canales de búsqueda de empleo en el mercado laboral colombiano 2003. *Economía, Getión y Desarrollo, 3*, 47-84.

Uribe, J. I., Viáfara, C. A., & Oviedo, Y. M. (2009). Efectividad de los canales de búsqueda de empleo en Colombia en el año 2003. *Lecturas de Econom{\'\i}a, 67*(67), 43-70.

Wellman, B. (1996). Are personal communities local? A Dumptarian reconsideration. *Social networks, 18*(4), 347-354.

Young, H. P. (2001). *Individual strategy and social structure: An evolutionary theory of institutions.* Princeton University Press.