Social interactions and information dynamics in self-employment in Mexico City

Marcos Valdivia López*

Introduction

Self-employment in Mexico City is an interesting case to study when considering network contagion. The figures of self-employment in Mexico City are ostensible larger than those in developed countries,[1] and self-employment activities in Mexico City are more closely related to informal economy and poverty. Several hypotheses have been advanced to explain this phenomenon. One set of traditional theories focus on what is called a segmented labor market or the presence of disguised unemployment in developing countries. In contrast, other studies rely on the empirical evidence of an important mobility between self-employment and other labor market sectors. These debates suggest that the central issue in self-employment is that entry-exit decisions are not only determined by external labor markets

---

[1] The self-employment rate in Mexico City is 20% in 2000.

(or price mechanisms) but also by internal labor markets where the social context (and institutions) can play a decisive role. Nevertheless, researchers have not explored if mechanisms of contagion through social networks operate in self-employment activity dynamics in developing countries. This paper fills in this gap by studying whether mechanisms of contagion through neighborhoods are operating in self-employment activities in Mexico City. The research uses a census tract based model that incorporates the effects of social interactions on the rate of self-employment.

Self-employment and social contagion

The standard literature of self-employment has been dominated by an external labor market approach. In the late seventies, Lucas (1978) formulated a model of self-employment in which individuals with higher *entrepreneurial productivity* (read as *ability*) are more likely to start their own firms. Since then, a literature that stresses that the conditions of entry are determined by liquidity constraints and assets has become the standard to understand self-employment in developed countries. Recently, some empirical studies have applied the standard theory of entrepreneurship to developing countries. For example, Maloney (2004) claims that self-employment in developing countries displays more similarities with the "voluntary" entrepreneurial small firms of advanced countries than with the profile of disadvantage that the traditional literature of informality (see below) in developing countries portrays about self-employment.[2] Moreover, this empirical literature emphasizes that self-employment is "voluntary" and reflects best decisions given the strong constraints that individuals face in developing countries (*Ibid.*).

In contrast to the standard literature, Eatwell (1997), along Joan Robinson's lines, called self-employment activities in developing countries as "disguised unemployment"; he considers that such activities reflect low levels

---

[2] This line of thought is not so far from that delineated in De Soto (1989).

of effective demand in a context of absence of long-term unemployment benefits that makes the low unemployment rate experienced in such countries illusory.[3] The phenomenon of self-employment in such countries is also linked to the dual hypothesis of the labor market that argues that there is a secondary sector with low wages and social security unprotected that absorbs the poor and disadvantage (Piore, 1979).[4] Nevertheless, empirical studies using panel data in Mexico and other countries in Latin America suggest that there is an important flow between formal and informal activities. People who decide to enter into self-employment activities are not only leaving unemployment conditions; most are leaving the formal sector or small-medium firms (Maloney, 2004; Calderón-Madrid, 2000).

It can be also argued that self-employment in developing countries acts as a cyclical buffer over business cycles (Galli and Kucera, 2003).[5] Data in Mexico runs along this hypothesis: the self-employment rate increased significantly during and after the 1995 crisis while formal employment suffers a decline (Calderón-Madrid, 2000).

Exploring the determinants of entry to self-employment activities must also include other categories of determination that go beyond the ones traditionally used by an *external market* perspective. Recent efforts to bring the standard entrepreneurial approach to developing countries have not taken into account the possible effects of social interactions (Maloney, 2004; Woodruff, 1999) even when sociological and anthropological literature about informality in developing countries consider social networks (and social capital) important components. From our perspective, the hypothesis that social interactions affect self-employment is even more appealing in contexts where informality and poverty are extreme.

---

[3] The official unemployment urban rate in Mexico is between 2 and 3 per cent.

[4] Another common reference with dualistic or segmented labor markets is the so called Harris-Todaro model (1970) for developing countries that is built up through considering wage differentials between sectors.

[5] The main point is that movements of informal employment are countercyclical: it increases in downturns absorbing workers form formal sector and decreases in upturns.

For example, panel data from Mexico (1998-1999) indicates that the mobility of persons between small firms is significant (Salas, 2003); in particular, the probability of a person leaving a small firm (2-5 workers) to experience self-employment is very high (0.43), while the probability to stay in self-employment activities is higher (80%) as observed when people are working in the largest firms (of more than 101 workers). This situation can be indicative of the impact of social interactions in producing self-reinforcing effects in each sector. It is reasonable to advance the hypothesis that people working in small units can be subject to social influence (and also skill formation) by their bosses such that it increases their odds to become self-employed, while individuals working in larger firms can make use of their social networks obtained on the job such that should they be fired or desire to look for another job, they would most likely transfer to another similarly sized firm.

Networks and internal labor markets

A literature in mainstream economics has emerged in recent years showing that social networks are important elements to determine employment and wage inequality. Examples are the studies of Topa (2001), who analyzes network effects on employment through physical distances, Arrow and Borzekowski (2004), who study the role of network connections in wage inequality, and Calvó-Armengol and Jackson (2004), who emphasize how network structures affect the dynamics of job acquisition and inequality.

However, this new social interaction literature in economics has not yet motivated further studies that analyze the phenomenon of self-employment. In contrast, sociological studies have provided better avenues in thinking about the relevance of social networks for entry in self-employment activities (Granovetter, 1995). Along these lines, empirical studies of the economic sociology of immigration have suggested that human and financial capital, as vindicated by the standard economic approach, are not enough to explain entrepreneurship in immigrants where the group of economics actors are culturally heterogeneous (Light and Rosenstein, 1995). Moreover, criticisms

can also be found in recent empirical studies in economics that show that the relationship between wealth and entry into entrepreneurship is weak, which differs from the standard economic belief (Hurst and Lusardi, 2004).

In an empirical study, Giannetti and Simonov (2004) investigate the strong correlation between individual and aggregate self-employment choices in Sweden and conclude that social norms are important determinants of the entrepreneurial activity. This is an important proposition because it brings to the discussion the hypothesis that self-employment might also be subject to contagion. Social norms which are disseminated through social networks are providing information or market tips to economic agents. In a broader sense, social norms are also part of the *social capital* category used by sociologists (Portes and Landolt, 2000). Given all the possible benefits that potential entrepreneurs can obtain from social networks, we can suggest that self-employment decisions are subject to strong peer group effects because such effects provide increasing returns to economic agents (read it as "utility" or general benefits). More importantly, if these interaction mechanisms exist, then it is possible to link a narrative of skill (or human capital) formation that is taking place locally and that can contribute to inequality in growth (Semmler, 2003).

Social influence can be an important variable to look at when the determinants of entry/exit to self-employment in developing countries are studied. Self-employment activities in developing countries can be strongly influenced by formal and informal institutional channels (De Soto, 1989; Galli and Kucera, 2003). The networks (formal and informal) that can influence the growth of self-employment activities may well represent local forces (peer group in the neighborhood, local street leaders, etc.) and also policy programs that provide a global perspective about the relevance of such activities.[6] To my knowledge, no study has explored the idea that self-employment in developing countries can be subject to contagion dynamics.

---

[6] For example, President Fox of Mexico (2000-2006) promoted self-employment activities ("changarros") among the poor and lower middle-class.

This paper contributes to filling in this gap by exploring a model of social interactions that explain the determinants of entry to self-employment at census tract level.

Spatial patterns of self-employment in Mexico City

It is important to have some empirical facts in order to advance a social contagion hypothesis. The existence of spatial agglomerations of self-employment activities are the stylized facts to justify that self-employment activities can be subject to network contagion. The purpose of the next section is to provide such spatial facts about self-employment through considering the Mexico City case. The next section uses the measure of spatial autocorrelation of a set of spatial features (*i.e.* census tract) to address empirically this issue. Spatial autocorrelation is a measure of the degree of dependency among spatial units and their associated data, and it is commonly related to the first law of geography that establishes "that everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). The literature of spatial statistics has developed some spatial autocorrelation statistics, and this research relies specifically on one of them, the Moran´s Index and in its local version.[7]

Spatial autocorrelation of the rate of self-employment in Mexico City

Self-employment activities increased during the nineties in Mexico City.[8] The rate of self-employment in Mexico City increased from 15.7 percent in 1990 to 19.56 percent in the year 2000.[9]

---

[7] See appendix for details.

[8] In this research only the part of the city that corresponds to the capital of the country is considered, that is, the Distrito Federal (DF).

[9] The self-employment rate is defined as the percentage of self-employed persons (working alone) over the population economically active (labor force). The population economically active is defined as the population over 12 years old that worked or sought job during the week of the interview.

FIGURE 1
Moran's Index of the self-employment rate in Mexico City
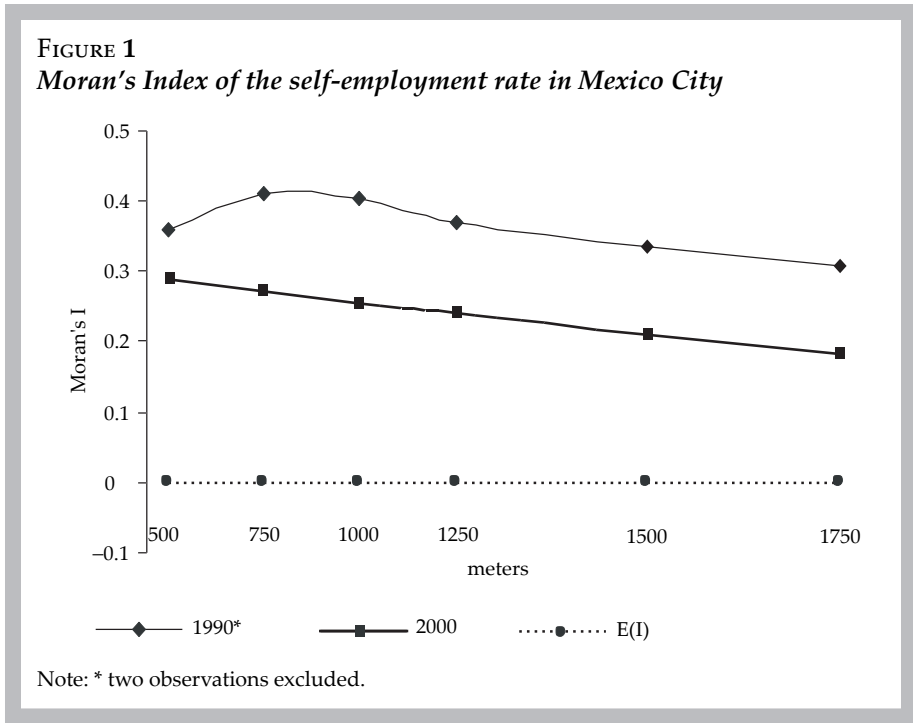

Note: * two observations excluded.

Figure 1, shows the correlogram of the Moran's Index of the rate of self-employment in Mexico City in 1990 and 2000 respectively at census tract level using different band distances. Significant and positive spatial autocorrelation is present in both years in any of the band distances.[10] The global spatial autocorrelation is less strong in 2000 than in 1990, this situation contrasts to the fact that the self-employment rate increased during the period analyzed. A reason for that might be that the area became more homogenous in the sense that local spatial disparities in self-employment were less intense in 2000. However, this cannot be inferred from Moran's Index because it is a global measurement that does not reflect local spots of spatial autocorrelation.

---

[10] All indexes are pseudo-significant: $p < 0.01$, inference based on 999 permutations.

In order to have a better picture about local changes, the Local Indicators of Spatial Autocorrelation (LISA) were calculated at census tract level (see appendix for technical details). Figures 2, are the LISA maps of Mexico City for 1990 and 2000 respectively. The maps just highlight the clusters that are (pseudo) significant ($p < 0.05$) as indicated by random permutations. In the calculations, I use a band distance of 1 500 meters to construct the relevant physical distance network structure of the census tracts.

From figure 2, it can be appreciated that Mexico City depicts, broadly speaking, a spatial polarization in the rate of self-employment that evolved during the nineties. In 1990, there was a clear city division between west and east that defined the type of cluster formation: High-High census tracts were localized in the west side of the city (*i.e.*, a census tract with an above average self-employment rate surrounded by census tract with above average self-employment), while Low-Low census tract were situated in the west side of the city in 1990 (*i.e.*, a census tract below average surrounded by census tracts below average). In 2000, the city does not depict a strong west *versus* east division as in 1990. Exploration of the spatial data detects the following local changes: *a*) a strong High-High cluster in the south west in 1990 located in the Iztapalapa area, which is populated by blue-collar families, reduced substantially its cluster size in 2000; *b*) an important Low-Low cluster in the south west in 1990 that is localized in a middle class region practically vanishes from the picture in 2000. These changes by themselves can account for much of the reduction in the global spatial autocorrelation. That is, even though the rate of self-employment increased almost five points in the whole region, it seems that the spatial polarization is reduced at the same time. Nevertheless, strong spatial clusters remain in both years. A High-High cluster localized in downtown and a Low-Low cluster in the Northwest remain as important areas that exhibit significant spatial local autocorrelation. Interestingly, downtown High-High cluster spread toward Southwest in 2000 reaching middle class zones that were characterized by having Low-Low spots in 1990. This generated the emergence of a Low-High new region of spatial autocorrelation (*i.e.*, a census tract with below average self-employment rate surrounded by census tract with above average

FIGURE 2

Mexico City LISA maps of self-employment rate using a band distance of 1.5 km. 1990 (left figure) and 2000 (right figure)
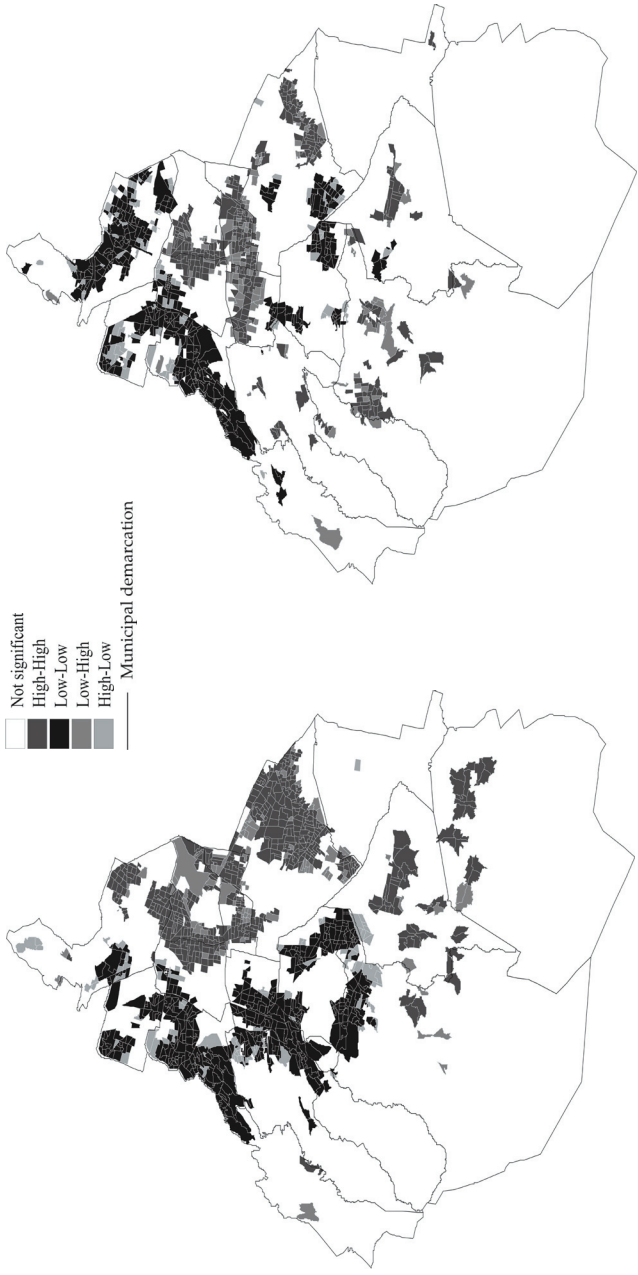
Notes: the intensity of the color in each census tract is associated to a pseudo significant LISA ($p < 0.05$). No color: no significant spatial autocorrelation; dark gray: Low-Low spot of spatial autocorrelation; medium gray: High-High spot of spatial autocorrelation; light gray: Low-High spot of spatial autocorrelation; lightest gray: High-Low spot of spatial autocorrelation.

Legend:
- Not significant
- High-High
- Low-Low
- Low-High
- High-Low
- Municipal demarcation

self-employment). This situation opens the possibility to think that a possible process of diffusion might have taken place in that region.[11]

As indicated above, in the center of the city (downtown) there is an important cluster of spatial autocorrelation where rates of self-employment are quite above the average of the city (High-High census tracts). This area (historic downtown) is characterized by the presence of a strong informal economy (street peddlers) and has a rich tradition of small entrepreneurial activities in the service sector. Next, I will focus on this area.

Downtown area contains 299 census tracts and the average population per census tract in 2000 is 3428 habitants, while the average labor force per census tract is 1 504 persons. Table 1, displays the statistics of self-employment of the region and these are compared with those observed in the whole city. Note that the rate of self-employment is quite similar to that registered in the whole city either in 1990 or 2000; the Moran's Index in 1990 is also similar.[12] But note that by 2000, the global indicator of spatial autocorrelation is larger in downtown than in the whole city; in other words, contrary to what happens in the whole city, the global spatial autocorrelation observed in downtown remains steady.

Figure 3 shows the LISA maps in 1990 and 2000 for the region studied using 1.4 km of distance. The main element in both figures is a strong division between a zone with High-High census tracts and a zone with Low-Low census tracts. Some of the highest growths in the self-employment rate are in the High-High cluster. This suggests, under a social contagion hypothesis, that the cluster is experiencing a positive feedback of increasing self-employment due to local interaction. Unfortunately, diffusion toward other areas is not perceived in the figure because the empirical analysis is restricted to downtown, but the High-High cluster extends Southwest (for

---

[11] It is part of the popular wisdom that the crisis of 1995 in Mexico (an in general the poor economic performance in Mexico during the 1990s) strongly impacted the middle-class. At least the aggregate data of this research suggests that middle-class zones started to be more engaged in self-employed activities. Moreover, the results run along the buffer hypothesis that indicates that self-employment is countercyclical.

[12] In 1.5 km the Moran's Index diminishes in downtown because the area analyzed is much smaller.

details, see figure 2). It is true that the polarization looks stronger in 1990 than in 2000, but the landscape in 2000 remains divided, even when the self-employment rate grows globally in the region.

TABLE 1
Global spatial autocorrelation of self-employment in Mexico City and downtown

| | Self-employment rate | | Moran's Index | | | | | |
| | | | 0.75km | | 1 km | | 1.5 km | |
	Mexico City	Downtown	Mexico City	Downtown	Mexico City	Downtown	Mexico City	Downtown
1990	15.65	17.28	0.4378	0.417	0.3626	0.3727	0.3685	0.2713
2000	19.56	20.30	0.2786	0.4284	0.2525	0.3816	0.2046	0.2785

Notes: Downtown E(I) = –0.0034; all Moran's Index with pseudo *p*-values < 0.05.

In the following section, I will introduce two models of social interaction that can explain the formation of spatial clusters in self-employment activities. Likewise, the models will be implemented empirically in the final sections of this essay through spatial econometrics and the framework of a census-tract model that combines Geographic Information Systems (GIS) and simulation.

A SIMPLE MODEL OF SOCIAL INTERACTIONS
AT CENSUS TRACT LEVEL

Two different model specifications to study social interactions in self-employment decisions are studied: model A that considers simple "imitation" and model B that introduces a threshold rule. These two models are built up from the next relationship: self-employment in a census tract i (S_i) is a

Downtown empirical LISA maps (1.4 km): 1990 (left fig) and 2000 (right fig)

Notes: the intensity of the color in each census tract is associated to a pseudo significant LISA ($p < 0.05$). No color: no significant spatial autocorrelation; dark gray: Low-Low spot of spatial autocorrelation; medium gray: High-High spot of spatial autocorrelation; light gray: Low-High spot of spatial autocorrelation; lightest gray: High-Low spot of spatial autocorrelation.

function of the following variables: *a*) local information of the rate of self-employment of the neighbors (S_j) that are located inside a radius of distance (R) from census tract i; and *b*) a vector of own characteristics of census tract i (X_i):

$$S_i = f(S_{j \neq i: \| j-i \| \leq R}, X_i) \qquad [1]$$

If we follow the idea of a *spatial reaction function* (Brueckner, 2003), we have an interpretation of [1] as a solution of a maximization problem at census tract level. Suppose that each census tract has the objective function:

$$U(S_i, S_{j \neq i: \| j-i \| \leq R}, X_i) \qquad [2]$$

Therefore, each census tract i chooses the level of self-employment that maximizes equation [2] so that, $S_i \equiv \partial U / \partial S_i = 0$, which solution is equation [1]. Each census tract i chooses the best response given both its own characteristics (that define its preferences) and the choices of the other census tracts. It can be assumed that each census tract is a representative agent. I will work with a linear specification of relationship [1]. I consider that the rate of self-employment in the census tract is a continuous variable in the range [0, 1] and that each census tract i is a representative economic agent. Social interactions are implemented by using the average rate of self-employment that it is observed in the census tract neighbors that are within a radius of distance. Model A considers "imitation" through conforming to the average rate of self-employment in the neighborhood.

$$S_i^{t+1*} = k + \Gamma X_i^t + \lambda \frac{1}{n_i} \sum S_{j \neq i: \| j-i \| \leq R}^t \qquad [A]$$

if $0 \leq S_i^{t+1*} \leq 100$, $S_i^{t+1} = S_i^{t+1*}$
if $S_i^{t+1*} < 0$, $S_i^{t+1} = 0$
if $S_i^{t+1*} > 100$, $S_i^{t+1} = 100$

Where S is the rate of self-employment in census tract i, k is a constant, Γ is a vector of parameters associated to own characteristic variables of census tract i (education, sex, age, % of computers, etc.) and λ is the parameter associated to the rate of self-employment of the census tracts j's that are the neighbors of census tract i. Neighbors n_i are defined as those census tract j that are located inside a radius of distance R from i, the distance is taken from the centroid of the polygon representing the census tract. Note that when $k = 0$ and $\Gamma = 0$, the level of self-employment only depends on social interactions; when $\lambda = 1$, we have a case of strict "imitation" (replicating the average rate of self-employment of the neighborhood). Equation [A] is complemented by an error term per census tract. This implies that each time that a census tract reviews its rate of self-employment, a small ε exists such that a census tract i also considers other reasons than those that are explicit in the equation (the error accounts for unobservable).[13] In contrast, in model B social interaction effects on self-employment are determined by threshold dynamics.[14] Model B has two different *laws of motion*:

$$S_i^{t+1*} = \begin{cases} k + \Gamma X_i^t + \lambda \dfrac{1}{n_i}\sum S_{j\neq i:\|j-i\|\leq R}^t & \text{if} \ \ \dfrac{1}{n_i}\sum S_{j\neq i:\|j-i\|\leq R}^t \geq \dfrac{1}{N-1}\sum S_1^t \\ k + \Gamma X_i^t & \text{otherwise} \end{cases} \quad [B]$$

where N is the total number of census tracts.

---

[13] In the econometric literature of social interaction, equations that are similar to [A] can be found in Topa (2001) (implemented also at census tract level) and in Glaeser, Sacerdote and Scheinkman (2003), who treat the relationship at the individual level in a cross section setting and where the social interaction refers to an aggregate of a reference group instead of an aggregate given by physical distance like in this case.

[14] Thresholds models, where collective behavior becomes influential in individual decisions only if a certain value (or level of *tolerance*) is surpassed, have been proposed to analyze a wide variety of social phenomena such as social revolts, residential segregation, informational cascades, etc. (Schelling, 1978; Granovetter, 1978). This type of modeling in the context of self-employment could take into account the costs associated to the perception that many (or few people) are enrolled in self-employed activities. Likewise, another important justification of proposing a threshold model in self-employment (that is not restrictive to developing countries) is related to the thresholds models of skill formation that have been proposed to explain locally increasing returns in growth theory (Semmler, 2003).

if $\quad 0 \leq S_i^{t+1*} \leq 100,$ $\qquad S_i^{t+1} = S_i^{t+1*}$

if $\quad S_i^{t+1*} < 0,$ $\qquad\qquad S_i^{t+1} = 0$

if $\quad S_i^{t+1*} > 100,$ $\qquad\quad S_i^{t+1} = 100$

Model B assumes that the underlying agents located in a census tract have a common perception that indicates that if the local rate of self-employment (neighborhood) is greater or equal than the global rate of self-employment (in the region), that would be enough to drastically increase the benefits for being engaged in self-employed activities. I assume that at each period, the global rate of self-employment is known (let us say that economic reports or local institutions make this information available). Each census tract i compares the global rate of self-employment with the rate of self-employment of its census tract neighbors and, if the rate of self-employment of the neighborhood is greater or equal than the global indicator, the census tract takes into account the rate of self-employment of its neighborhood; otherwise, the census tract updates its rate of self-employment without considering the rate of self-employment of the network. Equations [A] and [B] allow the introduction of two types of heterogeneity. First, both equations take into account census tract characteristics (such as sex, education, age, etc.). Secondly, *specific spatial location* of the census tract (the centroid of the polygon) and the radius of interaction among census tracts are the basic components of the physical network, which in the real world tends to be heterogeneous between census tracts. Geographic information systems provides data (Instituto Nacional de Estadística, Geografía e Informática, INEGI), to study empirically equations [A] and [B]. In particular, if equations [A] and [B] are considered simultaneously, a spatial econometric implementation exists using a Simultaneous Spatial Autoregressive (SAR) parametric framework. In particular, a *spatial lag model* coincides with model A in a cross section setting (Anselin, 2001, 2002):

$$S = \lambda WS + X\Gamma + u = (1 - \lambda W)^{-1} X\Gamma + (1 - \lambda W)^{-1} u \qquad [3]$$

W is a weighting matrix (n x n) that formalizes the network structure and u is a vector of random errors, S is an n by 1 vector of self-employment rate

observations, X is an n by k matrix of observations on the exogenous variables (*i.e.* census tract characteristics). Each element of W is row standardized such that $\sum_j w_{ij} = 1$, in that way premultiplying the vector of observations of neighbors (self-employment rate) by W corresponds to an averaging of the neighborhood values; in other words, equation [3] is a specification of "imitation" like in model A (but without subscript t).[15] Model B can have also an econometric implementation (at cross section level) with a spatial lag model (equation [3]). This can be done with restricting neighbors not only to distance but also when the average level of self-employment of the neighborhood is equal or greater than the aggregate level of the region. The neighborhood structure can be reproduced in the W matrix of equation [3]: the only difference (with respect to the model of imitation) is that census tract j takes a value of zero in W when the average rate of employment of the neighborhood is below the value of the aggregate rate of self-employment in the whole region considered.[16]

Due to econometric reasons that are considered in the next section, it is important to mention at this point, that the SAR framework has also another model in which the spatial dependence is incorporated in the error instead of the spatial lag variable as in equation [3]. If this were the case, the error in a linear specification to explain self employment rate could take the form of the following autoregressive model error:

$$S = X\Gamma + u \text{ and } u = \varphi W u + \varepsilon = (I - \varphi W)^{-1} u \qquad [4]$$

where ε is a vector of i.i.d errors with variance σ^2 and, W, S have the same characteristics than in the equation [3].

The differences on both models are apparent and, it is clear that the hypothesis of social interaction (imitation) only can be recovered in the spatial

---

[15] Note that in equation [3], $\partial S_i / \partial S_j = \lambda W$, which corresponds to the second partial derivative of equation [1].

[16] There is a cost in doing this and it is that the W matrix generates necessarily 'islands' (zero rows); this can complicate inferential procedures (Anselin, 2002).

implementation of equation [3] where the spatial effect is incorporated in the spatial lag variable.

Spatial econometric implementation of the models (downtown case)

In the last section, it was mentioned that the spatial lag model (equation [3]) can be considered as an econometric implementation of model A or B using cross section data. Consequently, equilibrium must be assumed when cross section data is used. In implementing the model, I have paid special attention to the exogenous variables: that is, the own characteristics of the census tract (vector X of observables in model A or B). Many variables associated to theories of self- employment mentioned before are considered. For example, I consider the percentage of the population per census tract that occupied their own house as proxy of assets.[17] In the same way, the percentage of the population between 20-24 years old (young people) and the percentage of the population between 60-64 years old (old people) are used as proxies of the risk aversion hypothesis: less "risk averse" individuals are likely to become entrepreneurs and more "risk averse" are likely to become workers (Parker, 2004). To account for the quintessential determinants of the external labor market, I consider the average years of education for people over 15 years old per census tract and also the percentage of households that have computers as a proxy of human capital. Of course, these last variables bring supply and demand effects of the labor market into the setting. Along with the literature that evaluates self-employment as disguised unemployment, I use a rate of economic dependency that indicates the number of children (between 0 to 14 years old) and older people (above 65 years old) for each one hundred economically active persons (people between 15 to 64 years old). Finally, I consider "the percentage of population who were living in the city 5 years ago" as a variable that indirectly can control for some of the

---

[17] The central issue in the standard theory of entrepreneurship is that individuals with greater assets are more likely to become self-employers (Evans and Jonavovic, 1989).

effects of sorting.[18] It is important for the reader to remember that the data used for the self-employment variable in this research corresponds only to self-employed people working alone.[19] All the mentioned variables are complemented by the social interaction variable that is the spatial lag variable in equation [3]: the rate of self-employment in each of the census tracts that are considered as neighbors (a distance criterion defines neighbors). The descriptive statistics of the mentioned variables in 2000 are displayed in table 2. As it is indicated above, all the variables except years of education and economic dependency represent percentages with respect to the population in the census tract.

Table 2 presents the mean and deviation of 299 census tract observations for downtown. In the third and fourth column of table 2, a join skewness/ kurtosis test for normality (an alternative to the Jarque-Bera test) is also presented (see D'Agostino *et al.*, 1990). This test is known to follow a χ^2 distribution with two degrees of freedom, and under the null hypothesis of normality, the expected value of the statistic is two. The results indicate clearly that all variables are far from being normally distributed (non-normality persist even when performing nonlinear transformations). The implications of this will be mentioned ahead.

If $\lambda = 0$ in the spatial lag model (see equation [3]), the econometric implementation of the models is reduced to an Ordinary Least Squares (OLS) estimation.

However, a previous section shows that the Moran's Index of the self-employment rate in the area studied presents significant spatial autocorrelation; OLS estimation can be affected by this situation.[20] Figure 4, shows the Moran's Index correlogram of the self-employment rate and the Moran's Index of the OLS's residuals in different band distance.

---

[18] Family income was excluded because it is not a census tract characteristic that can be considered exogenous like the other variables in the specification.

[19] With this regard, Woodruff (1999), in a working paper, has suggested that the determinants of self-employment in Mexico differ between a self-employer working alone and those with workers (or employers). The latter are the only ones that tend to behave along the lines of the theories of self-employment proposed in developed countries.

[20] OLS produce biased and inconsistent estimations.

TABLE 2
Descriptive statistics in downtown 2000
(299 observations)

	Mean	S.D.	Skewness/ Kurtosis test for normality (joint test)	
			Chi2	*p-value*
Self-employment rate (1990)	16.94	5.85	167.87	0.000
Self-employment rate	19.72	5.62	33.06	0.000
Years of education	10.17	1.46	14.75	0.001
Percentage of the population between the ages of 15 and 64	67.15	2.74	309.64	0.000
Percentage of the population who are men between the ages of 15 and 64	46.92	3.91	469.46	0.000
Percentage of the population between the ages of 20 and 24	13.65	3.62	542.05	0.000
Percentage of the population between the ages of 60 and 64	4.87	1.50	21.14	0.000
Percentage of the population who lived in the city in 1995	79.45	8.61	411.29	0.000
Percentage of the population living in their own house	50.29	17.09	47.64	0.000
Percentage of the households with computer	22.16	11.75	34.72	0.000
Number of persons economically dependent for each one hundred persons	46.21	5.81	171.42	0.000

Source: INEGI, Sistema para la Consulta de Información Censal (SCINCE), 1990 and 2000.

Figure 4
Moran's Index correlogram of the self-employment rate and ols residuals
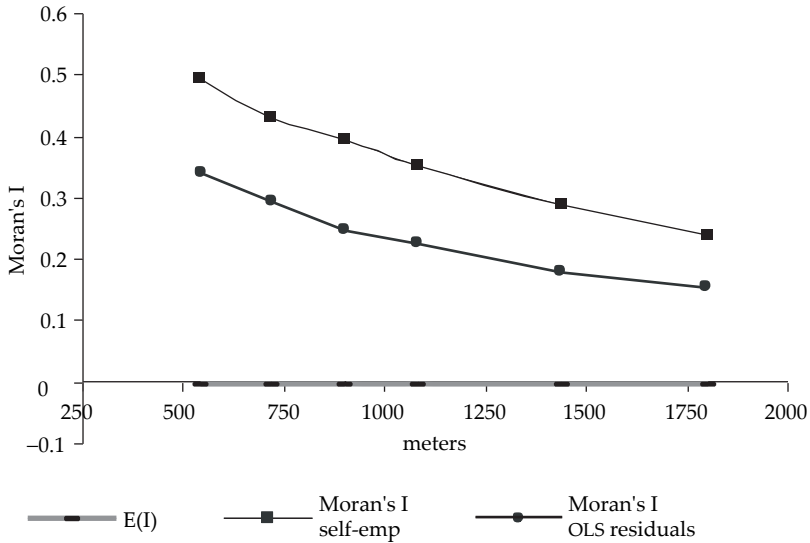
As expected, spatial autocorrelation is statistically significant also in the residuals of the ols regression (the regression without the lag variable). Note that the spatial autocorrelation is a monotonically decreasing function of physical distance. This situation alone could justify the use of a spatial lag model (equation [3]), however, the Moran's Index is also robust against misspecification problems (*i.e.* non-normality or hetereskodestacity). In fact, this is observed in the first column of table 4 where the diagnostic tests of the ols estimation of the model indicate non-normality in the errors and heteroskedasticity. Non-normality of the errors is not unexpected given that the covariates display strong non-normality (see table 2). It is important to mention that the violation of the normality hypothesis does not disqualify the model's specification when there are good theoretical reasons to think

that there is a spatial model that is adequate (Arbia, 2006).[21] Nevertheless, the problem of whether the spatial lag model is the best specification to account for the spatial autocorrelation remains, because, as it is known in the spatial econometric literature, a spatial error model (see equation [4]) can also account for the spatial autocorrelation as well. This is an important element to consider because the error can include omitted variables that are also spatially dependent.[22]

Therefore, several problems of identification can arise when a spatial (or social interaction) term is considered, and serious econometric work must be done to address these and other related problems. To respond to this issue, it is important to remark that in the last section, I showed that the spatial lag model (equation [3]) is an implementation of model A (or B) and, for that reason, it is not necessary to evaluate which econometric spatial model is more appropriate to account for spatial autocorrelation. Our motivations are theory driven not data driven. In spite of that and for information, I present the Lagrange Multiplier tests that allow for the distinction between spatial lag (LM_{lag}) and spatial error (LM_{error}) model alternatives (see Anselin, 2001).[23] Table 3, shows the main results for the tests.

Two different weight matrixes (W) of equation [3] are considered to calculate the tests: the first takes into account just local interaction (*i.e.*, model A) and the second considers the threshold model (*i.e.*, model B). As expected, all the standard Lagrange Multiplier tests accept the alternative.

---

[21] Likewise, presence of no/ normality and heteroskedasticity are frequent in this kind of settings because both are related problems (Arbia, 2006)

[22] When using a spatial lag model [3], many precautions must be taken because spatial autocorrelation can be due to either the neighbors' influence (*i.e.*, social interaction) or the errors (omitted variables) (see equation [4]). Moreover, the spatial autocorrelation can be due to both effects (see Case, 1991, for an implementation that nests both effects). Several methodological strategies can be cited to deal with this potential problem of specification (Case, 1991; Florax, Folmer and Rey, 2003; Anselin, 2002).

[23] These statistics have both an asymptotic χ^2 distribution with one degree of freedom. In strict sense, these tests would not be robust because they depend on errors that are normal. Nevertheless, we report the results because they provide useful information for this research; moreover, it is not uncommon to find in empirical essays the use of these tests even when the data shows non-normality (Mobley *et al.*, 2006).

TABLE **3**
Lagrange multiplier tests

Distance (meters)	Weight Matrix (model A)				Weight Matrix (model B)			
	LM_{lag}	LM_{error}	$LM_{lag\text{-}rob}$	$LM_{error\text{-}rob}$	LM_{lag}	LM_{error}	$LM_{lag\text{-}rob}$	$LM_{error\text{-}rob}$
540	53.86	51.16	8.57	5.87	57.8	35.18	38.57	35.18
p-value	0.000	0.000	0.003	0.015	0.000	0.000	0.000	0.000
720	88	74.23	15.83	1.87	53.97	31.79	38.80	16.62
p-value	0.000	0.000	0.000	0.172	0.000	0.000	0.000	0.000
900	97.15	87.01	17.67	7.52	40.96	40.47	29.07	28.58
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1080	101.78	102.97	13.05	14.24	45.36	55.94	32.78	43.36
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1440	102.55	119.11	11.77	28.33	n.d	n.d	n.d	n.d
p-value	0.000	0.000	0.001	0.000	n.d	n.d	n.d	n.d

Now, a simple rule sometimes used to distinguish between the alternatives when they are highly significant is to see which robust LM test has the higher statistic value (and which has the higher p-value); that is, if $LM_{lag-rob} > LM_{error-rob}$ then consider a spatial lag model otherwise go with the spatial error (Florax *et al.*, 2003). In general and for the purposes of this exercise, the spatial lag seems to be well justified as an alternative.[24] In general, the spatial lag model seems to be a better spatial option when the band distance is lower than one kilometer.

Table 4 summarizes the main results of the regressions considering only model A. The first column of the table displays OLS results of the specification (not considering the social interaction term) and in the second column is reported estimation with robust errors; likewise in the third and fourth columns, a reduced OLS version of the model is estimated. The next columns of table 4 report the estimation of the spatial lag model by maximum likelihood considering different band distances. The spatial lag estimations also report robust error estimation but it is only showed for the reduced model when the band distance is 540 meters. The spatial lag estimations are based on the algorithms outlined in Smirnov and Anselin (2001) implemented in Geoda 0.9.5-i(beta) and the robust versions were calculated in Data Analysis and Statistical Software (STATA) using the procedures of Maurizio Pisati (2001).[25]

Likewise, table 5 summarizes the main results of the regressions considering model B without reporting the robust and the reduced version of the models.[26]

First of all, table 4 and 5 indicate that the spatial lag model in both models fits better the data than OLS as it can be corroborated through the log likelihoods; but most importantly, the spatial interaction term is highly

---

[24] The tests are only informative because it is difficult to distinguish alternatives if we rely on them.

[25] Because two different programs are used in the estimation of the spatial lag model, the estimation of coefficients are not identical between them. However these discrepancies are not significant for the purposes of this exercise.

[26] I do not report these two last estimations because once they are considered in the model, the results behave similarly as in model A. In any case, the results can be obtained upon request.

Table **4**

Spatial econometric results of model A (local interaction): downtown 2000

Indepedent variable	OLS				Distance Band (meters) 540			
	Reduced model				*Reduced model*			
		robust		robust		robust		robust
Self-employment (Rate in neighbors)					0.461	0.350	0.497	0.361
s.e.					(0.043)	(0.052)	(0.044)	(0.050)
z test					10.721	6.752	11.291	7.170
Intercept	−5.238		23.854		−16.830	−13.376	9.914	21.147
s.e.	(12.175)	(15.755)	(5.375)	(7.919)	(10.674)	(13.830)	(4.913)	(6.955)
t or *z* test	−0.430	−0.330	4.438	3.012	−1.577	−0.967	2.018	3.041
Education years	−2.212		−1.097		−1.579	−1.689	−0.551	−1.682
s.e.	(0.432)	(0.572)	(0.274)	(0.363)	(0.381)	(0.488)	(0.248)	(0.505)
t or *z* test	−5.122	−3.870	−3.998	−3.024	−4.146	−3.463	−2.223	−3.327
Male 15-64	0.412				0.375	0.359		
s.e.	(0.127)	(0.277)			(0.112)	(0.273)		
t or *z* test	3.249	1.490			3.349	1.316		
Population 20-24	0.028				0.034	0.046		
s.e.	(0.129)	(0.222)			(0.113)	(0.212)		
t or *z* test	0.221	0.130			0.303	0.217		
Population 60-64	1.037				0.745	0.817		
s.e.	(0.234)	(0.346)			(0.205)	(0.334)		
t or *z* test	4.440	3.000			3.637	2.444		
Population was in 1995 in Mexico City	0.126				0.131	0.132		
s.e.	(0.068)	(0.090)			(0.060)	(0.075)		
t or *z* test	1.854	1.400			2.177	1.768		
Own House	−0.101		−0.067		−0.066	−0.074	−0.030	−0.050
s.e.	(0.019)	(0.026)	(0.018)	(0.033)	(0.017)	(0.024)	(0.016)	(0.029)
t or *z* test	−5.231	−3.820	−3.631	−2.016	−3.828	−3.096	−1.832	−1.729

Distance Band (meters)								
720			900			1 080		
		Reduced model			*Reduced model*			*Reduced model*
	robust			robust			robust	
0.785	0.698	0.842	0.794	0.711	0.863	0.601	0.608	0.643
(0.034)	(0.063)	(0.030)	(0.039)	(0.066)	(0.033)	(0.068)	(0.060)	(0.070)
23.088	11.002	27.999	20.359	10.695	25.950	8.838	10.209	9.165
–15.895	–18.196	3.057	–19.080	–18.624	2.130	–24.249	–25.215	7.087
(10.139)	(13.342)	(4.580)	(10.402)	(13.504)	(4.705)	(10.687)	(14.563)	(5.172)
–1.568	–1.364	0.667	–1.834	–1.379	0.453	–2.269	–1.731	1.370
–1.550	–1.494	–0.427	–1.462	–1.495	–0.426	–1.450	–1.376	–0.634
(0.362)	(0.461)	(0.235)	(0.371)	(0.472)	(0.242)	(0.384)	(0.515)	(0.254)
–4.286	–3.240	–1.817	–3.946	–3.169	–1.765	–3.781	–2.671	–2.500
0.329	0.355		0.308	0.327		0.420	0.438	
(0.106)	(0.264)		(0.109)	(0.268)		(0.111)	(0.273)	
3.104	1.344		2.830	1.223		3.783	1.606	
0.022	0.018		0.070	0.052		0.003	–0.029	
(0.107)	(0.206)		(0.110)	(0.207)		(0.113)	(0.213)	
0.206	0.088		0.638	0.250		0.023	–0.138	
0.666	0.683		0.614	0.686		0.676	0.683	
(0.194)	(0.343)		(0.199)	(0.346)		(0.204)	(0.343)	
3.427	1.991		3.078	1.983		3.308	1.989	
0.076	0.094		0.105	0.108		0.150	0.147	
(0.057)	(0.068)		(0.059)	(0.069)		(0.060)	(0.077)	
1.334	1.377		1.782	1.565		2.507	1.925	
–0.034	–0.041	–0.004	–0.042	–0.047	0.007	–0.068	–0.064	–0.033
(0.016)	(0.023)	(0.015)	(0.017)	(0.023)	(0.016)	(0.017)	(0.024)	(0.016)
–2.101	–1.795	–0.279	–2.503	–2.030	0.463	–3.991	–2.702	–2.036

Table **4, continued...**

Indepedent variable	OLS		Reduced model		Distance Band (meters) 540		Reduced model	
		robust		robust		robust		robust
Households with computer	0.159				0.162	0.156		
s.e.	(0.042)	(0.069)			(0.037)	(0.060)		
t or z test	3.756	2.310			4.368	2.598		
Population economically dependent	0.318		0.241		0.246	0.263	0.164	0.186
s.e.	(0.081)	(0.105)	(0.075)	(0.122)	(0.071)	(0.091)	(0.066)	(0.088)
t or z test	3.912	3.020	3.222	1.971	3.450	2.889	2.486	2.103
R-squared	0.35		0.23		0.49	0.47	0.39	0.40
Log likelihood	−875.31		−900.54		−844.16	−846.50	−869.42	−864.60
Jarque-Bera (value)	51.250		334.912					
p-value	0.000		0.000					
Breusch-Pagan test (value)	76.610		13.156		53.940		3.438	
p-value	0.000		0.004		0.000		0.329	
Koenker-Basset (value)	38.960		3.667					
p-value	0.000		0.300					
Likelihood Ratio Test							62.233	
p-value							0.000	

Notes: the spatial lag model was estimated in Geoda 0.9.5-i(beta). The spatial lag model with robust errors was estimated in stata. Numbers in parenthesis indicate standard errors, *t* value is for ols and *z*-value is for the spatial lag models.

	Distance Band (meters)							
	720			900			1 080	
		Reduced model			Reduced model			Reduced model
	robust			robust			robust	
0.171	0.167		0.169	0.167		0.140	0.136	
(0.035)	(0.057)		(0.036)	(0.058)		(0.037)	(0.063)	
4.853	2.938		4.680	2.881		–3.782	2.158	
0.192	0.220	0.103	0.209	0.223	0.118	0.264	0.259	0.183
(0.068)	(0.088)	(0.062)	(0.070)	(0.088)	(0.064)	(0.071)	(0.092)	(0.065)
2.821	2.502	1.648	2.994	2.527	1.846	3.710	2.806	2.800
0.54	0.54	0.46	0.51	0.53	0.43	0.49	0.51	0.41
–830.05	–826.02	–854.28	–834.53	–828.87	–857.61	–845.10	–840.39	–868.021
61.400		0.609	54.020		3.279	77.870		6.404
0.000		0.894	0.000		0.351	0.000		0.094
83.770		92.510	81.560			60.510		65.035
0.000		0.000	0.000			0.000		

TABLE **5**
Spatial econometric results of model B (threshold): downtown 2000

| Independent variable | OLS | Distance Band (meters) | | | | |
		540	720	900	1 080
Self-employment (Rate of neighbors)		0.173	0.182	0.157	0.162
s.e.		(0.022)	(0.023)	(0.024)	(0.024)
t or *z* test		7.864	7.913	6.542	6.750
Intercept	−5.238	−10.118	−4.777	−8.598	−5.733
s.e.	(12.175)	(10.955)	(10.988)	(11.230)	(11.167)
t or *z* test	−0.430	−0.924	−0.435	−0.766	−0.513
Education years	−2.212	−1.713	−1.852	−1.815	−1.949
s.e.	(0.432)	(0.392)	(0.391)	(0.401)	(0.397)
t or *z* test	−5.120	−4.370	−4.737	−4.526	−4.909
Male 15-64	0.412	0.418	0.376	0.418	0.423
s.e.	(0.127)	(0.114)	(0.115)	(0.117)	(0.116)
t or *z* test	3.244	3.667	3.270	3.573	3.647
Population 20-24	0.028	−0.024	−0.004	−0.021	−0.044
s.e.	(0.129)	(0.116)	(0.116)	(0.119)	(0.118)
t or *z* test	0.217	−0.207	−0.034	−0.176	−0.373
Population 60-64	1.037	0.792	0.788	0.719	0.697
s.e.	(0.234)	(0.212)	(0.213)	(0.220)	(0.219)
t or *z* test	4.432	3.736	3.700	3.268	3.183
Population was in 1995 in Mexico	0.126	0.118	0.08	0.123	0.118
s.e.	(0.068)	(0.061)	(0.062)	(0.063)	(0.063)

TABLE 5, continued...

Independent variable	OLS	Distance Band (meters)			
		540	720	900	1 080
t or z test	1.853	1.934	1.290	1.952	1.873
Own House	-0.101	-0.084	-0.064	-0.07	-0.067
s.e.	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)
t or z test	-5.316	-4.667	-3.556	-3.889	-3.722
Households with computer	0.159	0.171	0.183	0.16	0.166
s.e.	(0.042)	(0.038)	(0.038)	(0.039)	(0.039)
t or z test	3.786	4.500	4.816	4.103	4.256
Population economically dependent	0.318	0.298	0.283	0.273	0.245
s.e.	(0.081)	(0.073)	(0.073)	(0.075)	(0.075)
t or z test	3.926	4.082	3.877	3.640	3.267
R-squared	0.35	0.46	0.46	0.43	0.44
Log likelihood	-875.31	-848.52	-849.07	-855.73	-854.02
Jarque-Bera (value)	51.250				
p-value	0.000				
Breusch-Pagan test (value)	76.610	57.29	60.79	75.72	76.7
p-value	0.000	0.000	0.000	0.000	0.000
Koenker-Basset (value)	38.960				
p-value	0.000				
Likelihood Ratio Test (value)		53.57	52.48	39.16	42.57
p-value		0.000	0.000	0.000	0.000

Note: numbers in parenthesis indicate standard errors, t value is for OLS and z-value is for the spatial lag models.

significant and positive in either spatial lag model (A or B) and, its inclusion is affecting the estimation of the covariate coefficients. Nevertheless, the social interaction coefficient in model A (local interaction) is more than twice than the value of the coefficient in model B (threshold).

The spatial lag models still have problems of heteroskedasticity as indicated by the Breusch-Pagan test, but this is not unexpected given that the data presents strong non-normality.[27] Nevertheless, note that a robust estimation (see column called "robust" in table 4) do not change substantially the results. Moreover, the presence of heteroskedasticity can be in part due to the fact that some of the sociodemographic covariates used in the analysis presumably are correlated. To consider this, see in table 4 that heteroskedasticity is not present either in the OLS reduced version or in the spatial lag model reduced version as it is indicated by the Koenker-Basset (KB) test and the Breusch-Pagan test respectively. However, non-normality continues being present in the model because the KB test does not rely on the assumption of normality. It is important to mention that significant changes in the social interaction term are not detected even when the robust and the reduced models are considered; for that reason, we consider that the results will no be drastically modified if more sophisticated and complex spatial methods that deal with non-normality were implemented. Similar conclusions have been reached in other researches that deal with similar characteristics in the data used in this essay (see Mobley *et al.*, 2006).

The important result of this spatial econometric exercise has been to show that the spatial interaction term is significant, but also it is important to mention some results of the covariate coefficients. Contrary to what the literature of entrepreneurship would expect (Evans and Jovanovic, 1989), liquidity constraints (see *own house* variable) has a null or negative impact on the rate of self-employment (both models produce similar results). Education

---

[27] Heteroskedasticity is not uncommon when dealing with regional data because of the irregular nature of the areas (or regions) analyzed (Arbia, 2006). Heteroskedasticity can also indicate that different spatial regimes must be considered; however, we dismiss this possibility because the area analyzed is relatively small (a portion of Mexico City).

is strongly significant and negatively correlated to self-employment and that economically dependent persons (children and elders) are significant with positive coefficient. The negative impact of education is consistent with other empirical studies that use micro-data to analyze entry-exit decisions on self-employment in Mexico (Woodruff, 1999).[28]

Another interesting result is that the rate of old people (60-65 years old) is consistently positive and statistically significant; this situation reinforces the hypothesis that conditions of poverty and weak or absent social security for the elderly increases the rate of self-employment. Also note that the proportion of the population living in the city in 1995 tends to be statistically significant in most of the band-distances. This variable controls indirectly for sorting effects. In general, similar results are produced by models A and B in the statistical significance and sign of the parameters. But the social interaction term differs strongly in magnitude between models: local interaction is in average four times larger than the parameter obtained in the threshold model. The rest of the covariates do not exhibit large discrepancies in magnitude between models, but the parameter of education is stronger in the threshold model. Maybe the important suggestion in the regression data would be that global information in Model B plays a role of stabilizer in the sense that it diminishes the effect of local interactions reflected in the size of the coefficients. If the threshold model is a proxy of the way in which institutions (or policy effects by the authority) are being incorporated at individual level (census tract), then the smaller magnitude in the social interaction term in model B might be consistent with the idea that institutions reduce uncertainty in the economic environment (North, 1990).

It is important to remark that I am not claiming identification of the social interaction term with the spatial econometric implementation. Problems of identification in these kinds of models are hard in cross-sectional settings (Manski. 1997). Nevertheless, the spatial lag model is a suitable econometric

---

[28] Calderón-Madrid (2000) using panel data for Mexico between 1995-98 (and calculating hazard rates from leaving one sector to another) found that people with formal and higher education spend less time in self-employed activities.

implementation that provides us with relevant information that enriches the simulation procedures of our models. Empirical issues (and calibration of the models) will be further addressed in the next section in a simulation framework.

SIMULATION, CLUSTER DYNAMICS AND EMPIRICAL ADEQUACY OF THE MODELS

Dynamics properties of the models are not possible to study in a spatial lag econometric framework. But the spatial lag implementation provides information to restrict parameter space in the census tract simulations and, with that, the possibility to analyze the dynamic properties of the models. Specifically, the simulations in the next section are designed to see how well the models match the global Moran's Index empirically observed in the self-employment rate in downtown Mexico City in 2000.

Conditions of the model

The average rate of self-employment per census tract observed in downtown is 19.72 (the aggregate self-employment rate in the area is 20.3).[29] In the simulations, I target an equilibrium outcome that is close to the average rate of self-employment observed in 2000. In order to replicate this equilibrium outcome, I use the information provided by the spatial lag model in cross section, which presumes to be at equilibrium, to restrict parameter space in models A and B.[30]

It is observed from tables 4 and 5 that the best fit (see log likelihood) in both models happens in general in the distances: 540, 720, and 900 meters. In the simulations, I use the mean value of the parameters of the spatial lag model that are generated in these distances. I am assuming that the values of the parameters are not affected by distance. Because the basic idea of the

---

[29] These rates do not differ from the rates observed in the whole city –see table 1.

[30] I showed previously that the spatial lag model identifies appropriately model A and B.

simulations is to evaluate the dynamics of global clustering that each model produces, I will use the Moran's Index calculated at equilibrium to monitor the global spatial autocorrelation of the self-employment rate in the region at each period of time. An error is also considered in the simulations. To implement that, each census tract incorporates an error that comes from a normal distribution with standard error equal to the average of the standard errors of the spatial lag model and with mean zero. Also, all census tracts adjust asynchronously their rate of self-employment once each period of time.[31] The initial conditions used are the empirical observations of the rate of self-employment in 2000, but this is somewhat irrelevant because random initial conditions do not affect the general results. Finally, information of the own characteristics of the census tracts comes from census data in 2000.

Results

The simulations with noise do not produce a fixed equilibrium (either in the Moran's Index or the rate of self-employment) but one that oscillates very close around a gravitational point. I average the Moran's Indexes generated in 200 periods of time (more periods of times do not statistically affect the results) in order to have an estimate to compare with. First, I contrast in Figure 5, the models with and without an error term.

What random components do is essentially diminish the level of global spatial autocorrelation in the region simulated. The Moran's Index in both models drops more than twice if compared when noise is not present but it continues to show high and statistically significant levels of positive spatial autocorrelation.[32] Nevertheless, in our models the Moran's Indexes continue to show strong spatial autocorrelation because pure local interaction

---

[31] This is equivalent to a uniform activation of the census tracts, which means that a census tract i acts immediately before census tract j over the course of a period of time.

[32] The loss of spatial autocorrelation when noise is present is in accordance with what the theoretical literature of social interactions has suggested when random behavior is introduced in systems that resemble the one analyzed in this research: noise is breaking out the clusters formed in systems based on local information (see Young, 1998; Ianni and Corradi, 2002).

Figure 5

**Effect of the error term on the Moran's Index
under different model specifications**
(average of 200 periods)
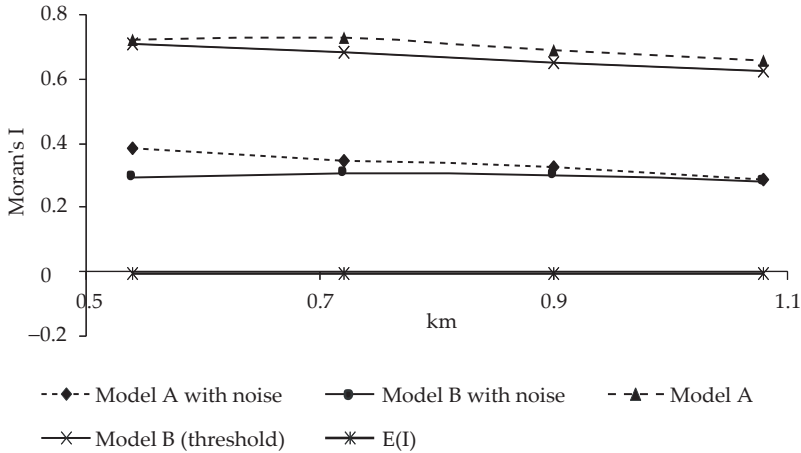
(model A) or global-local interaction (model B) in conjunction with the other variables (*i.e.*, census tracts' own characteristics) counterbalances the effect of noise maintaining the spatial clustering. The important result is that spatial autocorrelation of the self-employment rate continues being statistically significant even with random shocks. Now it is important to ask whether the models are reproducing the global spatial autocorrelation empirically observed in the self-employment rate. If we assume that the self-employment rate observed in 2000 is at equilibrium, then we can compare its Moran's Index with the one produced by the models. Figure 6 displays these comparative results between models and empirical data. The figure also contains a series of the models without the social interaction term (no local interaction or no global-local interaction effects) and that were produced with OLS estimation (*i.e.*, no social interaction term).

First, note that models of pure local interaction and threshold are much closer to the empirical data than the rest of the specifications. Then, a

Figure **6**
Moran's Index generated in different models
(average of 200 periods)
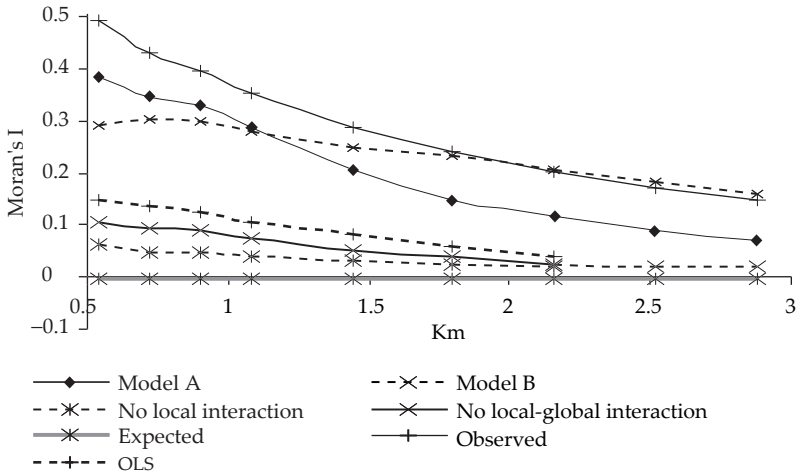
simple comparative static exercise indicates that the social interaction term is fundamental to reproduce the spatial autocorrelation observed. It is true that in short band distances (less than 1.5 km), the series without the social interaction term produce statistically significant spatial autocorrelation (own characteristics and error contribute to that), but they are quite low to be close to what is empirically observed. This gives an insight regarding the role of pure local interactions: it happens that local conformity under the presence of own characteristics variables triggers spatial autocorrelation.

Spatial autocorrelation of self-employment decreases in all models as the distance increases. But while it is clear that the models without the social interaction term reach spatial randomness with relatively short band distances, this situation does not occur with models A or B.

The dynamics of model A and B display interesting dynamics in figure 6, model A (local conformity) produces higher spatial autocorrelation than model B when the distance is smaller than 1 km; consequently, the series

are closer to the empirical spatial autocorrelation. This is consistent with the results of the econometric implementation (see last section) where model A has a better fit than model B in these band distances. But, interestingly, the behavior of both models is reversed if the band of distance is greater than 1 km: model B (threshold) not only produces higher spatial autocorrelation but its simulated outcomes are also close to the empirical ones.

From the point of view of reproducing local clusters, the latter results are consistent with local Moran's Index. In figure 7, I show typical LISA maps realizations of simulations in models A and B and OLS specification (initial conditions are random in all models). The simulated maps are contrasted with the empirical one. I highlight LISA statistics with 1% pseudo-significance level and the band distance used is 1 800 meters.

From figure 7, it is clear that models A and B reproduce better the empirical cluster than the OLS specification, which does not include the social interaction term. But model B (threshold) seems to reproduce better the empirical cluster. The local results are consistent with the global spatial autocorrelation generated by the models in these particular realizations: Model B generates the highest Moran's Index (0.183), which is closer to the empirical one (0.2404).

The main lesson of the simulations is that a model that considers global-local information (model B) fits real data better than a simple model of local information if a relatively large distance network structure is influencing the rate of self-employment (more than 1.5 km). This suggests, given the assumptions of model B, that self-employment rates at census tract level not only are influenced by local information (provided by the neighborhood), but also by global entities that process global information. Governmental propaganda, publicity and activism of local leaders can be means of transmission of global information.[33]

---

[33] For example, the political power of the leaders of the street peddlers in downtown Mexico City is widely recognized. Likewise, the federal governmental propaganda to promote self-employed activities (*changarros*) are also well recognized.

FIGURE 7
LISA maps of typical realizations under different models
(band distance 1800 meters)
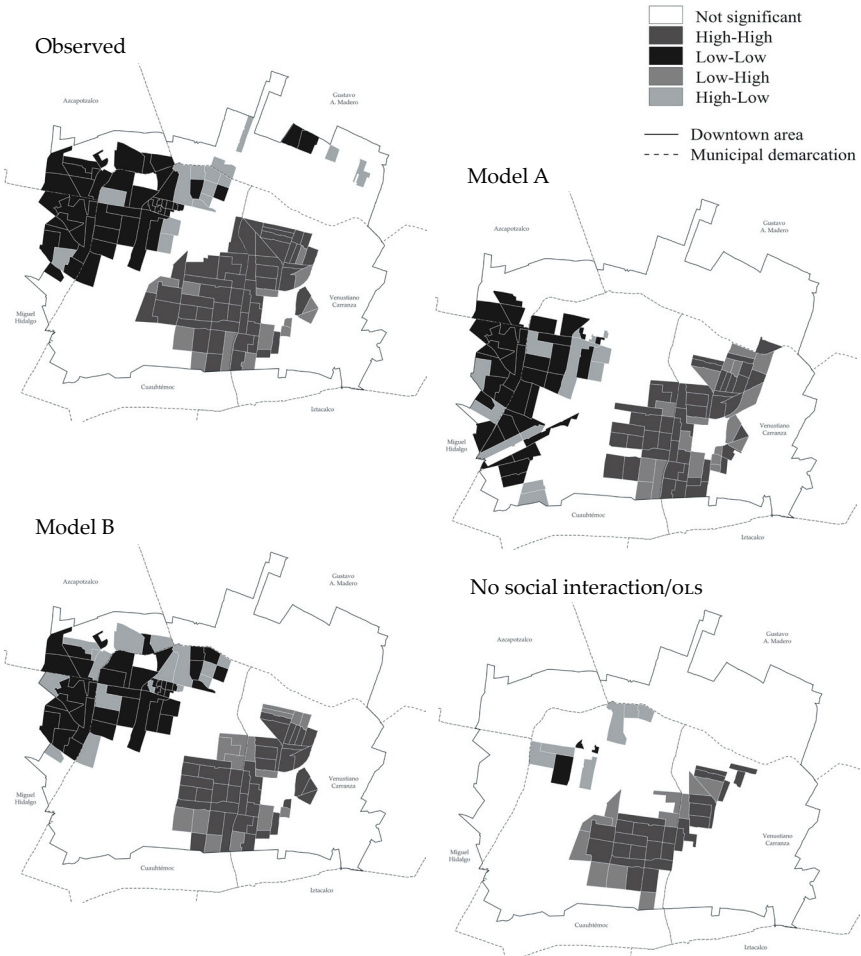
Notes: initial conditions are random in all models. The intensity of the color in each census tract is associated to a pseudo significant LISA ($p < 0.01$). No color: no significant spatial autocorrelation; dark gray: Low-Low spot of spatial autocorrelation; medium gray: High-High spot of spatial autocorrelation; light gray: Low-High spot of spatial autocorrelation; lightest gray: High-Low spot of spatial autocorrelation.

If both formal and informal institutions are processing and returning global information to the system, they can reduce the uncertainty about the global environment (*i.e.*, less global spatial autocorrelation). At the same time, they are perpetuating polarization in the clusters (*i.e.*, an increase of the local spatial clusters). This conclusion is consistent with the notion of local conformity and global diversity postulated by the literature of social interactions and institutions (Young, 1998).

Concluding remarks

This research contributes to understanding self-employment in developing countries because it advances a hypothesis so far not considered in the literature: entry-exit decisions in self-employment activities are also influenced by non-market interactions. In particular, this paper proposed through implementing a census tract based model with data of downtown Mexico City from 2000 that increasing returns from conformity are present in individuals when they are facing entry-exit decisions in the self-employment sector. Likewise, in this essay it is suggested that the effect of social influence must not be only restricted to local informational dynamics (neighborhoods) because channels of global information can also shape the effects of social influence. This situation remits us to the possibility that a threshold behavior might be present in the decisions of entry in self-employment activities; a situation that is reminiscent of social scientists understanding of collective action phenomena (such as joining a revolt, voting for a party, etc.).

An additional point is that the presence of social contagion in self-employment activities must not be restricted only to environments where informality and poverty are common (like Mexico City); the thesis certainly must apply in more advanced countries too. Further research in this area is necessary. Due to lack of information, this study relies on census tract data for the year 2000. However, the methodology of this paper might be adjusted to incorporate household level data and better information set regarding entry-exit decisions in self-employment. Further research must

address this possibility. Finally, the fact that social interactions can be present in self-employment activities can have strong policy implications to either boost entrepreneurship or to combat poverty traps; moreover, the fact that our results indicate that social interactions in self-employment entry-exit decisions operate through channels of local and global information, raises the issue of the viability of the authorities intervention to regulate the phenomena. However, in order to figure out the way in which policy could be effective, it is necessary to have more precise micro information of the local dynamics.

References

Anselin, L., "Local indicators of spatial association: LISA", *Geographical Analysis,* vol. 27, no. 2, 1995, pp. 93-115.

Anselin, L, "Spatial econometrics", in B. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Oxford, Basil Blackwell, 2001, pp. 310-330.

Anselin, L., "Under the hood. Issues in the specification and interpretation of spatial regression models", *Agricultural Economics*, vol. 27, no. 3, 2002, pp. 247-267.

Arbia, G., *Spatial Econometrics*, Berlin, Springer, 2006.

Arrow, K.J. and R. Borzekowski, "Limited network connections and the distribution of wages", The Federal Reserve Board, Finance and Economics Discussion Series, Working Paper no. 2004-41, 2004.

Brueckner, J.K., "Strategic interaction among governments: an overview of empirical studies", *International Regional Science Review*, vol. 26, 2003, pp. 175-188.

Calderón-Madrid, A., "Job stability and labor mobility in urban Mexico: a study based on duration models and transition analysis", Inter-American Development Bank, Research Network, Working Paper no. R-419, 2000.

Case, A.C., "Spatial patterns in household demand", *Econometrica*, vol. 59, 1991, pp. 953-965.

Calvó-Armengol, A. and M. Jackson, "The effects of social networks on employment and inequality", *American Economic Review*, vol. 94, 2004, pp. 426-454.

D'Agostino, R.B., A. Balanger and R.B. D'Agostino Jr., "A suggestion for using powerful and informative tests of normality", *American Statistician*, vol. 44, 1990, pp. 316-321.

De Soto, H., *The Other Path*, New York, Basic Books, 1989.

Eatwell, J., "Effective demand and disguised unemployment", in J. Michie and J. Grieve Smith (eds.), *Employment and Economic Performance: jobs, inflation and growth*, Oxford, Oxford University Press, 1997.

Evans, D.S. and B. Jovanovic, "An estimated model of entrepreneurial choice under liquidity constraints", *Journal of Political Economy*, vol. 97, 1989, pp. 808-827.

Florax, J.G.M., H. Folmer and S.J. Rey, "Specification searches in spatial econometrics: the relevance of Hendry's methodology", *Regional Science and Urban Economics*, vol. 33, 2003, pp. 557-579.

Galli, R. and D. Kucera, "Informal employment in Latin America: movements over business cycles and the effects of worker rights", International Institute for Labour Studies, Discussion Paper no. 145, 2003.

Giannetti, M. and A. Simonov, "On the determinants of entrepreneurial activity: individual characteristics, economic environment and social norms", *Swedish Economic Policy Review*, vol. 11, no. 2, 2004, pp. 269-313.

Glaeser, E.L., B.I. Sacerdote and J.A. Scheinkman, "The social multiplier", *Journal of the European Economic Association*, vol. 1, no. 2-3, 2003, pp. 345-353.

Granovetter, M., "Threshold models of diffusion and collective behavior", *American Journal of Sociology*, vol. 83, 1978, pp. 1420-1443.

Granovetter, M., "The economic sociology of firms and entrepreneurs", in A. Portes (ed.), *The Economic Sociology of Immigration*, New York, Russell Sage Foundation, 1995.

Harris, J.R. and M.P. Todaro, "Migration, unemployment and development: a two sector analysis", *American Economic Review*, vol. 60, no. 1, 1970, pp. 126-142.

Hurst, E. and A. Lusardi, "Liquidity constraints, household wealth and entrepreneurship", *Journal of Political Economy*, vol. 112, 2004, pp. 319-347.

Ianni, A. and V. Corradi, "The dynamics of public opinion under majority rules", *Review of Economic Design*, vol. 7, 2002, pp. 257-277.

Instituto Nacional de Estadística, Geografía e Informática (INEGI), Sistema para la Consulta de Información Censal (SCINCE) 1990 and 2000 (compact disc).

Light, I. and C. Rosenstein, "Expanding the interaction theory of entrepreneurship", in A. Portes (ed.), *The Economic Sociology of Immigration*, New York, Russell Sage Foundation, 1995.

Lucas, R.E. Jr., "On the size distribution of business firms", *Bell Journal of Economics*, vol. 9, 1978, pp. 508-523.

Maloney, W.F., "Informality revisited", *World Development*, vol. 32, 2004, pp. 1159-1178.

Manski, C.F., "Identification of anonymous endogenous interactions", in W.B. Arthur, S.N. Durlauf and D.A. Lane (eds.), *The Economy as an Evolving Complex System II*, United States, Addison-Wesley, 1997, pp. 369-384.

Mobley, L.R. *et al.*, "Spatial analysis of elderly access to primary care services", *International Journal of Health Geographics*, vol. 5, no. 19, 2006.

North, D.C., *Institutions, Institutional Change and Economic Performance.* Cambridge, Cambridge University Press, 1990.

Parker, S.C., *The Economics of Self-employment & Entrepreneurship*, Cambridge, Cambridge University Press, 2004.

Piore, M. (ed.), *Unemployment and Inflation,* White Plains, M.E. Sharpe, 1979.

Pisati, M., "Tools for spatial data analysis", *Stata Technical Bulletin*, no. 60, 2001, pp. 21-37

Portes, A. and P. Landolt, "Social capital: promise and pitfalls of its role in development", *Journal of Latin American Studies*, vol. 32, 2000, pp. 529-547.

Salas, C., "Trayectorias laborales entre el empleo, el desempleo y las microunidades en México", *Papeles de Población*, no. 38, 2003, pp. 121-157.

Schelling, T., *Micromotives and Macrobehavior*, United States, W.W. Norton & Company, 1978.

Semmler, W., "On the mechanisms of inequality", Center for Empirical Macroeconomics, Department of Economics, University of Bielefeld, Working Paper no. 134 2003.

Smirnov, O. and L. Anselin, "Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach", *Computational Statistics and Data Analysis*, vol. 35, 2001, pp. 301-319.

Tobler, W., "A computer movie simulating urban growth in the Detroit region", *Economic Geography*, vol. 46, 1970, pp. 234-240.

Topa, G., "Social interactions, local spillovers and unemployment", *Review of Economic Studies*, vol. 68, 2001, pp. 261-295.

Woodruff, Ch., "Can any small firm grow up? Entrepreneurship and family background in Mexico", Graduate School of International Relations and Pacific Studies, University of California, Working Paper, 1999.

Young, H.P., *Individual Strategy and Social Structure*, Princeton, Princeton University Press, 1998.

Appendix: Moran's Index and lisa statistics

Moran's Index is a global measure of spatial autocorrelation (one statistic is derived for the entire study area) that relates spatial proximity between regions. The expression is:

$$I = \frac{n \sum\limits_{i}^{n} \sum\limits_{j}^{n} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{(\sum\limits_{i}^{n} \sum\limits_{j}^{n} w_{ij}) \sum\limits_{i}^{n} (x_i - \bar{x})^2}$$

where n are the number of observations, w_{ij} is a measure of the spatial proximity between regions i and j, and xi is the value of the attribute of interest for location i. In this paper w_{ij} is binary and it is based on distances: if a census tract is inside the radius of distance considered, w_{ij} takes value of one, otherwise is zero. In order to normalize the outside influence upon each region, I use row-standardized weights such that the elements w_{ij} in each row sum to 1, consequently

$$\sum\limits_{i}^{n} \sum\limits_{j}^{n} w_{ij} = n$$

The value of Moran's Index ranges from –1 for negative spatial autocorrelation to 1 for positive spatial autocorrelation. No spatial autocorrelation is indicated by the expected value of Moran's Index, that is,

$$E_I = - \frac{1}{(n-1)}$$

To test statistically that there is spatial autocorrelation, one needs to test the hypothesis null of spatial randomness. It is assumed that the sampling

distribution of the Moran's Index has a normal distribution under the (null) hypothesis of no spatial pattern. Commonly two sampling assumptions are made for Moran's Index: normality or randomization. The first assumes that each region arises from normal distributions that have the same mean and variance in each region; in the second (randomization), the set of values is fixed but the locations associated to each value is not fixed. Consequently, it is assumed that all possible permutations of the regional values are equally likely. In such procedure, each time a new permutation is run, a pseudo-significance p-value is obtained; that means that the results are not exactly replicable. The tests of pseudo-significance were performed in the package GEODA095I-6.

If we plots the weighted average of the neighbors of the variable of interest against the values of the variable, we obtain the so called Moran Scatterplot; that is, plotting

$$\sum_{j} w_{ij} (x_j - \bar{x})$$

against $(x_j - \bar{x})$. The scatterplot indicates local spatially instability; that is, local deviations from global pattern of association. The scatterplot generates four quadrants between an observation and its neighbors: *a*) High-High which means that an observation (census tract) with a value above the mean is surrounded by observations (census tracts) with values above mean; *b*) Low-Low, an observation (census tract) with a value below the mean is surrounded by observations (census tracts) with values below mean; *c*) High-Low, which means that an observation (census tract) with a value above the mean is surrounded by observations (census tracts) with values below mean; and finally *d*) Low-High, which means that an observation (census tract) with a value below the mean is surrounded by observations (census tracts) with values above mean. If z is the vector of x_i in deviation from the mean and W is the row standardized weight matrix, then it is the case that the slope coefficient of the linear regression of Wz on z is the Moran's Index.

The latter brings us to the local Moran statistic:

$$I_i = \frac{(x_i - \bar{x})\sum\limits_{i} w_{ij}(x_j - \bar{x})}{\sum\limits_{i} w_{ij}(x_i - \bar{x})^2/n}$$

The sum of the local Moran's is equal to the global Moran ($\sum I = I$). Then there is a clear connection between the local indicator and the global indicator of spatial autocorrelation. LISA statistics can be subject to inferential procedures. In that way the local spatial instabilities visualized in the Moran Scatterplot can be tested, and significant local clusters can also be detected even in absence of global autocorrelation (Hot Spots). In this essay, I also assess pseudo-significance of the local Moran's through random permutations.[34]

---

[34] For a calculation of the variance I_i see Anselin, 1995.