TEXTO PARA DISCUSSÃO Nº 204

SELF-EMPLOYMENT IN BRAZIL AND ITS DETERMINANTS: A SPATIAL ANALYSIS

Sueli Moro
Flavia Chein
Ana Flávia Machado

Junho de 2003

Ficha catalográfica

331.5(81) Moro, Sueli.
M867s Self-employment in Brazil and its determinants: a spatial
2003 analysis / por Sueli Moro; Flavia Chein; Ana Flávia Machado. - Belo Horizonte: UFMG/Cedeplar, 2003.
24p. (Texto para discussão ; 204)
1. Setor informal (Economia) – Brasil – Modelos econométricos. 2. Brasil – População. I Chein, Flavia II. Machado, Ana Flávia. III. Universidade Federal de Minas Gerais. Centro de Desenvolvimento e Planejamento Regional. IV. Título. V. Série.

CDU

UNIVERSIDADE FEDERAL MINAS GERAIS
FACULDADE DE CIÊNCIAS ECONÔMICAS
CENTRO DE DESENVOLVIMENTO E PLANEJAMENTO REGIONAL

SELF-EMPLOYMENT IN BRAZIL AND ITS DETERMINANTS: A SPATIAL ANALYSIS

Sueli Moro
Professor at CEDEPLAR, Federal University of Minas Gerais, Brazil.

Flavia Chein
PhD student at CEDEPLAR, Federal University of Minas Gerais, Brazil.
We are grateful to the Msc students at CEDEPLAR:
Helger Marra Lopes and Sandro Monsueto.

Ana Flávia Machado
Professor at CEDEPLAR, Federal University of Minas Gerais, Brazil.

CEDEPLAR/FACE/UFMG
BELO HORIZONTE
2003

SUMÁRIO

ABSTRACT

In this paper, we intend to identify some factors underlying the different rates of participation of self-employed workers in Brazilian municipalities. In contrast with previous analysis, our study is sensitive to the spatial dimension of self-employment in Brazil. Relying on geo-referenced data from areal units, our approach avoids the restrictive assumptions of independence between area effects. Two different approaches to spatial analysis are exploited, Exploratory Spatial Data Analysis (ESDA) and Spatial Econometrics. We show that substantial geographical heterogeneity in rates of self-employment exists within Brazil. The results point to the presence of clusters of municipalities with similar rates of participation of self-employment in the economy. The fitting of a spatial lag model to the data demonstrates that significant part of variation in self-employment is explained by spatial dependence. In addition, the municipality's degree of urbanization, rate of employment in the secondary sector as well as GDP per capita among others, indeed influences its rate of self-employment. We conclude that spatial analysis can provide useful insights in analyzing self-employment in Brazil.

I. INTRODUCTION

A large amount of recent studies concerning development has also focused on labor market. The reason is that individual's main income is related to employment opportunities. In addition, both income and job "quality" have an effect on social welfare.

In this context, when the formal sector is unable to absorb the existing labor force, the informal sector turns out to be the main option to survival in developing countries.

The aim of this study is to shed some light on the determinants of the participation in the informal labor market in Brazil. The main purpose is to identify the factors underlying the different rates of participation of self-employed workers in Brazilian municipalities. By focusing our study on Brazilian municipalities, we intend to verify if some of the theoretical arguments used for countries can be replicated to more disaggregated levels.

As opposed to conventional analysis, our study uses the tools of spatial analysis focusing on the spatial context where the observations were generated. In other words, we are interested in the geographic or spatial dimension of self-employment in Brazil.

Brazil is an interesting case for researchers interested in geographical differences as social and economic heterogeneity matches geographical heterogeneity. We are mainly interested in answering the question: are the structural determinants of self-employment rates "invariant" across space or the structural conditions exert differing effects on self-employment in various sub-regions of Brazilian geography?

To the extent that we will be using data from areal units, there are some important methodological issues to be considered. Areal units are not independent. Hence, we cannot base our study on statistical methods that assume independency of observations, such as OLS.

Econometric relationships in space have two main features: spatial heterogeneity and spatial dependency. Spatial heterogeneity refers to the fact that econometric relationships may vary systematically over space. Spatial dependency or its empirical expression spatial autocorrelation is a situation where the sample observation in one location is correlated with observations in nearby locations. Thus, in this study we apply the formal tools of spatial analysis in an attempt to explain and describe variations in the rates of self-employment.

Two different approaches to spatial analysis are exploited in this study, Exploratory Spatial Data Analysis (ESDA) and Spatial Econometrics. The ESDA extends standard EDA by focusing on techniques to describe spatial distributions, discover patterns of spatial association, identify spatial outliers, etc. On the other hand, Spatial Econometrics methodology deals with the effects of spatial dependence and spatial heterogeneity in regression analysis. Our intention is also to verify in what extent the application of spatial analysis techniques may contribute to a better understanding of self-employment in Brazil.

We begin with a brief review of the literature on the informal sector and it's relationship to development. This is followed by an appraisal of the informal sector in Brazil. In section 4, we introduce the spatial techniques applied to the empirical part of the study. Section 5 presents the data base and the model to be estimated. Section 6 contains some descriptive statistics and the results of the estimation of the empirical models. Section 7 presents some concluding remarks and topics for future research.

II. THE INFORMAL SECTOR: MAIN CHARACTERISTICS AND IMPACTS

Over the last decades, the informal sector has been attracting the attention of researchers interested in measuring the social welfare. Researchers are motivated by the fact that the informal labor market is often characterized by lack of security and inferior jobs.

According to Barreto (2002) and Castells & Portes (1989), the low quality of the job position has a negative influence on welfare. This happens mainly because of the low productivity and inferior wages associated with the informal sector. In addition, when dealing with the informal labor market it's important to consider the loss of the institutional power, which was acquired by workers after several decades of struggle.

However, the informal economy does not include only subsistence activities undertaken by deprived individuals. The informal sector includes also a number of dynamic activities carried on by entrepreneurs with relatively high incomes (Castells & Portes, 1989).

In this sense, it's essential to distinguish informal economy, which is a specific form of production, from poverty, which is related to the distribution process (Castells & Portes, 1989). Therefore, although the majority of workers in the informal sector are poor, mostly in developing countries, the informal economy surpasses the social structure.

We ought to clarify what we consider as informal sector in this study. In view of the difficulties in obtaining information from census data about both familial work and small business, we consider as informal sector only workers who declare themselves as self-occupied, excluding the liberal professionals.

Noorderhaven et al. (1999), hypothesized about the proportion of self-employment in several countries. They based their conjectures on variables related to life-dissatisfaction and uncertainty avoidance. According to them, higher levels of dissatisfaction whether with one's own life, or with the implementation of democracy, are associated with higher (elevated) rates of self-employment. On the other hand, people living in countries characterized by strong aversion to uncertainty, are emotionally inclined to require more rules and procedures. Consequently, they tend to stay longer in their jobs. In these countries, the transition from formal employment (or unemployment) to self-employment is delayed comparatively to other countries.

Further hypothesis formulated by the authors are: (a) significant differentials between earnings in self-employment and permanent jobs are associated to higher rates of self-employment; (b) countries characterized by weak aversion to uncertainty show both wage differentials and self employment rates larger than countries where the aversion to uncertainty is strong; (c) elevated rates of unemployment in a country are associated with high levels of self-employment. This happens because the opportunity cost of switching from unemployment to self-employment is relatively low.[1]

Noorderhaven et al. also consider the relationship between rates of self-employment and prosperity. Higher levels of prosperity are related to lower rates of self-employment. In developing countries, both wages and pressure for efficiency are lower. Consequently, small firms tend to prevail

[1] Even after considering for the difficulties of starting a new business, in countries experiencing high rates of unemployment and self- employment.

in the economy. As the economy expands, the wages increase and there's a stronger pressure to exploit scale and scope economies.

However, the tendency to lower rates of self-employment can be weakened or even reversed in more advanced stages of economic development, when the services sector would become more important comparatively to the secondary sector.

III. DEVELOPMENT *VERSUS* INFORMAL SECTOR*:* THE BRAZILIAN CASE

The industrialization of Brazil, concentrated in both South and Southeast, contributed to the intensification of the regional disparities. Those industrialized regions experienced an increase in the size of the formal sector, followed by an improvement in wages, working and life conditions.

On the other hand, in the less industrialized regions, the informal sector contributes with a comparatively larger proportion of the labor market. The informal sector in these regions is chiefly composed of ordinary jobs in both commerce and services sector. Therefore, low earnings tend to prevail along with worse working and life conditions.

Jorge & Valadão (2002), in their research on the characteristics of the informal sector in Brazil noticed a prevalence of firms with low average profit in the Northeast while firms in the Southeast benefited from higher average profits.

The occupational structure contributes to the heterogeneity in the labor market and the technological progress stimulates its segmentation. Technological progress differentiates between low productivity occupations that do not demand much training from high productivity occupations that require skilled work force (Machado, 1993).

Concerning the macro-economic context, Cardoso Jr. & Fernandes (2002), analyzed the performance of the two sectors during the economic recessions, at the beginning of the 80's and 90's. They noticed that the relative proportion of formal workers tended to decline during those periods and to increase slightly in the subsequent periods of recovery. Nevertheless, the way the revitalization of the formal economy took place was different in the two periods. During the economic recovery of the 80's, that initiated in 84, the levels of formal employment raised above those from the years of economic recession. However, in the recovery that took place in 93, the previous tendency was barely attenuated, exhibiting signs of stabilization around 1996.

The authors found also that in periods of recession, the informal employment, tended to increase whereas during the recovery it declined slightly. In Brazil, informal workers are referred to as workers who don't have a signed labor card ("trabalhador sem carteira assinada") .

As to self-employed workers, their behavior was similar to the workers without a labor card. After the economic recession of the beginning of the 90's, the participation of this group in the total labor force increased drastically, matching the proportion of workers without a labor card.

It's precisely in this context that this analysis was conducted. The study is anchored in Noorderhaven's hypothesis of an inverse relationship between prosperity levels and self-employment.

IV. DATABASE AND EMPIRICAL MODEL

The sources of information for this study were, the Brazilian census for 2000, the Profile of the Municipalities for 1999 from IBGE[2], and the Human Development Index (HDI) calculated by PNUD/IPEA/IBGE for 2000.

The human development index was included only in the exploratory spatial analysis, because of multicollinearity problems. This index is a composite of variables representing three aspects of human development: longevity, education, and a decent standard of living. Longevity is measured by life expectancy at birth; education is measured by a combination of the adult literacy rate and the combined gross primary, secondary, and tertiary enrollment ratio; and standard of living, as measured by GDP per capita (PPP US) [3].

The dependent variable in the model was the rate of self-employment in Brazilian municipalities. The selection of independent variables was based on Noorderhaven's work. Although employment in the informal sector is voluntary, the individual's option for self-employment is influenced by the economic structure of the geographic area where he lives. Consequently, it's important, according to Noorderhaven et al., to identify indicators of economic development that must be controlled when analyzing motivation to self-employment.

Therefore, the independent variables in the model include measures of both, degree of development and urban organization in the municipalities.

The variables included in the model to account for the level of development were, the rate of urbanization, the GDP per capita, the number of houses in *favelas* and the number of libraries in the municipality. The last two were included as indicators of social development in the municipalities.

The degree of formalization in the economy was taken into account by the rate of employment in the secondary sector, calculated as the total employment in the secondary sector divided by the total workforce.

The indicators of the level of urban organization in the municipalities were variables related to the presence of plans[4], consortiums (health, education, housing, etc) and programs (income, professional training, etc) implemented by the local government.

The empirical model is the following:

$$SE = \beta_1 + \beta_2 URB + \beta_3 TX\ 2 + \beta_4 PC\ 96 + \beta_5 LB + \beta_6 FAV + \beta_7 DCONS + \beta_8 DPLAN + \beta_9 DPROG + \varepsilon$$

[2] CD-ROM Perfil de Informações Municipais 1999 – IBGE

[3] UNDP. Technical Notes about HDI. In: www. undp.org.

[4] Here the plans considered are mainly the Director Plan (Plano Diretor), Multi-year Plan (Plano Plurianual) and the Stategic Plan (Plano estrategico).

Where, **SE** is the rate of self-employment, **URB** is the rate of urbanization, **SEC** is the rate of employment in the secondary sector, **PC96** is the GDP per capita in 96, **LB** is the number of libraries in the municipality, **FAV** is the number of houses in *favelas*, **DPLAN** is a dummy variable that indicates the plans implemented by the local government [5], **DCONS** is a dummy variable for the presence of consortium and finally **DPROG** is a dummy variable that accounts for the presence of social programs carried out by the local government.

We expect a negative relationship between the rate of self-employment and the following variables: PC96 and URB, meaning that the more developed the municipality the lower rates of self-employment. The variable **SEC** was included in the model to capture the degree of formalization of the economy and to test the hypothesis that self-employment occurs mainly in primary and tertiary sectors. A negative value for the estimated coefficient of this variable would confirm this hypothesis. Concerning the variables indicating the level of urban organization, DPLANS, DCONS and DPROG, we expect that the more structured the municipalities the lower the rate of self-employment. For the variables representing the social development, LB and FAV, we expect a negative association to the rate of self-employment. The reason is that, the more developed the municipality the higher the number of both *favelas* and libraries. In Brazil, the *favelas* are present in almost every large urban center; consequently their presence may be associated to the size of the municipality[6].

V. SPATIAL METHODOLOGIES

The spatial analysis of data differs from standard analysis by taking into account the spatial distribution of the observations. The spatial structure in the data is usually embodied in a contiguity weights matrix, W, with elements w_{ij}, where the index ij corresponds to the neighbor i from the observation j. The presence of zero value in this matrix indicates the absence of spatial interaction among observations.

Exploratory Spatial Data Analysis

The exploratory spatial data analysis (ESDA) is a field of exploratory data analysis (EDA) that deals with georeferenced data. It is an ensemble of techniques to describe and visualize spatial distributions, detect patterns of spatial associations, suggest spatial forms of spatial heterogeneity and identify spatial outliers (Anselin, 1999). The tools of spatial analysis facilitate the identification of spatial regimes that are likely to exhibit distinct causal processes.

The specific ESDA technique used in this study is the Moran Scatterplot, which pertains to the group of the LISA statistics (Local Indicators of Spatial Association) [Anselin, 1995]. The LISA statistics differs from the usual measures of global spatial autocorrelation, as Moran's I and/or Geary's c, in the sense that it's able to detect local patterns of spatial autocorrelation.

[5] This dummy variable takes the value equal to one if the municipality has at least two plans and zero otherwise.

[6] The *favelas* are a consequence of the inequalities that characterizes Brazilian economic development.

The statistic Moran's I[7] may be interpreted by means of a linear regression, of y on Wy, where W is the spatial weights matrix. The slope in this regression, ρ, is an indicator of the extent of the spatial autocorrelation between each observation and the average of it's neighbors. It's value corresponds to the global Moran's I statistic and it's graphic visualization is the Moran Scatterplot. In a graphical representation of a Moran Scatterplot, the first and the third quadrants represent the positive spatial autocorrelation, in the sense that high (low) values for one observation are surrounded high (low) values for the average of the neighbors. The second and fourth quadrants indicate negative spatial autocorrelation, which means that high (low) values for one observation are surrounded by low (high) values for the averages of the neighbors[8]. The relative density of the first and third quadrants determinates the global intensity of the positive spatial autocorrelation. On the other hand, the observations in the second and fourth quadrants suggest patterns of spatial heterogeneity (non-stationarity) in the data.

Spatial Econometric Models

Spatial Econometrics is a sub field of econometrics that deals with the treatment of spatial dependence and spatial heterogeneity in regression models [Paelinck & Klassen (1979), Anselin (1988a)]. Both problems arise when the data used in the regression model has a locational component and violates the assumptions of standard linear regression models, respectively uncorrelated error terms and constant variance.

Spatial heterogeneity refers to structural instability in the form of non-constant error variance (heteroskedasticity) and changing coefficients across space. The heteroskedasticity in a regression model may be due to both regional heterogeneity and induced by the spatial autocorrelation in the model. A major distinction between spatial and non-spatial processes is that the diagonal in the error covariance matrix is not constant even with i.i.d. error terms. Therefore, tests for heteroskedasticity may be misleading (Anselin and Griffith, 1988). Spatial heterogeneity in a regression models is usually handled by means of standard econometrics techniques.

The spatial dependence in a regression model, or its empirical expression spatial autocorrelation, is a situation where the values of the dependent variable and/or the error term in one location are correlated with the value of the corresponding observations in nearby locations.

The spatial dependence can be incorporated in a regression model in two different ways: as a spatial lag or as a spatially correlated error term. The first alternative is formalized as a mixed regression model:

$$y = \rho Wy + X\beta + \varepsilon$$

[7] $I = N/S_0 \Sigma_i \Sigma_j w_{ij} \cdot (x_i - \mu) \cdot (x_j - \mu) / \Sigma_i (x_i - \mu)^2$

[8] This information must be read as: a high value of the variable in a municipality is surrounded by a high value for the average of the neighbors.

Where the parameter ρ reflects the spatial dependence measuring the average influence of the neighbors on observations in vector y; its estimation gives the proportion of total variation in y explained by the spatial dependence. The reduced form of this model shows that Wy is correlated with the disturbance even when the latter are i.i.d. (Anselin, 1999). The spatial lag term must be treated as an endogenous variable and the estimation method has to account for this endogeneity (Anselin, 1999).

The spatial dependency in the error term is expressed by a spatial autoregressive or moving average error process. The spatial autoregressive is formalized as follows:

$$Y = X\beta + \varepsilon \qquad \varepsilon = \lambda W\varepsilon + \xi$$

The error terms in this model are non-spherical, in the sense that the off-diagonal elements in their covariance matrix are spatially correlated.

The consequences of ignoring the spatial autocorrelation in regression models depend on the alternative hypothesis. If the alternative hypothesis is a lag model, the OLS estimator of the standard model will be biased and inconsistent. On the other hand, if the alternative hypothesis is an error model, the consequences will be the same as in the case of serial auto-correlated error terms, i.e. the OLS estimator will remain unbiased, but it will no longer be efficient (Anselin, 1999).

To the extent that ignoring the spatial autocorrelation in regression models is something analytically serious, the specification tests plays an important role in Spatial Econometrics. The Morans' I is the most popular specification test for spatial autocorrelation. However, it is sensitive to the presence of non-normality and heteroskedasticity of the error terms. Additional tests for spatial autocorrelation are Kelejian & Robinson (1998, 1999), Likelihood Ratio (LR) tests for inference on the spatial autoregressive coefficients [(Anselin, (1988a), Anselin and Bera, (1998)], amongst others. Lagrange Multiplier (LM) or Rao Score tests that allows for the distinction between spatial error and spatial lag models, outlined by several authors, LM test against spatial error alternative (Burridge, 1980); LM test against spatial lag alternative (Anselin, 1988b) and more recently the robust tests, a class of tests that are robust to the presence of local misspecification of the other form [Bera and Yoon, (1993), Anselin et alii, (1996)].

The estimation of models of spatial dependency must take into account that these models do not fit the classical framework under which most estimators are established. A full discussion of the advantages and features of the available estimation methods is, however, far beyond the scope of this paper. The reader may refer to Anselin (1999) for a more formal presentation of some of the available methods for estimating spatial models.

VI. RESULTS

ESDA

Descriptive statistics of the sample are provided in table 1. For the country as a whole, the average rate of self-employment in the workforce is 0.32. The Southeast region exhibits the smallest average (0.23) and the North region the highest one (0.39).

TABLE 1

Descriptive Statistics

Variable	Mean	Sd	Min	Max	Variable	Mean	Sd	Min	Max
Brazil					North				
SE	0.32	0.15	0.0067	0.99	SE	0.39	0.15	0.007	0.81
URB	58.8	23.41	0	100	URB	51.6	20.9	4	100
PC96	2973.4	2433.3	101	20938	PC96	1848	1258	101	12874
HDIM	0.70	0.08	0.47	0.92	HDIM	0.66	0.06	0.48	0.81
SEC	0.18	0.1	0	0.76	SEC	0.14	0.07	0.025	0.55
Northeast					Southeast				
SE	0.37	0.15	0.08	0.97	SE	0.23	0.08	0.08	0.99
URB	49.8	20.6	2	100	URB	70.3	21.1	0	100
PC96	1182	866	138	11532	PC96	4324	2944	323	20938
HDIM	0.61	0.05	0.47	0.86	HDIM	0.74	0.06	0.57	0.92
SEC	0.15	0.07	0.01	0.64	SEC	0.2	0.09	0	0.72
South					Central-W.				
SE	0.37	0.16	0.09	0.85	SE	0.26	0.09	0.1	0.99
URB	54.94	25.27	0	100	URB	69	17.5	20	100
PC96	4078	1806	177	17765	PC96	3364	1844	390	12697
HDIM	0.77	0.043	0.62	0.88	HDIM	0.74	0.04	0.6	0.85
SEC	0.21	0.13	0.02	0.76	SEC	0.16	0.07	0.04	0.56

The indicators of social development, HDIM and URB are above the national averages (of 0.69 and 0.58 respectively) in the South and Southeast regions. These regions also possess higher rates of employment in the secondary sector (21% and 20% respectively), comparing to the overall average (18%). The GDP per capita are much lower in North and Northeast regions.

The low probabilities of the Morans'I statistic point out to a positive spatial autocorrelation for all variables (table 2). These results may indicate the predominance of clusters of high and/or low values for the variables.

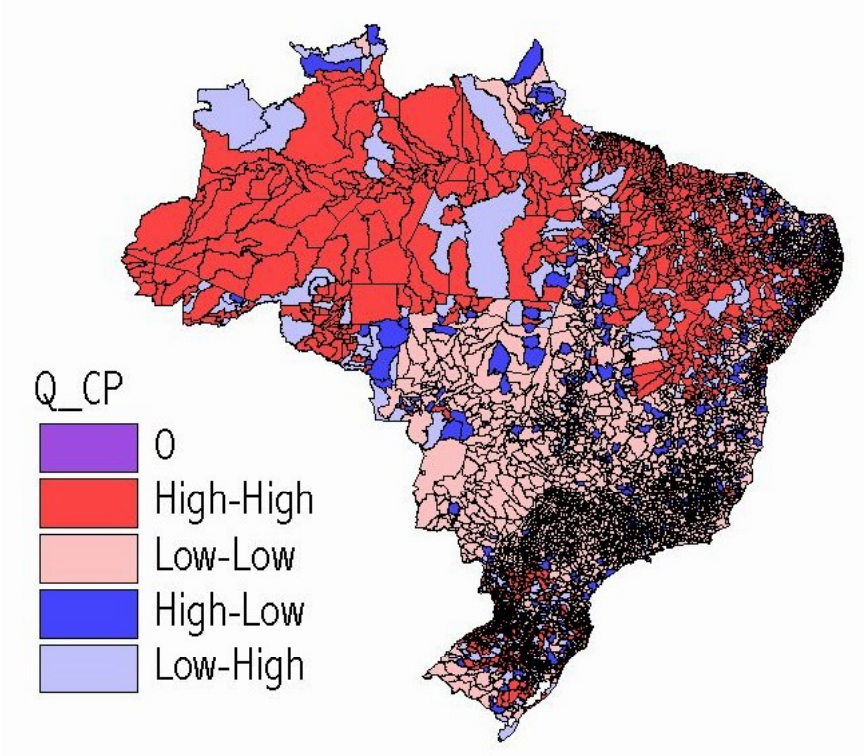
TABLE 2

Moran's I statistics

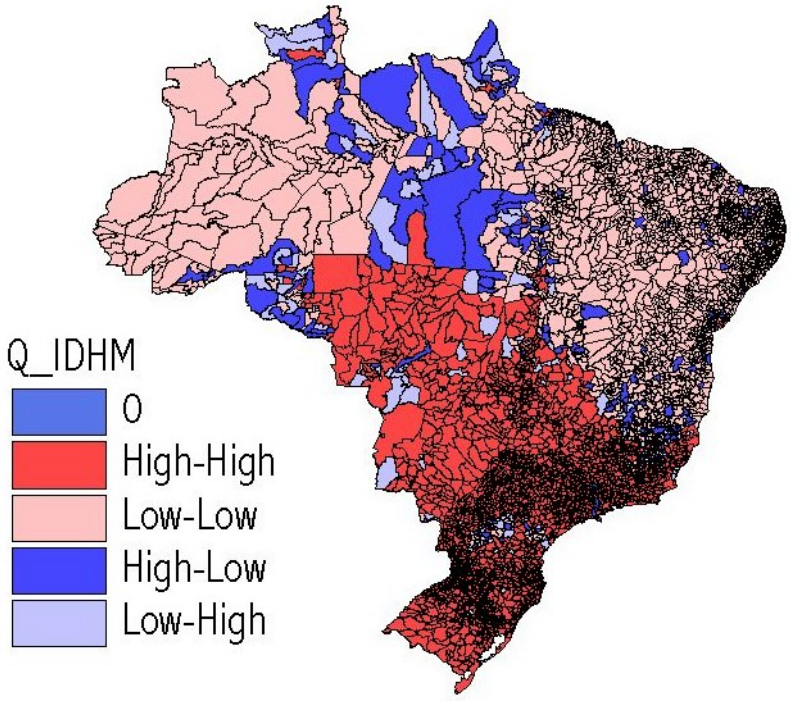
Variable	Morans'I	Prob
SE	0.56	0.000000
HDIM	0.69	0.000000
PC96	0.56	0.000000
SEC	0.53	0.000000
URB	0.43	0.000000
DCONS	0.46	0.000000
DPROG	0.10	0.000000
DPLANOS	0.15	0.000000

Visual inspection of the Moran Scatterplot maps (Figure 1) constructed from the Morans'I statistic, reveals that the municipalities tend to fall in two general categories. The regions South, Southeast and Central-west, are characterized by high rates of urbanization (high-high) and low rates of self-employment (low-low). The opposite behavior is observed in the regions North and Northeast. The latter regions also concentrate municipalities with smaller HDIM and GDP per capita, although some outliers can be observed (high values surrounded by low values of the neighbors), due to economic enclaves. In short, it seems that there's an opposite trend between the indicators of development and the rates of self-employment.

FIGURE 1
Moran Scatterplot maps

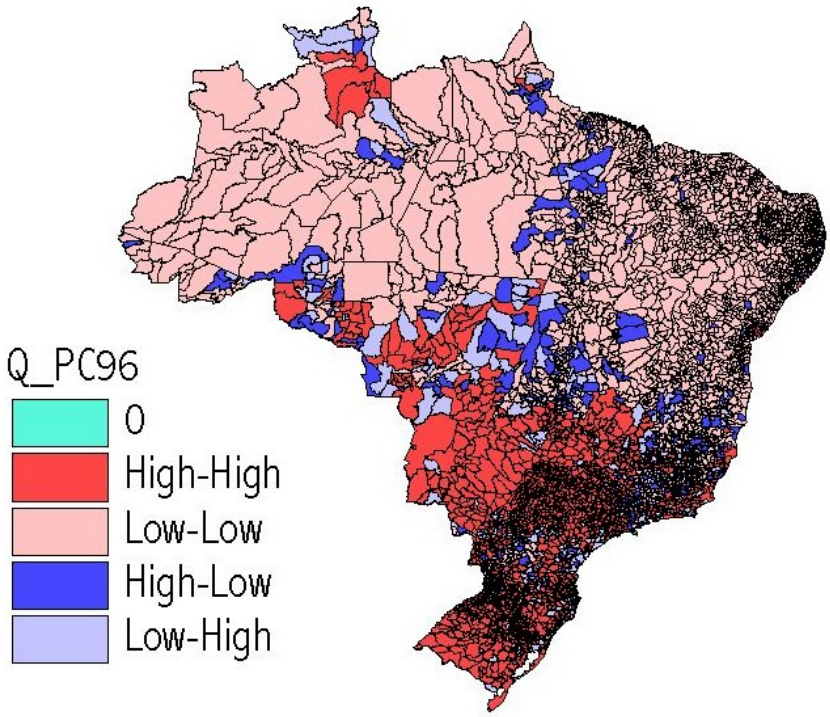


(a) Moran Scatterplot for self-employment - SE

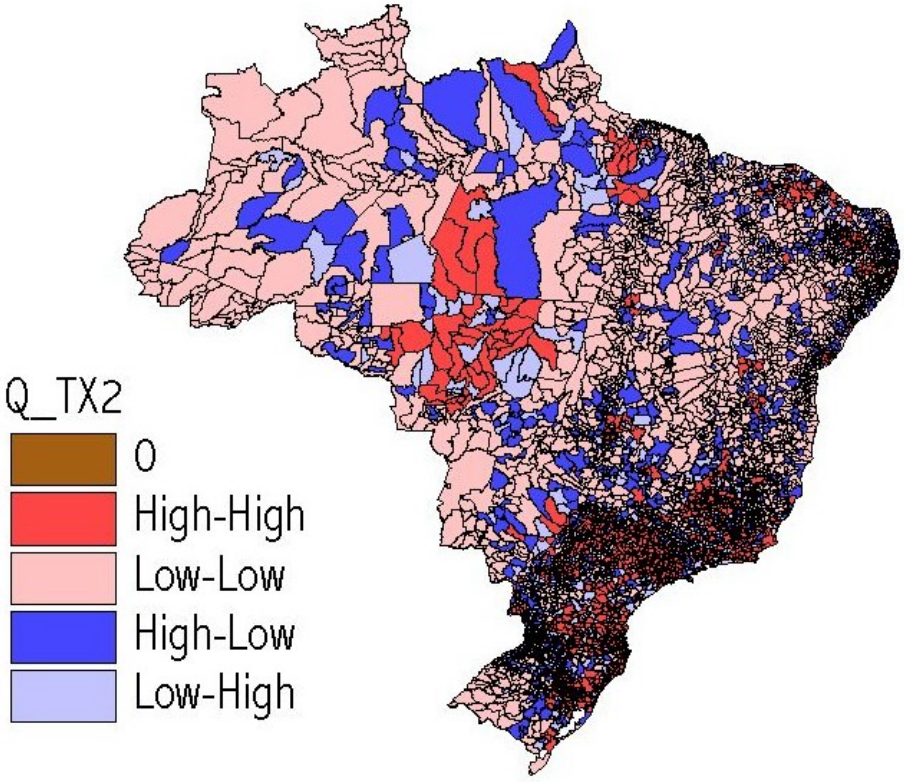


(b) Moran Scatterplot for rate of urbanization - URB.

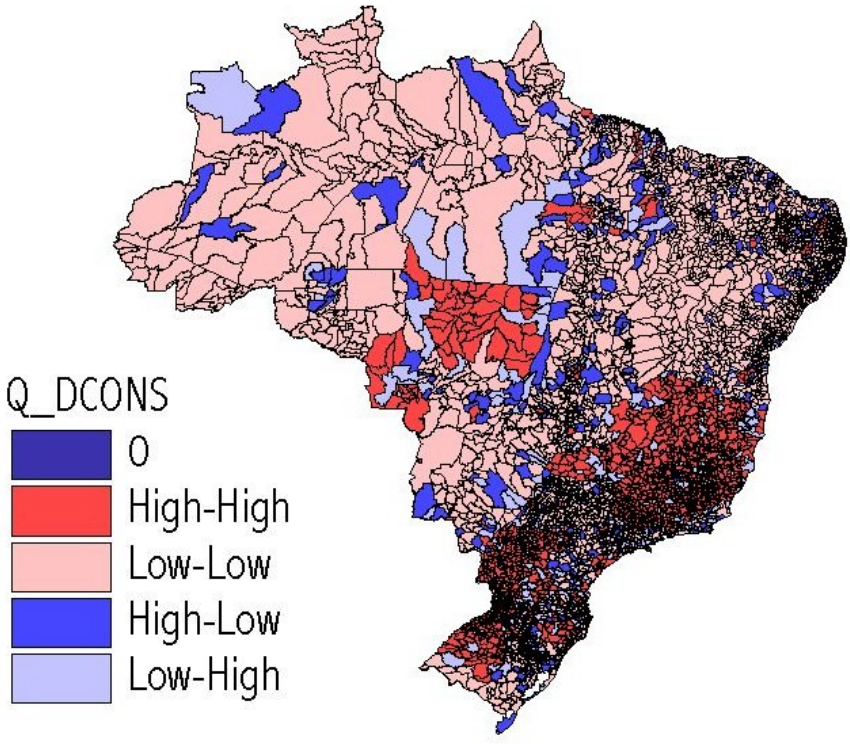
(c) – Moran Scatterplot for Human Development Index - HDI.



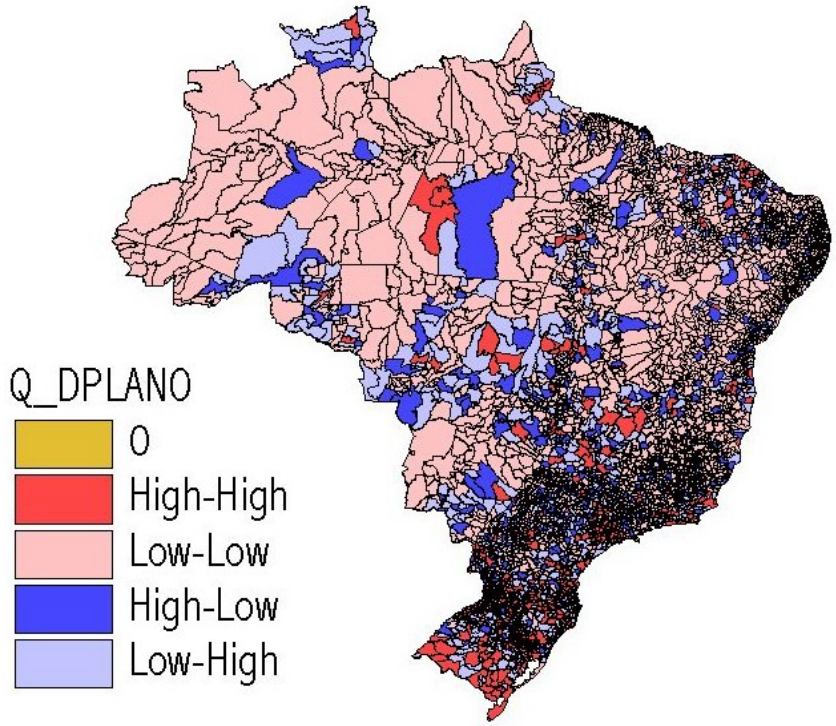
(d) Moran Scatterplot for GDP *per capita* – PC96



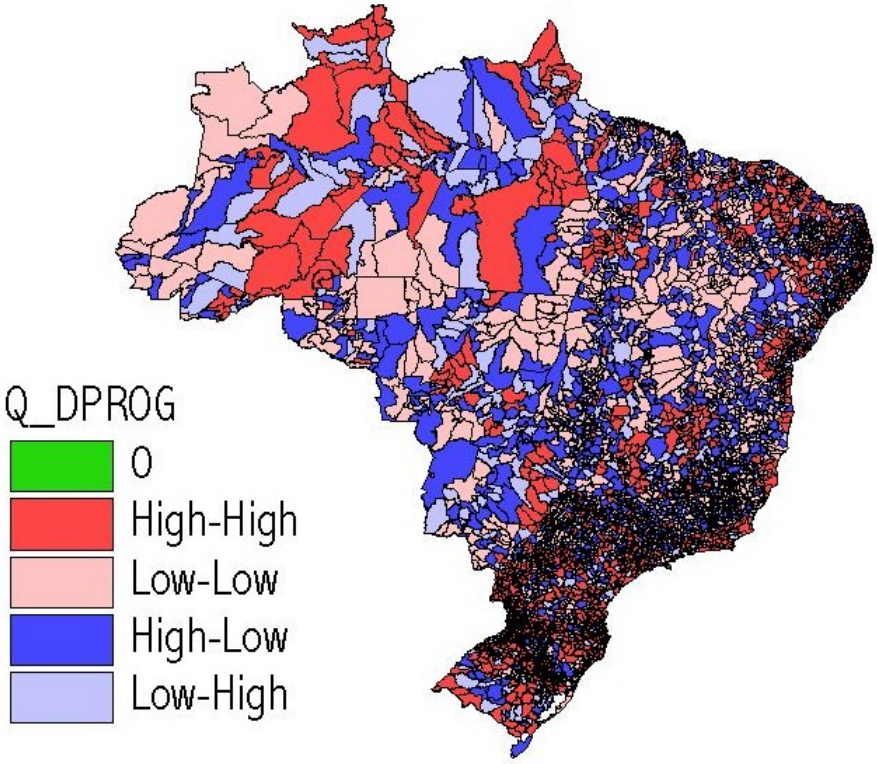
(e) – Moran Scatterplot for rate of occupation on secondary sector - SEC



(f) Moran Scatterplot for DCONS.



(g) Moran Scatterplot for DPLANS.



(h) Moran Scatterplot for DPROG.

Concerning the indicators of urban organization (DPLANS, DCONS e DPROG), the analysis is similar. The municipalities with consortiums are concentrated in Southeast region and some specific parts of the South and Central-west regions. In the remaining regions, only isolated municipalities appear to possess consortiums. The programs (income, professional training, housing, etc) show a uniform spatial distribution.

The results of the ESDA corroborate the regional disparities in Brazil. The South and Southeast regions have better socio-economic indicators whereas the North and Northeast regions show the worse results. The results also confirm the hypothesis that individuals living in less developed places tend to go to the informal sector.

The OLS model

Table 3 presents the results for two "non-spatial" OLS regressions. Model (1) refers to the estimation of the model for the country as a whole. Both heterokedasticity tests, White and Koenker-Basset, are highly significant. As we mentioned above, the presence of heteroskedasticity in a model using spatial data may be a consequence of both spatial heterogeneity and/or spatial autocorrelation in the variables.

In addition, the results of the ESDA demonstrated that substantial geographical variation (heterogeneity) exists within Brazil supporting the division of the country in two large homogeneous "macro regions": one composed by the North and Northeast, the less developed regions, [macro region (a)] and another including the more developed South, Southeast and Central-West regions [macro region (b)]. They stand for model 2, (a) and (b) results.

The test for structural stability of regression coefficients across regions (Spatial Chow test) rejects the null hypothesis, confirming that the pattern of effects across regions is different. Moreover, the tests for individuals' coefficients show that most explanatory variables exhibit significantly different effects in both "macro regions".

The results for the estimations show that the variables related to the level of development, (URB and PC96) are negatively related to the rate of self-employment for the country as a whole and for the less developed regions. These results confirm Noorderhaven's et al. hypothesis of an inverse relationship between prosperity levels and self-employment. However, the variable PC96 is not significant for the more developed regions. This result may be due to the fact that the standard error for this variable is much higher in those regions. The negative relationship between the rate of occupation in the secondary sector (SEC) and the rate of self-employment is consistent in all models and is coherent with the premise that self-employment occurs mainly in primary and tertiary sectors.

TABLE 3

The OLS results
Dependent Variable: rate of self-employment (SE)
Number of observations: 5507

Variables	Model (1) Generic Regression	Model (2) Spatial Regimes	
		(a)	(b)
Constant	0.57 (121.86)	0.56 (76.67)	0.54 (74.52)
URB	-0.003 (-45.47)	-0.002 (-40.78)	-0.003 (-40.79)
PC96	-0.003 (-4.66)	-0.011 (-4.39)	0.0006 (0.77)
SEC	-0.15 (-8.59)	-0.28 (-8.43)	-0.08 (-3.78)
LB	-0.0004 (-0.43)	-0.0008 (-0.26)	-0.0004 (-0.43)
FAV	$1.04*10^{-7}$ (0.37)	$-6.8*10^{-7}$ (-0.16)	$7.54*10^{-8}$ (0.27)
DCONS	-0.019 (-6.04)	-0.007 (-1.14)	-0.009 (-2.15)
DPLANS	0.02 (6.18)	0.008 (1.00)	0.03 (6.63)
DPROG	0.006 (1.90)	-0.011 (-2.23)	0.014 (3.51)
R2-adj	0.40	0.41	
Condition Number	8.63	10.12	
Jarque-Bera	1076***	1149***	
Koencker-Basset	446***	134***	
White	634***		
Chow-Wald		78.89***	

The coefficients of the variables indicating the level of urban organization exhibit changes in magnitude and significance. The coefficient of the variables DPLAN and DCONS are consistent in all models but the strength of the effects is different. The variable DPLANS is positively related to rates of self-employment, although non significant for the less developed regions. The interpretation of these results would be that the more structured the municipalities the higher the rate of self-employment. In this case, we would be contradicting Noorderhaven's et al. hypothesis of a positive relationship between higher levels of prosperity and lower rates of self-employment. However, positive and significant values for this variable are also associated to positive and significant values for the variable DPROG. The positive sign of its coefficient for the more developed regions may suggest that in those regions income and professional training programs are usually related to self-occupation. They usually motivate the entrepreneurship and the establishment of small business, which usually operate in the informality for quite long time. Besides, the housing programs embrace civil construction, a sector where self-occupation or hiring employees by task are usual.

The variable DCONS is negatively related to rates of self-employment indicating that the more organized the municipality the lower the rate of self-occupation. However, this variable is non significant for the less developed regions. This may be explained by the fact that in more developed regions, consortiums represent a viable option for small and medium size municipalities. On the contrary, in less developed regions, only the municipalities that reached some stage of development may be able to organize themselves in consortiums.

The coefficient of the variables FAV and LB were not significant. This may have happened because of the high number of missing observations for these variables. As a final point, we should consider that the participation of self-employment in the economic sectors tends to differ amongst these two large "macro regions". In the more developed areas of the South, Southeast and Central-west, self-employment is more likely to be associated with the tertiary sector. On the other hand, in most municipalities of the North and Northeast regions, a large proportion of self-employment is in primary sector, for instance the small farmers. Furthermore, the way the programs of housing, income and job generation, as well as professional training are implemented amongst the two "macro regions" differs considerably.

Specification tests and Spatial Lag-Model

Even after considering the spatial regimes in the OLS model (2), the Koencker & Basset heterokedasticity test is still significant. This result may have been induced by the presence of spatial autocorrelation in the model. Consequently, to assess the extent of spatial effects we carried out specification tests for spatial lag and spatial error dependence on the OLS model (2). The results are in table 4.

The low probability for the Morans'I statistic points to the presence of positive spatial autocorrelation in the error terms. The LM (error) rejects the null hypothesis of non-spatial autoregressive error terms. Likewise, the Kelejian-Robinson test, which does not rely on non-normality, points to some form of spatial autocorrelation in the error terms.

TABLE 4
Specification tests for OLS model

	Model (1) Generic Regression			Model (2) Spatial Regimes		
TEST	MI/DF	Value	Prob	MI/DF	Value	Prob
Moran's I (error)	0.403	50.34	0.000000	0.199	24.88	0.000000
Lagrange Multiplier (error)	1	25919.41	0.000000	1	613.47	0.000000
Robust LM (error)	1	191.31	0.000000	1	1.81	0.178284
Kelejian-Robinson (error)	7	3280.82	0.000000	14	1032.74	0.000000
Lagrange Multiplier (lag)	1	2673.303	0.000000	1	823.39	0.000000
Robust LM (lag)	1	345.205	0.000000	1	211.73	0.000000
Lagrange Multiplier (SARMA)	2	2.864.62	0.000000	2	825.20	0.000000

The LM (lag) also rejects the null hypothesis and points to the presence of a spatial lag in the model. According to Anselin (2002), in those cases where the specification tests indicates both lag and error, one should guide the decision by the results of the robust tests. The higher value of the Robust LM (lag) points to a spatial lag model. Table 5 presents the results of the estimation of a spatial lag model by Instrumental Variables (IV).

TABLE 5

Results for the spatial lag model

Dependent Variable: rate of self-employment (SE)

Endogenous Variables: SEC

Instruments: W_SEC

Variables	Model (2)	
	(a)	(b)
Constant	0.40 (29.78)	0.39 (36.49)
W_SE	0.43 (17.5)	
URB	-0.002 (-16.65)	-0.003 (-30.43)
PC96	$-1.31*10^{-5}$ (-5.46)	$1.11*10^{-6}$ (1.39)
SEC	-0.40 (-6.59)	-0.16 (-5.79)
DCONS	-0.007 (-1.26)	-0.009 (-2.40)
DPLANS	0.006 (0.82)	0.017 (3.59)
DPROG	-0.009 (-1.98)	0.007 (1.92)
Pseudo-R2	0.487	
Sq.Corr	0.490	
LM (error)	0.9883	

The first point to note is that the estimated spatial autocorrelation coefficient is positive and statistically significant, indicating that a significant part of variation in self-employment is explained by spatial dependence. This result implies that the geographic clustering of self-employment is due to the influence of self-employment in one place on self-employment in another, i.e. the average rate of self-employment of the neighbors. The reason for that can be the standard explanation of factor spatial mobility in regional economies.

In the case of labor force measures, the administrative boundaries do not reflect the underlying process generating the sample data. Workers are mobile and can cross municipality lines to find employment in neighboring municipalities. Therefore, self-employment rates measured by where people live could exhibit spatial dependency, since the mobility cost is low, i.e. low transport and accessibility costs.

Concerning the results for the remaining explanatory variables, the magnitude and signs are generally consistent with those observed in non-spatial analysis. The significance of the coefficients decreased for most variables, sometimes considerably. However, the significance of some coefficients (for instance, URB, DPLAN, SEC) was already so high that the decrease was not strong enough to change the main inferences from the OLS model.

The results for the LM (error) test for the spatial lag model confirm the absence of remaining spatial autocorrelation in the residuals.

VII. CONCLUDING REMARKS

The present study looked for evidences of the relationship between degree of development and rate of self-employment in the Brazilian economy. The point of reference was the article from Noorderhaven et al. where he investigated the differences in the rates of self-employment for more than twenty western countries and Japan, during the period 1974-1994.

We showed that substantial geographical heterogeneity in rates of self-employment exist within Brazil, supporting the division of the country in two large homogeneous "macro regions": one composed the less developed regions of the North and Northeast, and another including the more developed South, Southeast and Central-West regions.

The ESDA as well as the spatial regression analysis points out to the presence of clusters of municipalities with similar rates of participation of self-employment in the economy. The fitting of a spatial lag model to the data demonstrated among others that: (a) significant part of variation in self-employment is explained by spatial dependence, (b) the municipality's degree of urbanization, rate of employment in the secondary sector as well as GDP per capita indeed influence it's rate of self-employment.

This study represented merely a tentative and perhaps pioneering step towards the application of spatial analysis tools to geo-referenced data on employment. It points to the importance of considering spatial patterns when analyzing self-employment and it's determinants in such a heterogeneous country as Brazil. Future investigations should also consider additional explanatory variables as for instance, wage differentials between formal and informal sector, unemployment rates, among others.

REFERENCES

Anselin, L. and A. Bera (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In: A. Ullah and D. E. A. Giles, Eds., *Handbook of Applied Economic Statistics*, pp.237-289. New York: Marcel Dekker.

Anselin, L., A. Bera, R. Florax and M. Yoon (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics* 26, 77-104.

Anselin, Luc (1988a). *Spatial Econometrics: Methods and Models*. Kluwer Academic, Dordrecht.

Anselin, Luc (1988b). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis* 20, 1-17.

Anselin, Luc (1999). *Spatial Econometrics*. School of Social Science, University of Texas, Dallas.

Anselin, Luc. (1992). SpaceStat tutorial – a workbook for using SpaceStat in the analysis of spatial data. University of Illinois.

Barreto, R.G. Uma avaliação da relação entre informalidade e qualidade dos postos de trabalho. *Revista da ABET* (Associação Brasileira de Estudos do Trabalho). São Paulo. Ano II, n.2. July.

Bera, A. and M. J. Yoon (1993). Specification testing with misspecified alternatives. *Econometric Theory* 9, 649-658.

Burridge, P. (1980). On Cliff-Ord test for spatial autocorrelation. *Journal of the Royal Statistical Society* B 42, 107-108.

Cardoso Jr., Celso e Fernandes, Suiane (2000). A informalidade revisitada: evolução nos últimos 20 anos e mais uma hipótese para pesquisa. *Boletim de Mercado de Trabalho, Conjuntura e Análise,* n. 14, October. IPEA, p. 39-49.

Castells, Manuel and Portes, Alejandro (1989). World Underneath: the origins, dynamics, and effects of the Informal Economy. In: Portes, Alejandro; Castells, Manuel and Benton, Lauren A.(Eds.) *The Informal Economy: studies in advanced and less developed countries***.** Baltimore and London: The Johns Hopkins University Press, pp. 11-37.

Greene, William H (2000). *Econometric Analysis***.** 4.ed. New Jersey: Prentice Hall.

Instituto Brasileiro de Geografia e Estatística – IBGE (1999). CD-ROM Perfil de Informações Municipais 1999.

Instituto Brasileiro de Geografia e Estatística – IBGE (2000)**.** *Brazilian Census for 2000*. (Microdata).

Jorge, A. F. and Valadão L. (2002). O setor informal no Brasil: metodologia e principais resultados. *Revista da ABET* (Associação Brasileira de Estudos do Trabalho). São Paulo, Ano II, n.2. July.

Kelejian, H.H. and D. P. Robinson (1998). A suggested test for spatial autocorrelation and/ or heteroskedasticity and corresponding Monte Carlo results. *Regional Science and Urban Economics* 28, 389-417.

Kelejian, H.H. and D. P. Robinson (1999). The influence of spatially correlated heteroskedasticity on tests for spatial correlation. In: L. Anselin and R. Florax (Eds), *Advances in Spatial Econometrics*. Heidelberg: Springer-Verlag.

Machado, Ana F(1993)**.** *Trabalhador por conta própria: heterogeneidade e desigualdade de renda em regiões metropolitanas.* Master Degree Dissertation. Centro de Desenvolvimento e Planejamento Regional, Faculdade de Ciências Econômicas, Universidade Federal de Minas Gerais, Belo Horizonte.

Noorderhaven, Niels G. et al. (1999). Self-employment out of dissatisfaction: an international study. *Tinbergen Institute Discussion Papers*, number 99-089/3. 1999.

Paelinck, J. and L. Klassen (1979). Spatial Econometrics. Saxon House, Farnborough.

Portes, Alejandro; Castells, Manuel; Benton, Lauren A. (1989). Conclusion: the policy implications of informality. In: Portes, Alejandro; Castells, Manuel and Benton, Lauren A. (Eds.) *The Informal Economy: studies in advanced and less developed countries***.** Baltimore and London: The Johns Hopkins University Press, pp. 298 - 312.