Análisis Económico, vol. XXXV, núm. 88, enero-abril de 2020, pp. 183-208, ISSN: 0185-3937, e- ISSN: 2448-6655

Informality in the Metropolitan Zone of the Valley of Mexico (ZMVM): a spatial econometric and quantile error analysis

Informalidad en la Zona Metropolitana del Valle de México (ZMVM): un análisis econométrico espacial y de error cuantil

(Esta versión: 22/mayo/2019; Aceptado: 4/diciembre/2019)

*David Robles-Ortiz**
*Roberto Alejandro Ramírez-Silva***

ABSTRACT

The objective of this research is to analyze the behavior of the main causes of informality and its spatial relationship in the Metropolitan Zone of the Valley of Mexico (ZMVM). To achieve the target, we use the methodology of spatial econometrics, we propose a quantile error from a weight matrix (W). The results found show that neighborhood effects do not explain the informality within the ZMVM moreover, the education variable has a negative impact on informality. The originality of this work is the spatial econometric analysis of the spatial relationship in informality within the ZMVM. One of the main limitations of the present investigation is that it only covers 67% of the total ZMVM localities. The conclusion that emerges from this research is the educational factor is a crucial variable in the reduction of informality in the analyzed area and there is not strong evidence about contagion of informality between municipalities. It is recommended to focus efforts to analyze other metropolitan areas, which present the same problem of informality.
Keywords: Informality; spatial econometrics; matrix of weights (W); and quantile error.
JEL Classification: C10; C21; J40.

* Escuela Superior de Economía del Instituto Politécnico Nacional. Ciudad de México, México. E–mail address: drobleso@ipn.mx
** Escuela Superior de Economía del Instituto Politécnico Nacional. Ciudad de México, México. E–mail address:rramirezs@ipn.mx

RESUMEN

El objetivo de esta investigación es analizar el comportamiento de las principales causas de informalidad y su relación espacial en la Zona Metropolitana del Valle de México (ZMVM). Para lograr el objetivo, utilizamos la metodología de la econometría espacial, proponemos un error cuantil de una matriz de peso (W). Los resultados encontrados muestran que los efectos del vecindario no explican la informalidad dentro del ZMVM, además, la variable educativa tiene un impacto negativo en la informalidad. La originalidad de este trabajo es el análisis econométrico espacial de la relación espacial en la informalidad dentro de la ZMVM. Una de las principales limitaciones de la presente investigación es que solo cubre el 67% del total de las localidades de ZMVM. La conclusión que surge de esta investigación es que el factor educativo es una variable crucial en la reducción de la informalidad en el área analizada y no existe evidencia sólida sobre el contagio de la informalidad entre los municipios. Se recomienda concentrar los esfuerzos para analizar otras áreas metropolitanas, que presentan el mismo problema de informalidad.
Palabras clave: Informalidad; econometría espacial; matriz de ponderaciones (W) y error cuantil.
Clasificación JEL: C10; C21; J40.

INTRODUCTION

Mexico is among the fifteen largest economies worldwide[1], however, about 60% of its Economically Active Population (EAP) belongs to the informality sector. According to the International Labor Organization (ILO) the informal population is characterized by not having social security, see OIT (2014). Informality is a symptom of low productivity related to a low educational level and low economic development that is commonly associated with unemployment, ambulatory, deficiency of technology, and lack of social security for those who are in it – generating, also, a low tax collection.

Likewise, the Metropolitan Zone of the Valley of Mexico (ZMVM) occupies the third cite in population worldwide with more than 20 million inhabitants. The ZMVM is made up of 16 delegations from Mexico City, 59 municipalities from the State of Mexico and one municipality from the state of Hidalgo[2], housing around 18.5% of the total informal population in the country. A metropolitan area[3] is defined as the set of two or more municipalities, where a city of 50 thousand or more

[1] The GDP (current prices) is considered with data from the World Bank.
[2] In 2005, the territorial delimitation of the ZMVM was formed by an inter-institutional group, Secretary of Social Development (Sedesol), National Population Council (CONAPO), and the National Institute of Statistics and Geography (INEGI).
[3] The ZMVM is formed, from a process of territorial expansion, it is considered as part of the metropolitan area to those municipalities adjacent to the city, where the urbanized area of the city penetrates them, see Graizbord (2008).

inhabitants is located and there is a high degree of economic integration in terms of employment and consumption, see INEGI (2010). Currently, the ZMVM presents great challenges in terms of informality. First, there is a concentration of unemployment that motivates economic inequality, which promotes the incorporation of the population to informality. Secondly, differences are generated in the access to services and basic equipment of the population that is in both sectors in addition this generates environmental pollution, road congestion and lack of basic infrastructure that absorbs the demands of the two sectors, see Escamilla (2006).

Although, the phenomenon of informality has been object of study, however, it has not been analyzed in its spatial relationship within the ZMVM. According to Iracheta (2003), the latter is characterized by growing horizontally and being a great land consumer. Therefore, the two hypotheses to prove in this research are, first, question that arises is there a spatial relationship of informality within the ZMVM? and another that emerges from the previous one is, which the variables that impact it are? variables commonly explain the informality in the literature. In this way, the objective of this paper is to analyze the spatial relationship and its impact variables of informality within the ZMVM, where we expect to find spatial relationship factors in the study area. One limitation of this work is that it does not consider all localities (municipalities) of the ZMVM, only 67%, since the available data are extracted from official sample surveys.

The city has allowed to improve the living conditions of the population (in a certain way), given a greater complementarity between the agglomeration of the individuals that inhabit it and the economic activities that take place within it. Which has allowed informality to play a role in the social construction of delimited spaces and with characteristics, such as the "tianguis" –walking markets– or informal commerce areas, for example "Tepito", see Bonilla (2015). In other words, the fragmentation that occurs in the Mexican labor market shows the participation of factors that generate it and that interrelate among themselves where individuals can move from a formal labor market, with rigidities, to an informal one.

Within some factors that explain the high degree of informality in the ZMVM, we can mention the high level of deindustrialization that took place in the late seventies in the area, mostly within the capital of the country, thus dismantling the productive chains and causing a tertiary specialization[4] (formal and informal), see Pradilla (2005). Subsequently, the economic-financial crisis associated with the "Tequila" effect of 1994 brought with it the development of informality in the country's capital, according to the Organization for Economic Cooperation and Development (OCDE, 1995), unemployment increased by 8% in that area. In

[4] Pradilla (2005) points out that in the period 1980-2001, the tertiary sector increased its share of GDP from 66.01% to 76 per cent. The financial, insurance and real estate sectors stand out –from 8.38% to 21.1% in the same period.

addition, about 70% of the occupations that were formed between 1991 and 1997 remained non-salaried or through micro-businesses, concentrated in retail trade, a volume like that of industrial workers who lost their jobs due to the 1994 crisis, see Pacheco (2004).

Likewise, Sassen (1991) shows that the problem of informality is accentuated in the cities, because they concentrate a high-income sector with consumption patterns and sophisticated levels of needs, which increases the supply of low-wage jobs. It should be noted that in cities, especially in an urbanization process, inequalities are reinforced provoking informality (since one market requires the other).

For the above, this research analyzes the variables commonly cited in the literature as causes of informality and observe their impact under the approach of a spatial econometric in the ZMVM, with the use of micro data and official censuses. The econometric estimate gives two important results. The first result is the relevance of education among the population, that is, as the population's education increases, the likelihood of being informal decreases in the ZMVM. The second result is that there is no contagion between municipalities that make up the ZMVM in informality terms. This is important, because the existence of municipalities with high levels of informality, motivates other residents from other municipality to join informality as well.

This paper is organized as follows. The first part corresponds to explain the two principal approaches that explains the informality, for the purpose of identifying the variables to be used in the econometric model. The second section shows and characterize the phenomenon of informality in the ZMVM according the variables commonly used in literature about this topic as gender, education and age of the people that integrates this sector. The third explains in detail how the variables used in the present work were constructed and its description. In the fourth section the econometric methodology is exposed, as well as the weight matrices W1 and W2, and the econometric results obtained for discussion are shown. Finally, the conclusions and recommendations of the research are presented.

I. THE TWO MAIN APPROACHES: ESCAPE AND EXCLUSION

This section presents the two main approaches to informality analysis. The "exclusion" approach has been the most influential in explaining the phenomenon of informality. Castells (1989) as well as Harris and Todaro (1970) point out as a priority cause the rigidities that are generated in the labor market –aspects that range from the lack of technology, low educational level, among other factors– motivating to its fragmentation and creating a dual labor market. In a study focused on the new theories of the labor market carried out by Perrot (1995), the author indicates that

social polarization and dualization are observed in the labor market. This phenomenon is manifested in large cities with an imposing globalization, which materializes in a dual social structure, that is motivated at the same time by the changes in the occupational structure in the large metropolitan areas. Castell (1989) refers to the phenomenon calling it "dual cities."

On the other hand, under the "escape" approach, Perry (2007) argues that the limitations generated by the dual economy do not exist, rather there are competitive labor markets, where workers find similar conditions in both sectors (formal and informal), and choose their optimum level of acceptance with respect to the institutions and mandates of the State. In other words, there is a free transition between the informal and formal sectors, and vice versa –where the individual apparently decides to be informal by his own decision, and not by the barriers imposed by the labor market. However, it is important to mention that the problem of informality is a multi-causal issue[5], so studying it under one focus limits its understanding and therefore its solution.

II. Characterization of the informality in the ZMVM

This section characterizes the phenomenon of informality in the ZMVM, it analyzes those causes that commonly expose the literature as determinants of informality and shows how they behave in the ZMVM, variables to be used for the spatial econometric model as gender, education, age and marital status.

According to official data, the State of Mexico brings together around 14% of the total informal population in the country, and Mexico City only concentrates 6% of the national informal population. When estimating the size of informality in the ZMVM –with data from the Socioeconomic Conditions Module (MCS, for its acronym in Spanish)[6] from the National Household Income and Expenditure Survey (ENIGH, for its acronym in Spanish, 2015) – similar numbers are observed. Thus, of the 125 municipalities that conform the State of Mexico, 59 of them are part of the ZMVM representing 18.5% of the country's informal population. Of those entities that make up the ZMVM –the municipality of Ecatepec de Morelos has the largest number of informal population (9%), followed by the mayor of Iztapalapa and the municipality of Nezahualcóyotl, both entities with 8% and 5% respectively. Of all the informal population in the ZMVM the mayoralties of Mexico City host 34% of said population, the remaining municipalities account for 66%.

[5] Even, like Perry, op. cit., p. 21, will come to comment: "The term informality has different meanings for different people, but almost always has negative connotations: unprotected workers, excessive regulation, low productivity, unfair competition, evasion of the law, low payments or no tax payments and work "underground" or in the shade."

[6] For more knowledge of the structure of the MCS, see INEGI (2015).

About gender in informality, men represent 62% of the informal population in the ZMVM, while women account for only 38%. Only in the capital of the country 858, 230 women are estimated in informal activities against 127,474 men in the same sector. It is estimated that the municipality of Ecatepec de Morelos, one of the largest of the ZMVM, houses the largest population of informal men and women with 184,363 and 493,151 respectively, while the mayor of Iztapalapa hosts the largest informal female population.

However, when estimating the population with the MCS 2014 it is noted that the number of women in informality is higher than estimated with MCS 2015, around 74% of the non-economically active population is in informality in the central region of the country[7], which is indicative that we are facing a problem that is constantly changing. Various studies address the participation of women in informality, Cunningham (2001) assumes that the female gender prefers the informal sector for flexibility at the time, this allows them to take care of their children or perform domestic tasks, which is likely to increase if they have children of young ages, (Gong et al., 2004). Our estimates support the above, for the informal population of the ZMVM, predominate those who identify themselves as married, 63% are men, while the remaining 37% are women, where the municipality of Ecatepec de Morelos concentrates the largest informal population that argues to be married.

On the other hand, informality has been studied mainly under the exclusion approach, Harris and Todaro (1970), Pinto (1970), as well as Tokman (1987), assume that this phenomenon is caused by the rigidities prevailing in the formal labor market. One of them is the educational level of the population. In this area, it is estimated that most of the informal population of the ZMVM, around 1 860 465 individuals, have a level of secondary education, followed by individuals with preparatory around 1 503 351 persons. The inverse relationship between educational level and informality is remarkable, which is consistent with the assumption of a higher level of education less informality, important element to combating it, (Brandt, 2011). However, education alone is not decisive. Our estimates show that even at higher levels of preparation, such as post grade (master's or doctoral) are also located in this sector, although small is the population with this characteristic it is important to take it into account, since around 62,237 individuals, about 1 % of the population of the ZMVM report having informal employment. Even as explained by Schneider et al. (2010), informality is not unique in developing countries where the educational level of the population is higher, even though in Latin America it is 34.7 %, In the countries with the highest OECD income informality is present with 13.5 %.

[7] According to Bank of Mexico the central region is composed of the states of Mexico City, State of Mexico, Guanajuato, Hidalgo, Morelos, Puebla, Queretaro and Tlaxcala.

Another characteristic that commonly explains the phenomenon of informality is the age of the individuals who comprise it, in Mexico City and its Metropolitan Zone the average age of individuals who are part of informality is 39 years of age, while at the national level it is 32 years. Various studies, such as Flores and Valero (2012), argue that, at an early age, they are more likely to be part of informality, due to the little or no experience required for a certain type of employment, or the low entry requirements for the sector concerned.

III. DATA SOURCE AND DESCRIPTION OF VARIABLES

It is important to mention the source of data, the reason lies in the very nature of the topic, as informality is a controversial topic that is open to different perspectives, since it has led to have different definitions over time according to the various study approaches that have been presented. Household surveys are a rich source of data about economic behavior and its links to public policy, they not only provide information about households, but also at the level of the individuals that compose them, since they show their distribution among family members, see Deaton (1997).

For the present work we obtained data from the Socioeconomic Conditions Module (MCS, for its acronym in Spanish)[8]. A project carried out in Mexico by the National Institute of Statistics and Geography (INEGI, for its acronym in Spanish) and the National Council for the Evaluation of Social Development Policy (CONEVAL, for its acronym in Spanish) for the measurement of poverty in Mexico. The MCS takes the structure of the ENIGH, its difference lies in the representativeness of each. The MCS is not only representative at the national level, but also the survey offers information by entity –the ENIGH, which only has national representation– and is held every two years in order to provide useful information on healthy, income, education, and social security, as well as the economic activity of the members of the household, among other characteristics. The MCS is usually biennial, except for the MCS 2015[9] that was developed as an effort to avoid underestimating the income by the population[10].

Despite the criticisms made about the MCS 2015, as far as the income measurement is concerned, it is important to mention the reasons for which it was occupied for the present research. The first reason is to know the income behavior of the informal population with the new measurement proposed by INEGI, which allows us to have a different appreciation of the variable. Another reason is when pretending to study informality in the ZMVM, implies having information at the

[8] For more knowledge of the structure of the MCS, see INEGI (2015).
[9] The total sample of the MCS is composed of 64 thousand 093 homes, of which 15 thousand 133 were in locations of 2 thousand 500 inhabitants and more.
[10] Since, respondents tend to declare less than they receive as income.

state level and urban delimitation (with the MCS it is possible to obtain it). Another reason is the availability of information about the characteristics of the service granted by the social security system to which the ZMVM workers are affiliated. Finally, the International Labor Organization (OIT, 2013) points out that the methods for measuring informality are indifferent, the difference lies in identifying in a better way the method that meets the researcher's requirements.

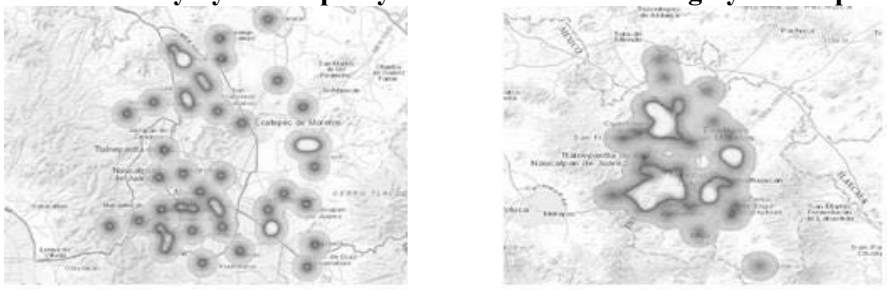
It is important to mention that 25 municipalities were not considered in the present work, since they were not included in the sampling of the MCS 2015[11], which represent around 30% of the sample. In the same way, the database has been homologated with that reported by the official authorities in the matter in order to have a reliable data, and thus avoid distortions in the results obtained. As it was discussed in the previous section, informality in the ZMVM is defined as the characteristic of that individual with an age of 15 years or more and who lacks social security. In other words, any individual that does not have a social security service with the age characteristics indicated above, at the time of the survey is considered as informal, although is affiliated with "Seguro Popular" and resides in the ZMVM. In Appendix A, it is possible to consult the criteria for the construction of the variables used in the present work.

As regards the educational level of the informal population was estimated with the maximum year approved in the school (by the individual) within the Mexican National Educational System. If for example our estimates show that a worker has 11 years of schooling, we would be in the understanding that they are two years of preschool, plus six years of primary education and three years of secondary education completed, therefore, the individual will have 11 years of education. Figure 1 show the concentration of informality and the educational level in the ZMVM.

[11] The INEGI indicates that the MCS being a probabilistic survey, which implies a sampling design, the observation units (in this case, the households) are selected at random that is, the sample does not fall in all municipalities. In addition, the number of homes selected was 2000 for the State of Mexico, in such a way that the sample is not large enough to be distributed in all the municipalities of the ZMVM.

Figure 1

Informality by municipality **Years of schooling by municipality**



Source: Own elaboration with ArcGIS software.

In the case of income[12], a variety of forms was considered, in which the individual can obtain income, such as paid work by own business, via transfers, among other sources. Likewise, in case of having benefits of bonus ("aguinaldo") and profit sharing[13], they are also considered as part of the income. The above is understood under the definition proposed by the ILO, given that the case of existing individuals working for formal companies may be presented where the workers can receive some extra income for the reason of the bonus and at the same time, they do not receive the social security benefit by the company.

Table 1 shows the variables used in the econometric model exposed later, in section IV– to explain the phenomenon of informality (Inf_i) in the ZMVM.

[12] It is important to mention that in the present work, the keys referring to annual income from stock returns and the keys contained in the section on financial and capital perceptions are not considered for the estimation of income. An individual cannot obtain income for these concepts and be informal, since this requires being in formal channels of the economy.

[13] The corresponding amounts for bonus "Aguinaldo" and distribution of profits (for 2014) are deflated at 2015 prices.

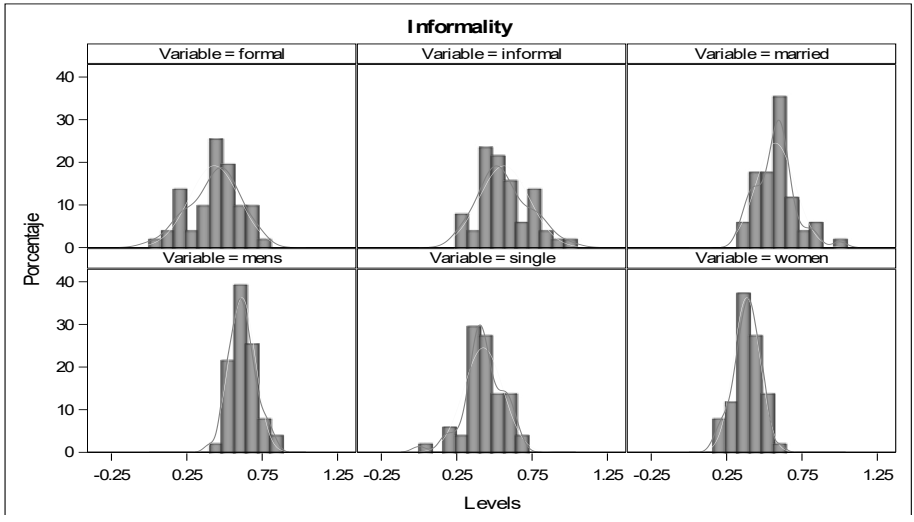
Table 1
Definition of variables

Variable	Definition
male	Percentage of average male population by municipality.
female	Percentage of average female population per municipality.
age	Average age of the population by municipality.
married	Percentage of average married population by municipality.
single	Percentage of average single population per municipality.
income	Average monthly income by municipality (at 2015 prices).
ays	Average years of schooling per municipality.
pob_2015	Estimated population by municipality (to 2015).
s_skm	Surface in square kilometers by municipality.
audm	Average urban density by municipality.
pgrm	Population growth rate by municipality.

Source: National Institute of Statistics and Geography (INEGI). Population Census -INEGI, 2010. National Household Income and Expenditure Survey (ENIGH).

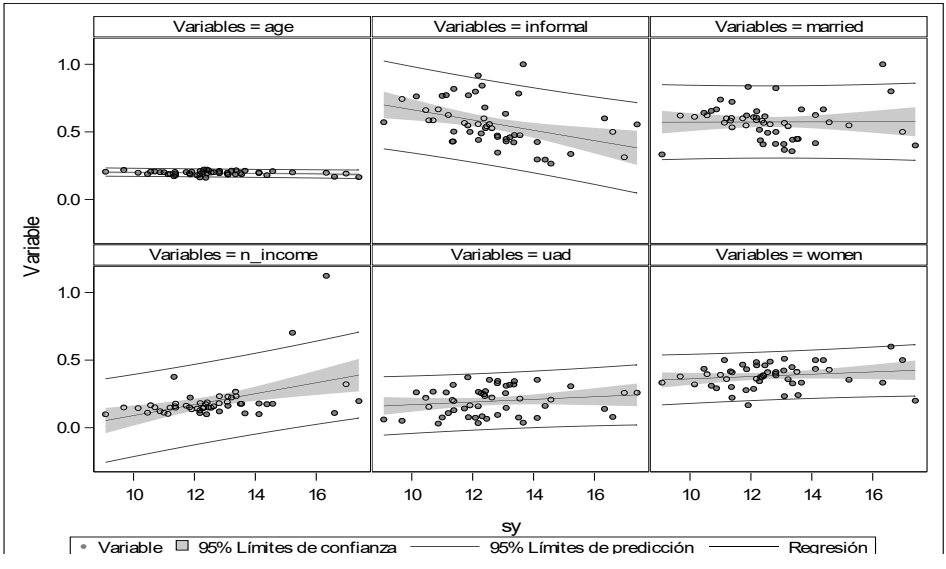
Figure 2 shows the histograms of the variables formal, informal, married, single, man and woman. It can be seen that for the case of the variables marital status and gender (male or female), the characteristic of being at the same time male and informal is more frequent within the ZMVM since they represent being heads of family and for that reason it is more common that they are married or under some relationship of couple and belong to the spectrum of informality. Figure 3 shows a panel of scatter-type graphics where it can be seen, clearly, that the variable years of study has a negative relationship with respect to the levels of informality within the ZMVM.

Figure 2
Histogram (variables)



Source: Own elaboration with SAS Software.

Figure 3
Scatter. years of schooling versus variables



Source: Own elaboration with SAS Software

IV. METHODOLOGY AND RESULT

Within the broad field of econometrics, the sub-branch that studies spatial interrelations (spatial heterogeneity and spatial dependence) is known as spatial econometrics, the last one is used to understand how one region influences another neighboring region. One of the pioneering works in the field of space econometrics was the work developed by Anselin (1988b). The author analyzes the number of homicides and their spatial relationship using as explanatory variables as information on the average income and average value of the houses in 49 counties of Columbus in the United States of America. It is important to point out that the conventional analysis of traditional econometrics[14] has focused mostly on the role of time as a key dimension of study without an adequate assessment of the spatial factor, see Pérez (2006).

Econometric Methodology: spatial and quantile

As regards the study of spatial econometrics, there are two models that are relevant in this area, the spatial autoregressive model (SAM) and the spatial autoregressive error model (SEM)[15]. The regional study that makes spatial econometrics, the variable y_i is defined with the sub-index (i) which refers to a locality belonging to a region D, either a continuous surface or a finite set of discrete points, in other words$\{y_i, i \in D; i = 1 \dots n\}$, see Baltagi (2003). For our case, D represents the ZMVM in the present work we will focus on spatial cross-sectional models, which are defined in the following equation:

$$y_i = x_i'\beta + \varepsilon_i \tag{1}$$

As soon as there is a spatial interaction of the phenomenon of study the spatial covariance conditions for$y_i, y_j \in D$ are defined below in equation (2).

$$cov[y_i, y_j] = E[y_i y_j] - E[y_i] \cdot E[y_j] \neq 0 \tag{2}$$

The mixed autoregressive-regressive model
The mixed autoregressive-regressive model is defined as follows:

[14] Traditional econometrics is very broad, just to mention some models: time series, GARCH volatility models, VAR models and Cointegration of variables, for a detailed description, see Lütkepohl (2006). In the models of space-time (data panel), a more detailed explanation can be found in Baltagi (2003).
[15] However, these models (the spatial autoregressive model and the spatial autoregressive error model) can be extrapolated to data panel models, fixed effects, using the two previously mentioned models and even qualitative data models Probit type, see Baltagi (2003).

$$y = \rho W y + X\beta + \varepsilon \qquad (3)$$
$$\varepsilon \sim N(0, \sigma^2 I_n)$$

The above equation can be represented in an alternating way as$y = (I - \rho W)^{-1} + (I - \rho W)^{-1}\varepsilon$ where the dependent variable (y) is a vector of order nx1. On the other hand, W is a weight matrix that acts as a space lag operator (nxn) and X is a matrix (nxk) where n = number of localities under study and k = number of explanatory variables in the model.

The spatial autoregressive error model (SEM)

According to Anselin (1988a) the spatial autoregression error model (SEM) which develops in two stages arises from a scarce spatial relationship of the phenomenon of study. In a first stage of this model we obtain the error term (ε), which is exposed in the following equation:

$$y = X\beta + \varepsilon \qquad (4)$$

Also, the equation (4) can be represented in the following way$y = X\beta + (I - \lambda W)^{-1}\mu$. Subsequently, in a second stage the parameter $\hat{\lambda}$ is obtained –based on a regression defined below:

$$\varepsilon_i = \lambda W \varepsilon_j + \mu \qquad (5)$$

The error term of equation (5) is expected to have the following distribution characteristics $\mu \sim N\left(0, \sigma_\mu^2 I_n\right)$. On the other hand, the obtaining of the parameters of the previous models is carried out by the maximum verisimilitude method, under a normal distribution in spatial errors, exposed in equation (6), using optimization methods such as BHHH, see Berndt *et al*., (1974). It can also possible to apply another kind of optimization method, see Kelley (1999).

$$InL = -\left(\frac{n}{2}\right)\ln(2\pi) - \left(\frac{n}{2}\right) ln\sigma^2 + ln|I - \rho W y| - \left(\frac{1}{2\sigma^2}\right)(y - \rho W y - X\beta)'(y - \rho W y\ - X\beta) \qquad (6)$$

In addition, in this paper we present the results of the local Moran Index, see Moran (1950)[16], this Index allows us to detect if there is any local spatial correlation pattern that is, for purposes of the present research, it allows us to evaluate whether there are high levels of concentration or presence of some informality cluster based on the rejection of the null hypothesis, this statistic tends towards a normal distribution.

Quantile residua

In order to detect residual autocorrelation factors not considered by the spatial auto-regression error model or by a regression of ordinary least squares. Next, the methodology of a quantile regression[17] represented in its generalized linear form in equation (7) is represented, as well as, an analysis with quantile errors weighted with the lag spatial matrices, later exposed.

$$y = X\beta_\tau + \varepsilon \tag{7}$$

The coefficients (β_τ; $0 < \tau < 1$) are obtained from equation (8), by means of a verisimilitude function that is maximized given a specific τ, using the methodology of Berndt *et al*., (1974). A more detailed study of other methods of maximizing a function of verisimilitude can be found in Kelley (1999). Although the expression is in terms of minimizing distances, however, optimization can be seen in maximization terms, by changing the sign of the expression.

$$\hat{q}_\tau = \underset{q \in R}{\operatorname{argmin}} \left[(\tau - 1) \sum_{y_i < q} (y_i - q) + \tau \sum_{y_i \geq q} (y_i - q) \right] \tag{8}$$

In this way, the conditional quantile regression is given by $Q_{Y|X}(\tau) = X\beta_\tau$ it can be obtained for the parameters β_τ and under a particular distribution (Y), minimizing the distance between the observation i and its respective quantile, solving for the equation (9).

$$\hat{q}_\tau = \underset{q \in R}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau (y_i - q) \tag{9}$$

[16] The Moran Index is defined as $I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{y' W y}{y' y}$ where w_{ij} is the element (i, j) of the weight matrix (W) and the variable (y) represents the dependent variable of the study.

[17] The quantile model was exposed for the first time by Koenker and Bassett (1978).

Expressed differently from the previous equation, in terms of a vector of coefficients, we have:

$$\beta_\tau = \underset{\beta \in R^k}{\operatorname{argmin}} E\big(\rho_\tau(Y - X\beta)\big) \tag{10}$$

For the present work, we propose an analysis of spatial errors obtained from a quantile regression for τ = 0.50, previously exposed in equation (8) since for the case of 0.25 and 0.75, the coefficients obtained are not significant. In such a way that allows us to analyze some patterns of spatial autocorrelation of the phenomenon of informality in the municipalities and delegations of the ZMVM, from the following equation:

$$\varepsilon_i = \delta \sum_j w_i \varepsilon_j + u_i \tag{11}$$

Proposal of weight matrices W1 and W2

For the present research, two weight matrices are used representing the space lag operators. The generalized definition of a weight matrix (W) is it expressed in equation (12). It is important to note that it is not possible to consider auto neighborhoods, so the elements of the main diagonal of each matrix is equal to zero that is, only neighborhoods are considered for i ≠ j.

$$W = \begin{bmatrix} 0 & w_{1,2} & \cdots & w_{1,n} \\ w_{2,1} & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & w_{n-1,n} \\ w_{n,1} & \cdots & w_{n,n-1} & 0 \end{bmatrix} \tag{12}$$

The weight matrix (W1) represents a neighborhood (queen type) in the region between municipalities and delegations of the ZMVM, considering less than 25 kilometers represented as (d)[18]. The matrix W1 is defined from an operator function where there is a neighborhood if the distance between a municipality (i) and

[18] A 25 kilometer Benchmark distance (d) was taken for the present work, which considers factors of highway sinuosity, this distance is an approximation of the center of Mexico City to the municipality of Ecatepec de Morelos (in the State of Mexico), since both locations have a high flow of people every day, within the ZMVM.

another municipality (j) is less than d and zero otherwise, the above is expressed in the following equation:

$$W1 = \begin{cases} I\left(d_{ij} < d\right) \\ 0 \end{cases} \tag{13}$$

On the other hand, we propose another type of neighborhood very particular, represented by the matrix W2, defined in equation (14) which not only consider factors of neighborhood distance, but also educational factors[19]. In such a way that allows us to explore how educational factors affect the informality phenomenon among the neighboring municipalities of the ZMVM, when the years of schooling of a municipality (j) are lower than the average (12.6 years) of the ZMVM.

$$W2 = \begin{cases} \dfrac{I\left(\left(d_{ij} < d\right) + \left(ae_j < \overline{ays_{ZMVM}}\right)\right)}{\sum_{j=1}^{n} I\left(\left(d_{ij} < d\right) + \left(ae_j < \overline{ays_{ZMVM}}\right)\right)} \\ 0 \end{cases} \tag{14}$$

For the calculation of the distances (d_{ij}) between each locality considered in the present work, the Haversine trigonometric formula is used. The latter calculates the shortest distance between two points on the surface of the earth and is defined as follows:

$$d_{ij} = 2 \cdot r \cdot arcsen\left(\sqrt{sen^2\left(\frac{\varphi_j - \varphi_i}{2}\right) + \cos(\varphi_i)\cos\left(\varphi_j\right) sen^2\left(\frac{\lambda_j - \lambda_i}{2}\right)}\right) \tag{15}$$

Here r represents the average radius (6371 km) of the earth, φ represents latitude and λ length, for the municipalities (i) and (j) of the ZMVM respectively.

[19] The variable years of study was considered (the obtained econometric results are exposed in subsection 4.3), since this variable turned out to be explanatory in the four econometric specifications, see Table II. The variable AUD (average urban density) also proved to be explanatory. However, the distance factors are already considered in the same matrices W1 and W2, respectively.

Econometric results

In this sub-section the results obtained are presented from the econometric model proposed for the present investigation in its generalized linear form exposed in equation (16). Where the dependent variable Inf_i represents the percentage of the informal population in each municipality or delegation of the ZMVM that is a function of the independent variables, previously defined in Table 1.

$$\begin{aligned}\text{Inf}_i = \beta_0 + \beta_1 \text{male}_i + \beta_2 \text{female}_i + \beta_3 \text{age}_i + \beta_3 \text{married}_i \\ +\beta_4 \text{single}_i + \beta_5 \text{income}_i + \beta_6 ays_i + \beta_7 \text{pob}_{2015_i} + \beta_8 \text{s}_{\text{skm}_i} \\ +\beta_9 \text{audm}_i + \beta_{10} \text{pgrm}_i + \text{u}_\text{i} \end{aligned} \quad (16)$$

Table 2 shows the results obtained by Ordinary Least Squares[20] later, those obtained under the methodology of spatial econometrics for the two main models, the spatial autoregressive model (SAM) and the spatial auto-regression error model (SEM), and finally a quantile regression for a $\tau = 0.50$. Regarding the regression of Ordinary Least Squares, the variables that only explain informality are the years of schooling and the average urban density. The impact of the first variable is negative, this result shows that more years of study in the population in each municipality informality is presented at lower levels. Similarly, another variable that negatively impact informality is the average urban density,[21] which can be interpreted as the higher density in the municipalities of the ZMVM, the lower percentage of informal population is observed.

However, the last result seems to be not very clear, since it was found that the highest levels of informality are not those presented in Ecatepec de Morelos or the mayors of Iztapalapa and Gustavo A. Madero (demarcations with high population levels). Where it would be expected that these demarcations were promoters of the phenomenon and presented in the first place in percentage of informality. The foregoing is, of course, in terms of percentage of population since being entities with a high number of populations, also its composition is diverse. In

[20] These coefficients are obtained by the following closed formula $\hat{\beta} = (X'X)^{-1}X'y$ where X represents a n,k order matrix of explanatory variables, including the intercept, and (y) is a n,1 order vector that represents the dependent variable. The above formula is obtained from a linearization of the errors under a normal distribution.

[21] The average urban density is defined as $\text{aud} = \frac{\sum_{i=1}^{n} p_i^2}{\sum_{i=1}^{n} p_i}$ where p_i is represented as the population of the urban AGEB i and s_i the surface of the urban (i) AGEB, respectively. The Basic Geostatistical Areas (AGEB) are units that limit a part or the whole of an urban locality, in sets of blocks that generally go from one to 50. These are perfectly delimited by streets, avenues, walkers or any other easy feature to identify in the land and whose land use is housing, industrial, commercial, and services.
Source: Glossary of the Environment and Natural Resources Secretariat (SEMARNAT).

other words, the population that make up these demarcations has the characteristic of being informal, but also another part that is not, so the result must be interpreted with some caution. The above mentioned is exposed in the first column of Table 2. On the other hand, based on the spatial autoregressive model (SAM), the results obtained are very similar to those found by ordinary least squares, see the second column of Table 2. Again, the variable of years of schooling impacts negative to informality likewise, the variable dmu has a negative impact, as well as the rate of population growth for each municipality of the ZMVM. This last result is directly related to the dmu, so it is understandable its negative impact on informality.

As regards the results obtained with the spatial autoregression error model (SEM), the results are like those of the spatial autoregressive model (SAM), see the third column of Table 2. However, the variables that make the difference are the negative impact on age and in the gender of women which shows that informality is centered on an average age of 38.7 years and in a third quartile of 40.5 years. Regarding the gender woman it can be understood that the impact on the last result is negative, because of the role played by the woman as a housewife.

Table 2
Econometric results

	Ordinary Least Squares (OLS)	**Spatial autoregressive model (SAM)**	**Autoregressive error model (SEM)**	**Quantile Regression (τ=0.50)**
Variables	**Coefficients Standard error**	**Coefficients Standard error**	**Coefficients Standard error**	**Coefficients Standard error**
Constant	9.864 (201.64)	71.754 (97.64)	76.95 (100.07)	-8.32 (109.43)
male	70.0 (162.26)	0.500 (0.000)	1E-05 (0.000)	200*** (62.38)
female	71.55 (162.27)	-0.287 (0.195)	-0.319* (0.189)	211.75*** (62.47)
age	-0.008 (0.008)	-0.010 (0.006)	-0.011* (0.006)	-0.025*** (0.007)
married	-80.19 (122.53)	-70.557 (97.667)	-75.321 (100.11)	-201.14** (85.06)
single	-80.02 (122.49)	-70.370 (97.633)	-75.11 (100.07)	-200.95** (85.03)
income	7.55e-07 (1.87e-06)	1.12E-06 1.47E-06	1.10E-06 (1.36E-06)	0.000*** (0.000)
ays	-0.036** (0.016)	-0.038*** (0.012)	-0.034*** (0.012)	-0.057*** (0.013)

pob_2015	2.10e-10 (8.29e-08)	3.83E-10 (6.60E-08)	1.54E-08 (6.30E-08)	0.000 (0.000)
s_skm	0.000 (0.000)	7.9E-05 (2E-04)	-1.95E-05 (0.002)	0.000* (0.000)
dmu	-0.001** (0.001)	-0.001*** (478)	-0.002*** (0.001)	-0.001 (0.000)
tcp	-0.014 (0.011)	-0.016* (0.008)	-0.017** (0.008)	-0.035*** (0.009)
Number of observations	51	51	51	51
Lambda	-	-	0.851*** (0.141)	-
Rho	-	0.787 (0.000)	-	-
Likelihood test Probability	- -	7.599 0.006	10.141 0.001	-

Note: * represents that it is statistically significant at the level of 10%, ** at 5% and *** at 1%, respectively.

Source: Own elaboration with STATA 12.1 and with the library "spatreg" of the same software.

Regarding the selection of one of the two spatial econometric models, spatial autoregressive model (SAM) and the spatial auto-regression error model (SEM), the results are acceptable at first sight for the spatial auto-regression error model (SEM), given that the lambda parameter is also significant, considering the result of the likelihood test, the latter also turns out to be significant. What would confirm that the informality in the ZMVM, in what refers to the phenomenon of spatial autocorrelation, could not be explained by means of dependence of one region with respect to another. Or what is said, a neighboring locality (municipality or delegation) does not influence another in its informality levels, see the lower part of the second and third columns of Table 2.

As regards the results obtained by means of the quantile regression for $\tau = 0.50$, only the variables dmu and pob_2015 do not explain informality. Where we can see that the sub-sample factors are better predictors compared to the total sample, see the fourth column of Table 2. From the last results obtained, an alternate regression is made based on the proposed model of the equation (15) where we estimate the parameter $\hat{\delta}$ using quantile residues as a function of the quantile residues with spatial lag with the weight matrices W1 and W2 (previously defined).

According to the above, in the first and second column of Table 3, only using the matrix W1 delta is significant. Which shows that only distance factors explain residual spatial autocorrelation, within the localities that are part of the ZMVM, but

not factors of distance and education together. More concretely, we can see that although the years in education have a negative impact on informality, but it does not turn out to be a variable of attraction (in its low levels), and that it influences to increase informality in other nearby localities. From the last obtained results, the Moran indexes were obtained using both matrices W1 and W2. The results are shown in the third and fourth columns of Table 3, where both indices are not significant that is, no empirical evidence was found that there is a local residual spatial autocorrelation within the ZMVM, see Figure 4.

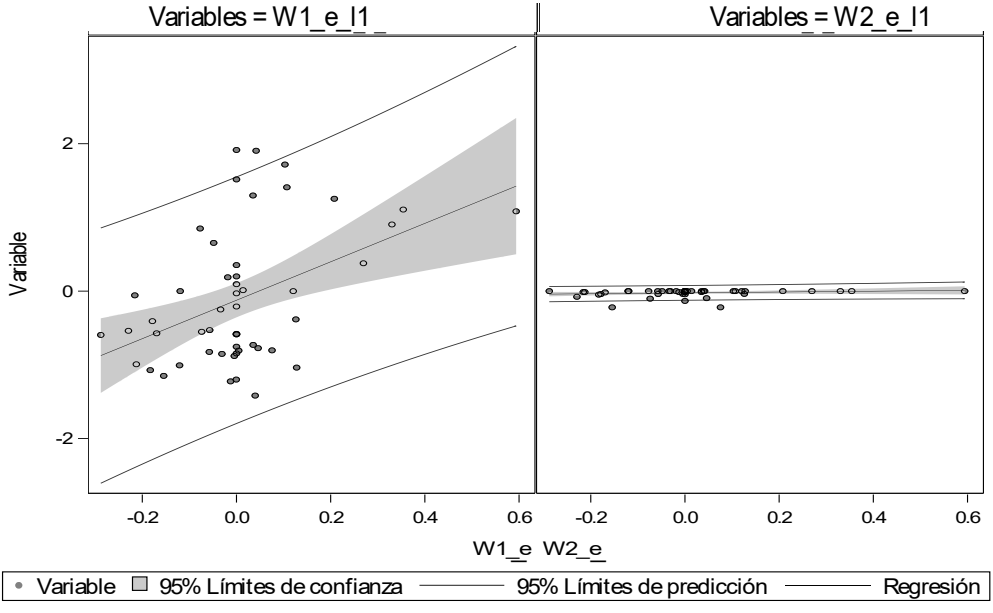
Table 3
Quantile errors results

Delta W1 Coeff./Std. Error.	Delta W2 Coeff./Std. error.	Moran index for W1	Moran index for W2
0.073*** (0.021)	0.396 (0.391)	0.073 Probability 0.398	0.161 Probability 0.394

Note: * represents that it is statistically significant at the level of 10%, ** at 5% and *** at 1%, respectively.

Source: Own elaboration with STATA 12.1 and with the library "spatreg" of the same software.

Figure 4
Quantile error vs. lagged quantile error for W1 and W2



Source: Own elaboration with SAS Software

According to the econometric results, previously exposed, we can summarize that the phenomenon of informality is present within the entire ZMVM, but in different magnitudes. Where there are high levels of schooling, the presence of informality is lower and vice versa. However, there is no strong spatial relationship since there may be localities with low levels of education that are neighboring high-level study localities –so the neighborhood factor does not influence (in a positive or negative way) another locality in its informality level, as a norm generalized within the ZMVM.

CONCLUSIONS AND RECOMMENDATIONS

From the need to be able to answer what are the variables that impact informality in the ZMVM? and understand how the spatial relationship occurs within this same region. According to the results obtained, the negative impact of the years of study on informality levels in the ZMVM stands out. This result shows where the phenomenon should be attacked. On the other hand, a negative impact was also found on the average urban density, which can be seen as an immediate consequence that with a greater density of space, informal markets are saturated by their low added value provoking that the informal sector of the population looks for sources of formal income.

On the other hand, the spatial specification of residual autocorrelation turned out to be better explanatory, which allows us to observe that there is a scarce spatial relationship of the phenomenon in the ZMVM. That is, the informality of a municipality or mayoralty does not influence another (within the same region) in generalized terms. Despite the high social and economic integration that exists among them.

Regarding the results of the quantile model, only the population and density variables are not explanatory in this model, which shows that sub-sample specifications are the ones that best explain the phenomenon of informality. Last results come to reinforce the limitation of this work. Therefore, these reveal the challenge of the municipalities, which share high levels of informality in how they can be coordinated in order to carry out joint strategies in the solution of the phenomenon, together with the restrictions legal, budgetary and political immediacy they face.

About the analysis of quantile spatial errors (using weight matrices). The latter show us that only distance factors explain autocorrelation in spatial errors. In addition, the Moran index also confirms a low relation between municipalities or delegations of the ZMVM, in relation to informality. It is corroborated that only sub-sample factors explain informality in the region –in terms of spatial interrelation.

Following this idea, it calls for similar studies to be carried out in the different metropolitan areas of the country to provide more information to combat

this problem. The research also invites the development and implementation of a Metropolitan Commission to promote formality, which avoids possible contagion impacts between the same municipalities or delegations of the ZMVM. This is how to establish programs at the municipal level and even at the colonial level, since the phenomenon of informality does not behave in a homogeneous way between delegations or municipalities.

REFERENCES

Anselin, Luc, (1988a). "Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeinity", *Geographical Analysis*, Vol. 20, No. 1. https://doi.org/10.1111/j.1538-4632.1988.tb00159.x

Anselin, Luc, 1988b. *Spatial Econometric: Methods and Models*. London: Kluwer Academic Publisher.

Baltagi, Badi H., 2003. *A Companion to Theoretical Econometrics*. Malden, MA.: Blackwell Publishing Ltd. DOI:10.1002/9780470996249

Berndt, Ernst R., Hall, Bronwyn, Hall, Robert and Hausman, Jerry, (1974), "Estimation and Inference in Nonlinear Structural Models", *Annals of Economic and Social Measurement*, Volume 3, number 4, National Bureau of Economic Research, Inc. p. 653-665

Bonilla, R. R., (2015). "Informalidad y precariedad laboral en el Distrito Federal. La economía de sobrevivencia", *Economía Informa,* número 391. Marzo-abril.

Brandt, N. (2011), "Informality in Mexico", *OECD Economics Department Working Papers*, No. 896, OECD Publishing, Paris, https://doi.org/10.1787/5kg3nzlp1vmq-en.

Castells, M., (1989). The informational City. Information Technology, Economic Restructuring and the Urban – Regional Process. Oxford, UK: Basil Blackwell,

Cunningham, W. (2001). "Breadwinner versus caregiver: Labor force participation and sectoral choice over the Mexican business cycle", In Elizabeth G. Katz and Maria C. Correia (eds.) *The economics of gender in Mexico: Work, family, state, and market*, Washington, DC: The World Bank.

Deaton, A. (1997). The analysis of household surveys: a microeconometric approach to development policy. Washington, DC: The World Bank.

Environmental Systems Research Institute, INC. (Esri). (2014). ARCGIS, Redlands, California. USA.

Escamilla, I., (2006). "Terciarización y segregación ocupacional en la periferia expandida de la Ciudad de México", en Adrián G. Aguilar (Coordinador). *Las grandes aglomeraciones y su periferia regional. Experiencias en Latinoamérica y España.* México, Instituto de Geografía, UNAM. Miguel Ángel Porrúa.

Flores, D. y J. Valero (2012). "Evolución y causas de la informalidad en México", en J. Calva (coord.), *Reforma fiscal integral. Análisis estratégico para el desarrollo*. Vol. 5. Juan Pablos Editor y Consejo Nacional de Universitarios, México, D.F.

Gong, X., Van Soest, A., & Villagomez, E. (2004). "Mobility in the urban labor market: a panel data analysis for Mexico", *Economic Development and Cultural Change*, 53(1), 1-36. October.

Graizbord, Boris. (2008). Geografía del transporte en el área metropolitana de la Ciudad de México. México: El Colegio de México A.C.

Harris, J. y M. Todaro., 1970. "Migration, Unemployment and Development: A two-sector analysis", *American Economic Review*, Vol. 60, N°1, pp.126-142.

INEGI, (2010). *Delimitación de las Zonas Metropolitanas de México*, 2010. Aguascalientes, México: Instituto Nacional de Estadística y Geografía.

INEGI, (2015). *Módulo de Condiciones Socioeconómicas. Encuesta Nacional de Ingresos y Gastos de los Hogares*, 2014. Aguascalientes, México: Instituto Nacional de Estadística y Geografía. Descripción de la base de datos.

Iracheta Cenecorta, A. (2003). "Gobernabilidad en la zona metropolitana del valle de México", *Papeles de población*, vol. 9(36), pp. 211-239.

Kelley, C.T., (1999). *Iterative Methods of Optimization*. Philadelphia: Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611970920

Koenker, R. and G. J. Bassett. (1978). "Regression quantiles", *Econometrica*, vol. 46, no. 1, pp. 33-50. DOI: 10.2307/1913643

Lütkepohl, H. (2006). *Introduction to Multiple time Series Analysis*. Berlin: Springer Verlag.

Moran, P.A.P., 1950. "Notes on Continuos Stochastic Phenomena", *Biometrica, Vol.* 37 (1), pp.17-23.

OCDE., 1995. *Estudios económicos de la OCDE, México*. París: Organización para la Cooperación y el Desarrollo Económicos.

OIT. (2013). La medición de la Informalidad: Manual estadístico sobre el sector informal y el empleo informal. Ginebra: Organización Internacional del Trabajo.

OIT. (2014). *El empleo informal en México: situación actual, políticas y desafíos.* Programa de Promoción de la Formalización en América Latina y el Caribe (FORLAC). Organización Internacional del Trabajo.

Pacheco Gómez Muñoz, M. E. (2004). Ciudad de México, heterogénea y desigual un estudio sobre el mercado de trabajo. México: El Colegio de México.

Pérez Pineda, Jorge A., (2006). "Econometría espacial y ciencia regional". *Investigación Económica*, vol. LXV, número 258, pp. 129-160.

Perrot, A., 1995. Les nouvelles théories du marché du travail. Paris: La Découverte.

Perry, G. et al., 2007. Informalidad: escape y exclusión: Salida y exclusión. Washington, DC: Banco Mundial.

Pinto, A. (1970). "Naturaleza e implicaciones de la heterogeneidad estructural de la América Latina", *El Trimestre Económico*, Vol. 37(145), pp. 83-100.

Schneider, Friedrich; Buehn, Andreas; Montenegro, Claudio E. (2010). *Shadow economies all over the world: new estimates for 162 countries from 1999 to 2007*. Policy Research working paper; no. WPS 5356. Washington, DC: World Bank. http://documents.worldbank.org/curated/en/311991468037132740/Shadow-economies-all-over-the-world-new-estimates-for-162-countries-from-1999-to-2007

Pradilla Cobos, Emilio, (2005). "Metropolitan Area in the Mexico Valley: Megacity without planning", *Ciudades.* (9), pp. 83-104.

SAS Institute Inc. (2014). SAS, Cary, NC, USA.

Sassen, S. (1991). *The Global City*. New York, London: Princeton University Press.

Stata Corp LLC. (2012). Stata, V. 12, Texas, USA.

Tokman, V. E. (1987). "El sector informal: quince años después", *El Trimestre Económico*, Vol. 54, número 215, pp. 513-536, julio-septiembre.

Appendix A: construction of variables

	*Constructed variable**	*Definition*	*Values used*	*Generated variable*	*Type*	*Description*
Informality	Atemed	Affiliated persons or not, to some institution that provides medical attention	1=Yes 2=No	Social security (seg_so)	Dichotomous variable	If the event is 1, individual does not count social security.
	inst_1	Medical institution IMSS	1			Variables that allow to identify the social security institution. Those people receive Seguro Popular are informal.
	inst_2	Medical institution ISSSTE	2			
	inst_3	Medical institution ISSSTE (state type)	3			
	inst_4	PEMEX, National defense or Navy	4			
	inst_5	Other	5			
	Age	Years since the birth up to the day of the interview	age>=15		Numerical	Age equal or over 15 years old.
	Segpop	Affiliated to the Seguro Popular	1=Yes 2=No	Seguro Popular segpop	Dichotomous variable	

Income	ing_cor	Current revenue	Average entities	Average monthly income ing_cor	Numerical	Income average sum for work, rentals, transfers, and other type of income.	
Years of school attendance	Level aprob.	Education all levels approved	{0,…,9}	Years at school	Numerical	None	0 years
						Pre-primary	2 years
						Elementary school	8 years
						Junior high school	11 years
						High school	15 years
						Normal	19 years
						Technical career	15 years
						Career	19 years
						Master´s Degree	21 years
						Ph. D.	25 years
Marital status	edo_conyug	Conjugal condition of the member of the household	{1,…,6} 1=married 0=lives alone / single	Marital status edo_civil	String variable	Marital status	
Sex	Sex	Characteristics that distinguish organisms on the basis of their reproductive function	{1,2}	Sex	Dichotomous variable	Sex of the person	

Source: Own elaboration with data from: INEGI (2010) and MCS (2014).