Equilibrio Económico, Revista de Economía, Política y Sociedad. Vol. 14 (1) Semestre enero-junio de 2018
Núm. 45, pp. 7-42 ISSN: 2007-2627

Informal Earnings in the Labour Market: The Mexican Case

Luis Huesca *
Linda Llamas *

Abstract

This research is aimed at providing recent empirical evidence on the determinant factors prevailing in Mexico with regards to the formal-informal labor market earnings. In order to deal with issues of self-selection in polychotomous models the Bourguignon's et al. (2007) novel methodology is used. The empirical application solves the puzzle of finding reasons why informal work seems to be rationed by demand, as well as a reason to confirm that occupational choices are not selected at random, i.e. decisions are rational. Occupational self-selection process points out a direct relation between age of the worker and informal entrepreneurs.

Keywords: *Earnings, Informal sector, Occupational choice, Wage gap*

* Centro de Investigación en Alimentación y Desarrollo, A.C., & CIRPÉE visiting scholar, Université Laval, Québec, Canadá. Address: Hermosillo, Sonora, Mexico, Carr. a La Victoria Km. 0.6, zip. 83304.
lhuesca@ciad.mx
linda.llamas@gmail.com

Los ingresos informales en el Mercado de trabajo: El caso Mexicano

Resumen

Esta investigación muestra evidencia empírica reciente sobre los factores determinantes de las remuneraciones en el mercado laboral formal-informal prevaleciente en México. Con el fin de abordar los problemas de autoselección en modelos politómicos se implementa la novedosa metodología de Bourguignon et al. (2007). La aplicación empírica determina por qué el trabajo informal parece estar distribuido por la demanda, así como una razón para confirmar que las opciones de ocupación no son elegidas al azar por los trabajadores, es decir, son racionales. La autoselección ocupacional indica una relación directa, entre la edad del trabajador y su vínculo de pertenecer al grupo de emprendedores informales.

Palabras clave: *Ingresos, Sector informal, Selección ocupacional, Brecha salarial*

JEL Classification: J24, J31, J42, O15, O17

Recibido el 7 de septiembre de 2015. Recibido en forma revisada el 30 de junio de 2017. Aceptado el 21 de julio de 2017.

I. Introduction

Failures in policies to promote formal labor markets are crucial to understand why the informal sector in developing countries remain as structural, with specific features such as in the Mexican case, where informality reached sixty percent of the workers in 2013 (Inegi, 2014). This labor condition offers a certain explanation on behalf of the insufficient capacity of the Mexican economy to generate enough employment for their workers, not to mention the increase of precarious jobs created within the formal sector.

The least skilled individuals in Mexico (as in most of the Latin American countries) earn most of their income from the wage-earning source. Because they cannot rely on unemployment insurance, the employment stability tends to be more valued rather than to obtain higher wages. Hence, it is important to investigate the labor market specific qualities and the new role that the formal-informal sectors carry out in both the productive and labor activity.

The functioning of the informal sector in the labor market has been commonly categorized throughout the time as an absorption for those productive activities that have not yet managed to be inserted into the legal and structured economic dynamics (Hart, 1971; ILO, 1972; Leontaridi, 1998; Azuma and Grossman, 2002), and recently including both, internal and external mobility of workers such as the case of rural China (Wu, 2010). The inherent factors for informality have been related with very low levels of labor productivity or structural imperfections associated to the economic system, being the last one a regular basis for developing economies.

The goal of this study is to offer an overview in the labor market that is opposed to traditional thesis of segmentation. These theories postulate that institutional

factors, such as unions or the public intervention in determining wages produces segmentation in the market for higher formal and lower quality informal jobs.[1]

Thus, this study analyse these two segments in the labor market and contributes to an scarce empirical literature on labor economics, applied for the first time –to authors' knowledge– to the Mexican case. Previous analysis of this particular experience has only been focused on earnings, but only as a binary occupational choice, pooling wage earners and self-employed. The novel and still fresh polychotomous self-selection methodology has been applied in this research [Hereby BFG (Bourguignon et al., 2007)] in combination with the path-breaking work of Oaxaca (1973) which is based on the multinomial logit model. This technique includes multiple correction terms and goes beyond the classical Heckman self-selection binary one (Heckman, 1979). BFG allows to detect a selection bias in the earnings equations to the position of workers, as well as to link the selection bias into the allocation of workers to each alternative of occupation by decomposing the earning differences.

The order of the paper is as follows. Next section provides a review of the literature and justifies the framework; the third section describes briefly the structure of the urban labor market in Mexico. The fourth section presents the methodology and the analysis of the database. Fifth section presents the empirical results and summarizes the findings of the document.

II. Definitions and theoretical framework

It is well known that the term of informality or informal sector was initially introduced at the beginning of the 1970s by Hart (1971, 1973), analyzing the labor market of Ghana and shortly after by the International Labor Organization (ILO) with its study made in Kenya (ILO, 1972), which was the first of its type that

[1] On this matter take a look at the research done by Fields (1975), Mazumdar (1983), Dickens and Lang (1985), McNabb and Ryan (1990), and Magnac (1991).

systematically allowed to quantify the activity of the informal sector and advised its government to recognize its great utility. The essence of its concept continues to be the same up to date, except that different perceptions are currently applied (Jusidman, 1993; Cervantes et al., 2008; Macias and Cazzavillan, 2010; La Porta and Shleifer, 2014) concerning the functioning of the labor market in developing countries.

Literature regarding Mexican informal sector is continuously growing. For instance, Macias and Cazzavillan (2010) uses a structural equation model and macroeconomic time series data to track the informal economy's evolution for 1970-2006 period; estimates reveal economic factors such as taxation, low salaries and excessive regulation as main contributors to increase informal economy. These factors accounted for 40% of GDP during the first two decades (1970's and 1980's) and decrease to reach 30% of GDP during the 1990's onwards. Previous study from United Nations Human Settlements Program (Habitat, 2006) also indicates that Mexican informal sector was stable over the period 1993-2002, after comparing the size of informal jobs as a share of GDP; albeit these results imply that informal sector has not been growing at a higher rate than the rest of Mexican economy, it exposes the lack of capacity to reduce this condition.

The informal activities have more prevalence in the businesses managed by self-employed individuals without any registry and wage-earners without social security benefits.[2] Its highest expression appears in the former workers with levels ranging from 70% in Africa, to 60% in Latin America and 59% in Asia (ILO, 2002). At early 2000's, the relative participation of informality for Mexico ranged from 36% to 45% for the self-employed and a greater participation for wage-earners accounting for a share up to 50% for the wage-earners (Cervantes et al., 2008).

[2] According to Van der Sluis et al. (2004) the literature pertaining to the measurement of returns to schooling in *entrepreneurship* or its closest equivalent concept, self-employment is still poorly defined, both for developed and developing countries.

Cervantes et al. (2008: 47) show that informality in Mexico accounts for 65% of their labor force.

The growth of the informal activity in Mexico generates future social problems. Both, wage-earners and self-employed workers are not covered with medical services and do not contribute to social security, weakening the effectiveness of future health and pension systems. According to Levy (2010), continuous transition between formality-informality has also substantive implications for the effectiveness of social programs, since wage earners and self-employed workers do not have permanent social security coverage. Moreover, Levy's position is that mexican social programs fosters informality and subsidize low productivity jobs, so that policy reforms must consider their impact on informality.

In order to analyse this *status quo*, occupational categories such as wage-earners and self-employed have been used in many studies and for different countries (Taylor, 1996; Cohen and House, 1996; Maloney, 2002, 2003; Tannuri et al., 2004; De Acevedo, 2004; Zhang, 2004; Huesca, 2005; and Moreno-Treviño, 2007). Following the legal and statistical approach (Roubaud, 1995, and Cervantes et al., 2008) it is considered for the informal sector those working as self-employed which has no name or official registry and those wage-earners without receiving social security benefits.[3]

Using a static perception, the work of Cohen and House (1996) shows the coexistence of a great number of very capable and skilful workers who have not been formally trained in the formal and informal sectors in Sudan, explaining why the activities better paid are for the self-employed workers and for the employees informally working in micro-enterprises. Tannuri et al. (2004) analyze the labor participation and earnings in the Bolivian formal-informal sectors, using quantile techniques with adjustment for selection bias. They found that the informal self-employed are the most affected category by selection and that human capital

[3] A special typology of informality has been shown in Cervantes et al. (2008), which agrees with the one used in this study.

endowment is best specified for the formal sector at the median and higher quantiles.

De Azevedo (2004) applied a polychotomous technique to explain earnings in the slums in Rio de Janeiro, Brasil, and found lower returns to education and lower salaries in the informal sector. This latter evidence in Bolivia is a signal of the good opportunities offered by the formal labor market in this country. Maloney (2002) observes a high heterogeneity of mexican workers between both sectors as well, estimating higher returns for education in certain informal points. Zhang (2004) examines self-selection and wage differentials for urban China using a polychotomous technique in a four sectoral model, dividing labor selection among government, state-owned, urban collective and private-individual enterprises.

Last but not least, it seems for Mexico that higher levels of schooling have a net gain if informal workers would change into the formal sector, while workers with lower levels of education would have a negative effect on their earnings had they decided to become formal workers (Moreno-Treviño, 2007). Another research computing earnings differentials among these sectors but adding the gender component was made by Huesca and Camberos (2009), and its findings using counterfactuals and a semi-parametric kernel approach depicts how self-employment is better paid for men than for women between 1992 and 2002, meanwhile women seem to have greater gains as formal wage-earner than men.

Indeed, the value added of the informal economy as a share of GDP imply that labor productivity is low in Mexico, as it accounted for an average of 12.5 percent of GDP from 1993 to 2002 (Habitat, 2006), that is, a yearly average of 1.38%. Likewise, La Porta and Shleifer (2014) point out that formal firms are more efficient when they are managed by educated entrepreneurs, while small and inefficient firms whether formal or informal, are generally run by uneducated entrepreneurs. The authors notice that entrepreneurs from informal economies are unproductive and the transition to formality is driven by economic growth.

If the informality trend obeys to efficiency reasons, the proper theoretical framework leads to the use of Heckman and Sedlacek seminal research (1985), supporting the existence of a competitive labor market where the workers interact simultaneously and compete, developing their activities in the sector that makes them better-off the most, in terms of earnings due to their skills in the productive process. The subject analysis is explained by two basic features: 1) Despite the predictions of its possible reduction, the informal economy has not only been increasing at a world-wide level, but it is present within the labor market in new ways and, with unexpected patterns as well, and 2) regardless the intense debates about the definition of their qualities and which type of policies would allow to legalize their activity, it is widely recognized that the informal economy is a sector that has allowed to promote development and growth and it has fought poverty too (Tokman, 1994; Sethuraman 1997; De Soto, 2000; and Charmes, 2000).

III. The structure of the urban labor market in Mexico

According to information of the National Survey of Urban Employment[4] (ENEU by its acronym in Spanish) the Mexican urban labor market assembles in greater proportion wage-earning workers, with a total of 72% of the occupied population during 1992 and 2004.[5] Then, self-employed workers follows with an average share of 18%. The workers in part-time or piecework represented about 5%, whereas the employers represented 5% of the market force. The rest of categories are residual and participates with less significance. This period is relevant for analysis due to the following facts: North American Free Trade Agreement –NAFTA– was signed; a change in ruling the political power at a national level starting in 2000; and a more stable macroeconomic indicators.

[4] This Survey was carried out by the Instituto Nacional de Estadística, Geografía e Informática –INEGI–.
[5] From this later year, ENEU project disappeared and INEGI changed to a new labor survey named ENOE, which covers all the territory and not just the urban areas in the country. This is the reason why this research covers only the period of 1992-2004.

The informality of the Mexican labor market

The informal segment of the labor market can be defined as that part of the occupied population or under-employed that is not subject to any contractual relation and upon the lack of unemployment benefits, is able to work in exchange of very low remunerations; otherwise, it decides to become self-employed instead of thickening the rows of unemployment. Given these characteristics, informality has become a mantle that not only hides the true magnitude of the unemployment situation in Mexico and Latin America, but it has also become a common characteristic for practically all the poor in the underdeveloped countries. Technological changes produced in the world-wide economy has been biased to more skilled labor (Acemoglu, 2002), but it is a reason as well that could have been one of its effects as indicated by the statistics of ILO (ILO, 2001).

In the case of the growing self-employment work category, the hypotheses can be summarized in two facts: 1) the increase in self-employment work is a simple reflection of the entrepreneurial activity, and therefore presents itself as a better labor option. In these regards, an economical expansion would increase this sort of activities, since the individuals would be more arranged to invest and to risk into a self-employment business, instead of continuing working as wage-earners; 2) working as self-employed would be only a reflection of the lack of opportunities to become a regular wage-earner; for that reason, it would become the last option.[6]

IV. Methodology and data

The data bases of ENEU for 1992 and 2004 are used for the empirical application at each second quarterly in order to create a cross section of the 39 urban areas

[6] Other additional causes exist, such as the worker's civil status, age, connection with the familiar structure, the social and organizational structure where he lives (Heckman and Hotz 1986; Taylor, 1996).

of the sample. This sample comes up to 85% of the urban zones in the country and includes all the cities with a population over 100,000 inhabitants or more.[7] A complete structure and basic statistics of the variables considered in the empirical analysis can be seen in the Appendix (tables A1, A2). Following the papers of Heckman and Hotz (1986) for Panama and Buchinsky (1998) for the United States, only men were selected to reduce earnings dispersion and to refer specific uniform attributes of the workers, which mainly refines the analysis for two bonded reasons in this research: 1) A high proportion of women work in a small-business without receiving any payment (more than 12% in our data); and 2) the proportion of women working as self-employed in the formal sector is undersized in the samples, with no more than 3%. Earnings include both wage-earners and self-employed separately for the econometric model.

Earnings are expressed in American dollars (USD) -adjusted to 2000 constant prices for the second quarter, using the consumer price index by wage levels from the Banco de Mexico[8]- dividing their value by the average exchange rate at the same quarter (9.50 pesos per dollar).[9]

Table A2 included in the Appendix, displays a summary of statistics for the complete sample and the variables considered. The estimation for the density of earnings allows depicting the pattern of behavior in the different probability density functions (FDPs) for any distribution. Densities are considered as a nonparametric *Kernel* technique that smoothes its shape, avoiding "the noise" that induces the use of a sample instead of the total population. Thus, a function $\hat{f}(x)$ is considered on a vector of earnings $x = (x_1, \ldots, x_n)$ under the

[7] See the technical manual of the ENEU for a complete description of the cities in the years previously referred.
[8] Instituto Nacional de Estadística Geografía e Informatica (INEGI) provides a complete series of these indexes (PI), where levels are expressed in minimum wages (MW) but for a different year base; we change the base to 2000=100 and adjust its variation to the next sort out of MW brackets: PI $\leq$ 1; 1 < PI $\leq$ 3; 3 < PI $\leq$ 6; 6 < PI.
[9] Stata 13 was used to compute all the estimations and the two step method ado-file -*selmlog*- programmed by Marc Gurgand version 2007.

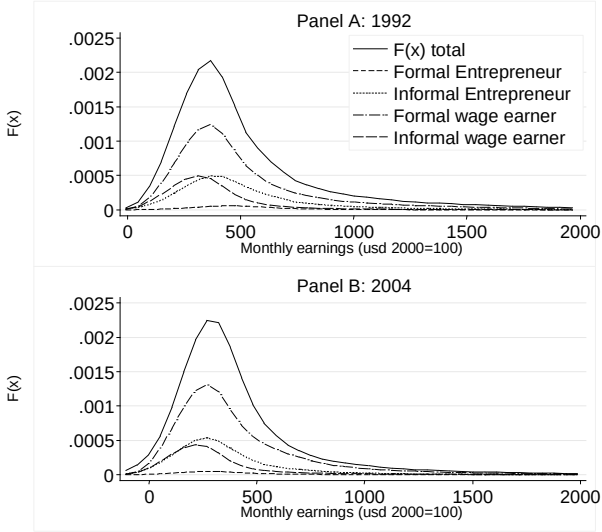
assumption that the sample has been extracted from its own original population density $f(x_j)$. An alternative method that allows decomposing not only the sub-groups that underlie and compose the total FDP, but also to detect the changes operated within each population sub-group, is the one proposed by Jenkins and Van Kerm, (2004). First, the estimation of the densities by sub-groups $k = (1,\ldots,K)$ is obtained in function of its participation weighed them down as follows:

$$f(x) = \sum_{k=1}^{K} \upsilon^k f^k(x) \qquad (1)$$

density function $f(x)$ for each unit of income *x* in the expression (1) represents a weighed sum of the FDPs for each sub-group *k*, where υ^k stands for the population share of the group *k*, and f^k as the FDP of the group *k*. Changes within each population sub-group are not conducted for the purpose of this study, but they are only aimed to estimate decompositions at the formal-informal level. At first glance, densities of distributions –in 1992 and 2004– are calculated using an adaptive kernel estimator (Van Kerm, 2003)[10]. Two features are clear: 1) Average earnings is bridging the gap between the informal entrepreneurs (self-employed) and the formal wage earners in 2004; and 2) formal entrepreneurs and informal wage earners seem to have the lowest average earnings at the conditional mean. Expression (2) is used to decompose the densities of the subgroups for the labor segments of self-employed and wage earners within the formal-informal sectors shown in Figure 1.

[10] The advantage of this estimator is that it avoids smoothing the distribution in those zones of high earnings concentration while maintaining reduced variability of the estimated positions where the information of the data is small, such as in the highest ranks of earnings (Jenkins and Van Kerm, 2004; and Silverman, 1986).

Graph 1. Density decomposition of formal-informal earnings in Mexico, 1992 and 2004. (Grid=110, optimal width)



Source: Authors' calculations using microdata from ENEU, 1992 & 2004.

The polychotomous correction of selection bias.

The model requires to be corrected for selection bias and, consequently, not to obtain estimates that could be consistent, but biased otherwise (Heckman, 1979; Heckman and Sedlacek, 1984). The selection bias must be understood basically because self-selection of individuals in the information and the data can exist (that is, identical individuals when using samples defined with a non-random criterion and participating in two sectors). Also, due to self-handling of the data and information inducing selection bias properly into the sample.[11]

[11] For instance, selection bias causes the same non-observable characteristics that make an individual choose their job, and that could also be increasing his wage in the same sector over the one perceived by another individual with the same characteristics, but removed from the population at random.

This causes not only biased, but also inconsistent coefficients. Accordingly, the two step generalized methodology proposed by BFG for polychotomous cases is used, allowing OLS implementation in the calculations.[12]

It is intended to answer the following questions: Will the differences in earnings between the formal and informal sectors of the labor market in Mexico be statistically significant? Which are the socioeconomic and occupational factors that affect earnings differentials? Will the earnings gap of the labor sectors be higher among the top earners, or among those located at the bottom of the distribution?

Description of the BFG model in the Mexican case.

An earnings equation is used taking as dependent variable, the log of hourly earnings controlling by the following attributes: males working at least 20 hours a week and with positive earnings[13]; education introduced continuously, age to capture experience and quadratic age; a variable detecting the household-head; a dummy variable that measures size structure; four categorical variables for the sector of economic activity; a variable identifying the productive sector of tradable and non-tradable goods; finally, six dummies categorizing the regions in the country.[14] Using the description of Oaxaca (1973), a formal-informal sectoral partition in *i* and *j* is observed respectively; the earnings are represented in log

[12] As opposite to the bivariate case, when the number of events exceeds two categories, previous techniques (Lee, 1983 and so forth) present restrictions on the structure of the error terms and, generally, an inappropriate application –since those methods have been elaborated with the requirement of using an univariate transformation order–. A correction for multivariate cases was developed by Dubin and McFadden (1984); although this technique could not reach to evaluate a model strong enough to admit maximum likelihood estimators, with complete information for the case in which the number of choices were higher than two. This provides a model where the *J* sector must be required to establish a *J*-1 selection terms.

[13] This subsample follows the specification of previous articles applied for the same formal-informal approach and gender topics as well (Marcouiller et al., 1997; Buchinsky, 1998; Meza, 2001; Gong and Van Soest, 2001) for México and the United States, in order to homogenize and to deal with the variance in the models. A model without imposing the 20-hours condition was implemented in our case and we obtained similar results, except for the standard errors, which were higher for that case (Estimates are available upon request).

[14] See Hanson (2004) for more details about the regionalization.

terms by w_{ij} , with the basic assumption that the sense of the absolute discrimination goes from sector *i* towards *j*, that is, $F(w_i \geq w_j)$; therefore, the density of the earnings is represented now as $F(w_{ij})$. The expected differential between the unconditional means of the distributions is obtained as follows:

$$E(w_i) - E(w_j) \qquad (2)$$

The earnings equation to be estimated by OLS for each subsector *i* (formal) and *j* (informal) is represented in a standard fashion conditioned by the global mean of the sample distribution; it is traditionally written as follows:

$$\ln w_i - \ln w_j = \beta_i^{'} x_i + \beta_j^{'} x_j + \varepsilon_{ij} \qquad (3)$$

where $\beta_{ij}^{'}$ and x_{ij} stands for vectors in both coefficients and attributes of the individual respectively, being ε_{ij} the residual term displaying the usual statistical conditions with $E(\varepsilon_{ij} \mid x_{ij}) = 0$, and $V(\varepsilon_{ij} \mid x_{ij}) = \sigma_{ij}^2$. Despite the known sub-samples, for a *J* segment of occupations combining a set of observable characteristics for each worker, as well as the knowledge for the *j* segment to which it belongs, it brings about an idea into the covariance nature and the need to estimate an unbiased vector of parameters for $\beta_{ij}^{'}$. According to the BFG technique, a multinomial Logit must be estimated assuming that the information follows a Gumbel distribution G (η) independent and identically distributed (*iid*). The following model is considered using the notation for sectors *i* and *j* with *M* alternatives for the individuals, with the following notation:

$$y_s = x_s \beta'_s + u_s \qquad (4)$$

$$y_s^* = z_s \gamma_s + \eta_s, \quad s = 1,\ldots,M \qquad (5)$$

Where x_s and z_s compose a vector of independent variables and the disturbance term u_s which confirms the usual conditions $E(u_s \mid x, z) = 0$, and $V(u_s \mid x, z) = \sigma_s^2$. The impact on the dependent variable y_s is observed just for the case where the alternative *M* is chosen which happens when:

$$y_s^* > \max_{j \neq s}(y_j^*) \qquad (6)$$

The BFG technique demonstrates that it is feasible to use OLS consistently, once the polychotomous vector of probabilities has been included in the estimation; then, the existence of a high correlation between the explanatory variables and the disturbance terms of equation (5) becomes corrected. BFG (2001: 2) shows that the condition (6) is equivalent to the replacement of the elements into the equation (5) as $z_s\gamma_s > \varepsilon_s$ where $\varepsilon_s = \max_{j \neq s}(y_j^* - \eta_s)$.

It is confirmed that the vector (η_s) is *iid* and Gumbel distributed; thus, their respective cumulative and density functions are $G(\eta) = \exp(-e^{-\eta})$ and $g(\eta) = \exp(-\eta - e^{-\eta})$ (See McFadden, 1973). It is in this part of the model where it applies the multinomial Logit specification in the traditional way:

$$P(z_s\gamma_s > \varepsilon_s) = \frac{\exp(z_s\gamma_s)}{\sum_j \exp(z_j\gamma_j)} \qquad (7)$$

Bourguignon, et al. (2001: 6) proves that using the coefficients $(\hat{\gamma}_s)$ in the estimation of expression (7) will provide the predicted probabilities $m(\hat{P}_s)$ of the model for a generalization; although, there would appear more than one solution for the vector *z* in order to obtain the different probabilities in the expression

(7).[15] The operation is made weighting and dividing (7) by the expression $\exp(z_s\gamma_s)$ for *M*=*j* as the base category. Finally, the following expression is used to consider the vector of unbiased coefficients β'_s :

$$y_1 = x_1\beta_1 - \sigma_1[\rho_1 m(P_1) + \sum_{s>1}\rho_s \frac{P_s}{P_s - 1} m(P_s)] + v_s \qquad (8)$$

where $m(P_s)$ are the probabilities obtained from the formula (7) and $(\sigma_1\rho_1),\ldots,(\sigma_1\rho_j)$ as the coefficient terms for the polychotomous correction of selectivity bias; v_j is an orthogonal error parameter towards the other terms, having a mean expectation equal to zero. This last property allows using directly the OLS procedure in the estimation.[16]

Besides the coefficient of the base category on the *M* set, the outcome of the previous selection coefficients will reflect the effect on the probability differences among the true coefficient on category *z*. A four selection model has been estimated for this case, obtaining the polychotomous probabilities of the workers not considered in sector *i* (formal) –wage-earner or self-employed–; and the workers who do not belong to sector *j* (informal). In both sectors *i* and *j* we use the Oaxaca (1973) approach and separate models were estimated for each labor category as follows,

$$\ln(w_i) - \ln(w_j) = \Delta x\ \beta_i^{'} + x_j \Delta\beta^{'} - \sigma_{\eta u}\rho_{ij}^{'} \qquad (9)$$

[15] Since this equation is sub-identified the solution is to choose one of the *M's* alternatives to standardize to zero the coefficient in this category.

[16] One problem that arises from this occupational selection process technique is related to the Independence of Irrelevant Alternatives (IIA) as stated by Hausman and McFadden (1984). On this matter, Bourguignon et al. (2007) proves that selection bias correction, based on the multinomial logit model, can provide fairly good correction for the outcome equation, even when the IIA hypothesis is violated.

where $\sigma_{\eta u,}$ stands for the covariance between the error term of the earnings equation; $\rho^{k}_{ij,}$ stands for the selectivity terms. Replicating the process at the conditional mean, changes at the quantile level between the characteristics and the coefficients are also estimated, taking the log of earnings and introducing the vector of $(\sigma_{\eta}\rho'_{ij})$ variables into the equation as $(\sigma_{\eta u}\rho^{k}_{ij})$; thus, the equation (9) is represented using θ for quantile notation as:

$$\ln(w_i^{\theta}) - \ln(w_j^{\theta}) = \Delta x^{\theta}\beta_i^{'\theta} + x_j^{\theta}\Delta\beta^{'\theta} - \sigma_{\eta u}\rho_{ij}^{k} \qquad (10)$$

Considering simultaneity in the sectoral models, ρ_{ij}' and ρ^{k}_{ij} measures the effect of the sample selection on earnings, derived from the absence of a wage offered to the individuals out of the formal sector and vice versa. Therefore, the statistical significance of the coefficients $\sigma_{\eta u}\rho^{'}_{ij}$ reflects the impact meaning on the sample selection, allowing estimating, at the same time, the earnings equations coefficients by OLS.[17]

Following Cohen and House (1996) research, for the use of the polychotomous Logit –as settled by the BFG technique– the position of the individuals in the labor market settles down in two sequential decisions, where each one displays multiple criteria of selectivity[18]. In order to estimate both, the informal segment and wage differentials, empirical literature on labor market segmentation in developing countries pool together self-employed workers with informal wage earners (see Cohen and House, 1996; Marcouiller et al., 1997; Saavedra and Chong, 1999 for

[17] Combining both segments (formal-informal) it is possible to obtain a picture of both the nature of the covariance between their sample mean as well as the stochastic error terms which captures the unobserved changes, being at the same time, independent of the vector x_{ij}, which finally allows to obtain unbiased estimates of the vector of parameters $\beta^{'}_{ij}$.

[18] The first is conceived implicit, since it refers to the decisions considered by the employer to hire certain individuals; information that is included in the Z attributes vector, at the same time. The second is the individual self-decision to work in a labor segment that maximizes his utility function.

instance). They, however, have ignored formal-informal sector differences and simply have focused their analysis on differences among wage earners and self-employed. Combining such categories can be risky because it can only be valid if no fundamental underlying differences among the distinct groups are found.

V. Empirical results

From the equation (5) following the human capital approach, the dependent vector y_s^* disaggregates the individuals in M_i which will refer to i=1, if an individual is a formal self-employed, i=2 if he is an informal self employed; i=3 if the individual works as a formal wage-earner; and i=4 if works as an informal wage earner. The equation is controlled by the following covariates: schooling (*school*); household head (*head*); if the worker lives in a household with numerous members (*dmembers*); if their location is at a tradable export sector (*prodtype*) and a group of dummies for regional and economic activity sectors, omitting region 5 (*South*) and sector 4 (Services) to avoid the dummy trap and collinearity. The level of confidence for the estimation in the models is 95%. Next, the results of the models for the 1992 and 2004 periods are analyzed in tables 1 and 2. Relative Risk Ratios (RRR) from the multinomial Logit model is shown in table 1 as an indicator of the occupational incidence relative to each covariate. Formal wage-earner is the omitted variable, thus, schooling levels indicate that for every additional degree the individual reduces its likelihood of working as an informal wage-earning or self-employed and formal self-employed as well for both years.

Experience as an additional human capital attribute allows inferring a different behaviour of the workers compared to the previous argument, as long as its RRR is greater than 1 for the informal self-employed in both years. Household head individuals display a strong effect for both years, whether they are working as informal or formal entrepreneur; household size has a relevant impact for all the

labor segments, except for the informal wage earners in 2004, even though the risk ratio is close to one. By all means, construction (dsec2) has a strong impact in the informal wage-earner segment, occupation typically involved with low skilled workers receiving reduced salaries in Mexico; while jobs in commerce (dsec3) show a high risk of informality, where even the informal wage earners workers actively participate in 2004.

Table 1. Relative risk ratios of worker's participation in Mexico, 1992 and 2004 (maximum-likelihood estimates –first stage-)

Variables	1992			2004		
	Wage earner	Self-employed		Wage earner	Self-employed	
	Informal	Formal	Informal	informal	Formal	Informal
Education	0.809*	0.801*	0.821*	0.865*	0.855*	0.865*
Age	0.851*	0.986**	1.028*	0.854*	0.975**	1.024*
Age2	1.002*	1.000*	1.000*	1.002*	1.000*	1.000*
Head	0.929**	3.369	1.245**	0.787**	3.158	1.148**
dmembers	1.192**	1.385***	1.039**	0.998**	1.278***	1.014**
prodtype	0.476**	1.802	0.593**	0.293**	1.437	0.829***
dsec1	0.346*	1.165***	0.313*	0.453**	0.207*	0.165*
dsec2	1.326	0.803	0.537**	1.679	0.248***	0.783***
dsec3	0.887**	1.277***	1.584**	1.440***	0.349**	1.074**
dsec5	0.238**	0.121***	0.020*	0.579***	0.069**	0.405**
dregion1	0.690**	0.863	0.615**	0.698**	0.707	0.579**
dregion2	0.833**	1.022	0.799**	0.669**	1.418	0.663**
dregion3	0.833**	2.845	0.895**	0.972**	2.484	0.896**
dregion4	0.786**	1.154	0.760**	0.922**	0.638	0.837**
dregion6	0.543**	0.853	0.716**	0.542**	0.743	0.561**
Pseudo R^2		0.1404			0.1346	
Log likelihood		-62,895.536			-92,663.605	
LR Chi (45)		20,537.87			28,824.46	
Prob > χ^2		0.000			0.000	
Obs.		46,789			69,311	

Note: Significance level at 1%*; 5%**; and 10%***
dregion1: Border; dregion2: North; dregion3: Capital; dregion4: Center; dregion5: South; dregion6: The Yucatán Peninsula.
Source: Authors' calculations using the data base of ENEU, 1992 and 2004.

In order to acquaint for the impact on earnings and determine its possible potential loss for the workers, table A3 from Appendix shows the estimates for the earnings corrected equations using the BFG technique. As usual, the dependent

variable is the natural logarithm of monthly earnings, and four equations are estimated including four selection terms representing the chosen labor segment, derived from the predicted values of equation (7).

It is observed that the coefficients display the expected signs for both sectors and labor category. Education returns are similar for both sectors where it can be seen an increasing tendency during the period (see graphs 2 and 3). Estimates are in line with studies for Mexico that have used the same data of this paper and information from the National Survey of Household Income and Expenditure -ENIGH- (Marcouiller et al., 1997; Barceinas and Raymond, 2003; and Huesca, 2004).

Coefficients that measure self-selection are statistically significant, except for the formal self-employed segment in 2004, suggesting that in this labor segment the existence of non-observable factors in the process of wage determination is not a problem. The fact that *sigma-Rho* coefficient shows a negative sign, would mean that this occupational classification would receive a lower wage, had the worker decided to participate in the counterpart sector.

Being a formal worker as wage-earner would be the best choice in order to obtain a higher wage. In order to test the presence of unobservable on the occupational choice process, an F test was conducted to prove that all $(\sigma_\eta \rho_{ij})$ are cero. The null hypothesis that the $(\sigma_\eta \rho_{ij})$ are cero across all four segments is rejected, although the effects of unobservable are greater for wage earners than for self-employed (See table 2). We run the proportionality hypothesis as well. This hypothesis is shown in tables 2 and 3, inferring that the wage setting process and self-selection will be proportionately recognized if both sectors have the same slope of selectivity-corrected coefficients, but the intercepts may vary according to each sectoral wage setting process (Heckman and Sedlacek, 1985: 1087).

A jointly significance test is carried out first, where the coefficients are meant to be equal to zero; second, a model for each sector is applied and the test is carried out again, assuming that the Sigma-Rho coefficients in the four choice specification are equal for each other.[19]

Table 2. Coefficients of selectivity variables in earning equations, 1992
(Proportionality hypothesis between formal/informal sectors)

	Wage earner		Self-employed	
Variables	Formal	Informal	Formal	Informal
$\sigma_\eta\rho_1 = \sigma_\eta\rho_2 = \sigma_\eta\rho_3 = \sigma_\eta\rho_4$	38.70* F(4, 26128)	16.37* F(4, 8743)	40.91* F(4, 1316)	44.96* F(4, 12167)
$\sigma_{\eta i}\rho_{1i} = \sigma_{\eta j}\rho_{1j}$	4.28**		6.45**	
$\sigma_{\eta i}\rho_{2i} = \sigma_{\eta j}\rho_{2j}$	43.88*		22.16*	
$\sigma_{\eta i}\rho_{3i} = \sigma_{\eta j}\rho_{3j}$	59.95*		47.67*	
$\sigma_{\eta i}\rho_{4i} = \sigma_{\eta j}\rho_{4j}$	25.95*		65.06*	

* Indicates significant at the 1% level, ** significant at the 5% level; *** Significant at the 10% level.
Source: Authors' calculations using the data base of ENEU, 1992.

Table 3. Coefficients of selectivity variables in earning equations, 2004
(Proportionality hypothesis between formal/informal sectors)

	Wage earner		Self-employed	
Variables	Formal	Informal	Formal	Informal
$\sigma_\eta\rho_1 = \sigma_\eta\rho_2 = \sigma_\eta\rho_3 = \sigma_\eta\rho_4$	58.71* F(4, 40607)	8.21* F(4, 11452)	40.91* F(4, 1433)	44.96* F(4, 16331)
$\sigma_{\eta i}\rho_{1i} = \sigma_{\eta j}\rho_{1j}$	8.12**		7.33**	
$\sigma_{\eta i}\rho_{2i} = \sigma_{\eta j}\rho_{2j}$	47.17*		32.60*	
$\sigma_{\eta i}\rho_{3i} = \sigma_{\eta j}\rho_{3j}$	29.93*		45.27*	
$\sigma_{\eta i}\rho_{4i} = \sigma_{\eta j}\rho_{4j}$	35.15*		67.16*	

* Indicates significant at the .01 level, ** significant at the .05 level; *** Significant at the .10 level.
Source: Authors' calculations using the data base of ENEU, 2004.

The *F*-test rejects the proportionality hypothesis discarding the possibility of equal impact of self-selection in the four choice occupations, confirming the non-randomly allocation of workers.

[19] In the first case, a rejection of the test will imply that any sector would have its own wage setting process, and for the second test, the existence of differences between workers into the same occupation but located in different sectors will prove that the proportionality hypothesis does not hold, and coefficients sigma-lambda will show that self-selection effectively determines the supply of the skills in the market by itself; in other words, individuals are free to choose where to work.

This output can be compared to Wu's (2010) results, where in rural China a link between employment choices among four labor categories was found. Using migrant earnings, a significant self-selection was computed where those workers with good health status, men, and the youths with higher schooling levels are more likely to migrate, having negative selection those workers related to non-farm work.

Earning's differential and decomposition.

In spite of being slightly higher the returns to education for the formal sector at the conditional mean, a corrected earning difference in support of the informal self employed workers has been computed for both years. As a result, the share of the gap is found to be mostly unexplained (62.7%) for 1992, meanwhile it reversed for 2004 (47.4%), that is, mostly due to productivity differences –non-discrimination component.

Table 4. Decomposition of corrected earnings differential in Mexico* between formal-informal sectors, 1992 and 2004.

Indicators	Wage earner		Self-employed	
	1992	2004	1992	2004
Formal earnings (log-mean)	1.165	1.121	0.915	0.717
Informal earning (log-mean)	0.988	0.889	0.998	0.919
Earning gap	0.177	0.232	-0.083	-0.202
- Attributes	0.165	0.088	-0.052	-0.096
- Absolute discrimination	0.012	0.144	-0.031	-0.107
- % explained	6.9	62.1	37.3	52.6
- Relative discrimination % unexplained	93.1	37.9	62.7	47.4
Total	100.0%	100.0%	100.0%	100.0 %

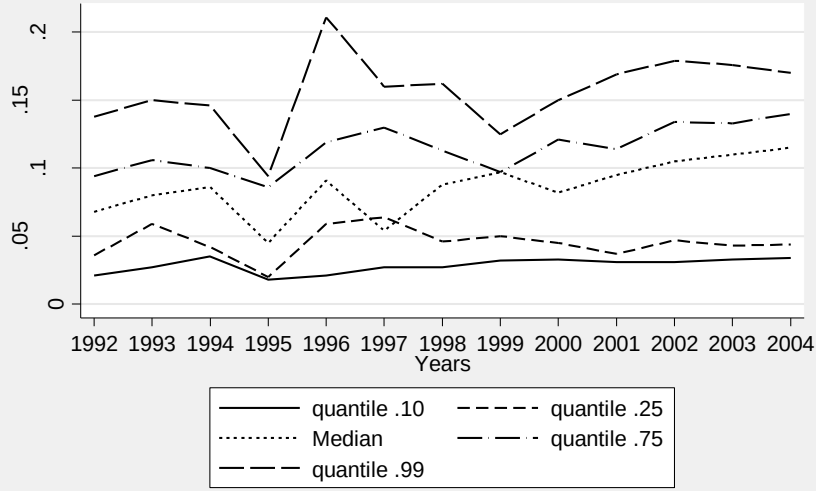
* Log of earnings corrected for selection bias using a polychotomous normal model.
Source: Authors' calculations using the data base of ENEU, 1992 and 2002.

Semi-parametric estimates by quantiles

This section comes across to quantify the earnings impact analyzing the situation with a distributional criteria in order to not only determine the changes on the

conditional mean, but also by selected quantiles in the distribution. The procedure is reproduced with the BFG technique using quantile regressions (Koenker and Bassett, 1978; and Poterba and Rueben, 1994) and for the formal-informal sectors considering the wage earning category in 1992 and 2004.[20] As expected, it is verified for both labor sectors (and wage earners) that education returns are smaller when the wage equation is conditioned towards the lower quantiles; when we select our equation towards the top, the yields are higher (see graphs 2 and 3). This event confirms the earnings premia of the human capital, although with remarkably differences regardless its sector of origin.[21]

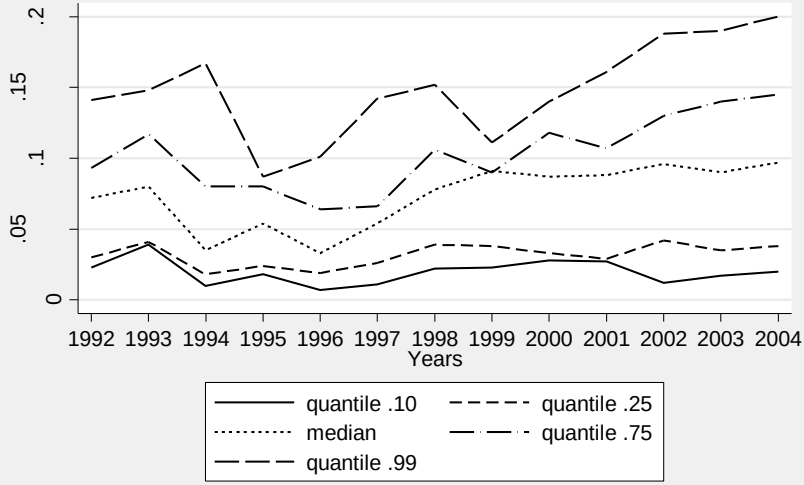
Graph 2.
Returns to education for the formal sector in Mexico 1992-2004.



Source: Author's calculations using ENEU respective years.

[20] The full set of estimates are not shown for space matters; however they are available upon request.
[21] Maloney (2002) describes that at time of crisis, the less qualified the workers, the more vulnerable they become despite the labor segment, a reason to think of deterioration within the least skilled workers.

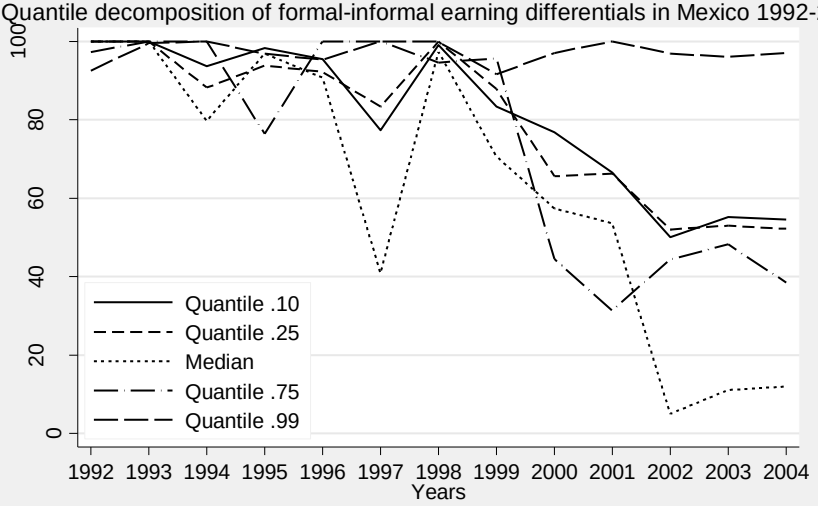
Graph 3.
Returns to education for the informal sector in Mexico 1992-2004.



Source: Author's calculations using ENEU respective years.

Formal sector seems to perform counter-cyclical, while informal sector tends to behave pro-cyclical. Departing from the Mexican crisis in 1995, growth in the Mexican economy plummeted and just one year later, returns to education grew for the formal sector at the majority of the quantiles; meanwhile, it can be seen for the informal sector that returns did not increase until the economy began to recover its growth path in 1999. The differential decomposition by quantiles of the wage earners is observed in graph 4.

Graph 4.
Quantile decomposition of formal-informal earning differentials in Mexico 1992-2004.



Source: Author's calculations using ENEU respective years.

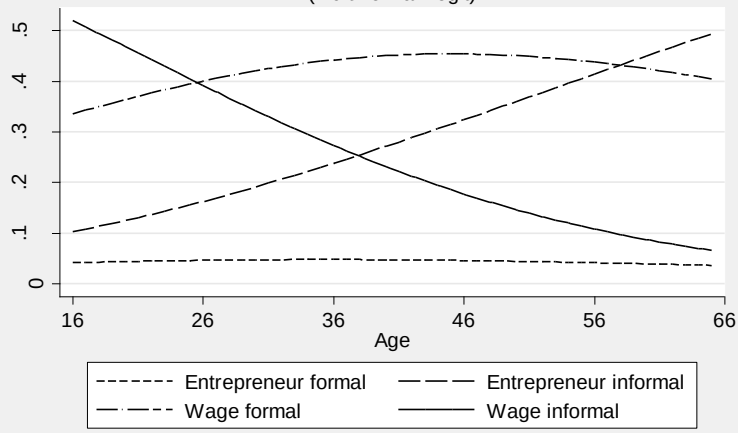
As Tannuri et al. (2004) assert, this indicates that people with more formal-like characteristics earn less than people with more self-employed-like characteristics, when they engage into informal jobs and are located at the top of the conditional earnings of the distribution. Unobservable characteristics are relevant for Mexico at any quantile; in contrast, unobservable characteristics in Bolivia are not relevant in the earnings specification, where individuals would earn more if they decide to move towards formal activities.

Occupational choice selection

An ultimate not less important topic of analysis in this research to wind up the empirical application, is to depict the occupational self-selection process of selection followed by the earners in the labor market. Using the MNL specification –from section 5–, predicted probabilities by segment with experience and schooling are depicted in Graphs 5 and 6.

Graph 5. Occupational selection and experience in Mexico, 2004.

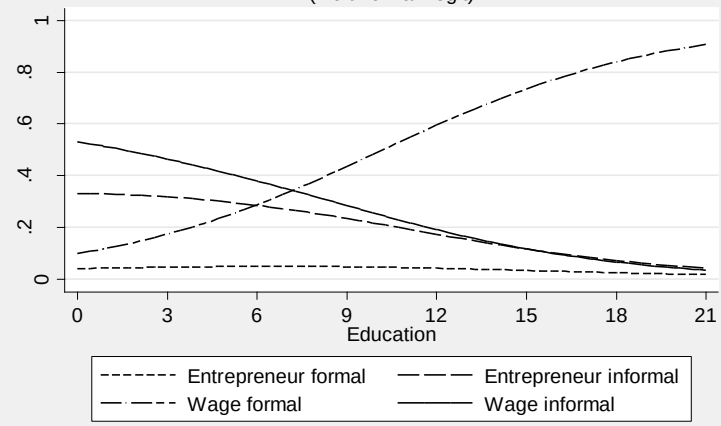
(Multinomial Logit)



Source: Author's calculations using ENEU, respective years.

Graph 6. Occupational selection and education in Mexico, 2004.

(Multinomial Logit)



Source: Author's calculations using ENEU, respective years.

VI. Conclusions

The attempt to formulate the final linkage of informal and formal labor markets is still incomplete, despite efforts made in the past. Different causes within countries, regions and domestic economic conditions are some of the reasons why formalizing activities do not always lead to success. Informal sector behavior is counter-cyclical, meanwhile for the formal is pro-cyclical. Evidence suggests, as one of its main causes, a lack of worthy and well remunerated jobs generation, as long as the risk of losing their formal occupation imposes the workers a reason to set them in motion to look for another alternatives, or to migrate out of the country, as the last choice.

Findings reveal that the informal sector is better paid as self-employed moved toward the top of the distribution of earnings. This represents a signalling that both segments of the labor market include a similar collective of workers, who are able to develop their job with a parallel productivity.

Neither the informal wage-earner nor the formal entrepreneurs, represent an employment choice for the highly qualified individual. The process limits this sort of workers to be hired as formal wage earners or as informal self-employed, where choices are not selected at random, that is, selection is voluntary.

From the pool of workers who selected the right choice are those in the informal self-employed occupations, had they decided to move to a different labor segment, then they would have lost income (the correction term was negative); on the other hand, informal wage earners presented a positive correction terms during the period.

For future research, it is interesting to replicate this methodology with longitudinal data, to compute mobility by interacting other relevant labor and institutional variables to obtain more robust estimates of the selection process among the sectors.

References

Acemoglu, Daron (2002). "Wage inequality and Technical Change", *Journal of Economic Literature*, Vol. 40, No. 1, pp. 7-72.

Azuma, Y. and H. Grossman, (2002). "A Theory of the Informal Sector" NBER, Working paper 8823.

Barceinas, F. and J. L. Raymond-Bara (2003). "Hipótesis de Señalización Versus Capital Humano. El Caso de México." *El Trimestre Económico*, vol. LXX(1), 277, (January-March): 109-94.

Bourguignon, Francois, M. Fournier and M. Gurgand (2001). "Selection Bias Correction Based on the Multinomial Logit Model" Working Paper no. 20023-04, CREST/INSEE, Paris, December.

Bourguignon, Francois, M. Fournier and M. Gurgand (2007). "Selection Bias Corrections Based on the Multinomial Logit Model: Monte-Carlo Comparisons." *Journal of Economic Surveys*, Vol. 21(1), 174-205.

Buchinsky, M. (1998). "Recent Advances in Quantile Regresssion Models: A Practical Guideline for Empirical Research." *Journal of Human Resources*, 33, (Winter) 88-126.

Cervantes, Niño, José Juan, E. Gutiérrez, G., and L. Palacios (2008) "El concepto del economía informal y su aplicación en México: factibilidad, inconvenientes y propuestas", *Estudios Demográficos y Urbanos*, El Colegio de México, Vol. 23, núm. 1 (67), pp. 21-54.

Charmes, J. (2000). "The Contribution of Informal Sector to GDP in Developing Countries: Assessment, Estimates, Methods, Organizations for the Future", paper presented at the 4th Meeting of the Expert Group on Informal Sector Statistics (Delhi Group), Geneva, August 28-30.

De Soto, H. (2000). The Mistery of Capital. Why Capitalism Triumphs in the West and Fails Everywhere Else. Bantam Press.

De Azevedo, J. P. (2004). "An Investigation of the Labour Market Earnings in Deprived Areas: Evaluating the Sources of Wage Inequality in the Slums",

Paper presented at the EGDI and UNU-WIDER Conference, *Unlocking Human Capital Potential: Linking the Informal and Formal Sectors*, Helnsinki, Finland, September 17-18.

Dickens, W. T. and K. Lang. (1985). "A Test of Dual Labor Market theory", *American Economic Review*, vol. 75, 792-805.

Dubin, J. A. and D. L. McFadden. (1984). "An Econometric Analysis of Residential Appliance Holdings and Consumption." *Econometrica*, 52 (March): 345-62.

Cohen, B. and W. J. House. (1996). "Labor Market Choices, Earnings, and Informal Networks in Khartoum, Sudan." *Economic Development and Cultural Change*, 44 (April): 589-618.

Fields, Gary (1975). "Rural-Urban Migration, urban Unemployment and Underemployment, and job search activity in LDCs." *Journal of Development Economics*, 2, 165-87.

Gong, X. and Arthur van Soest (2001). "Wage Differentials and Mobility in the Urban Labor Market: A Panel Data Analysis for Mexico", IZA-Discussion Paper, no. 329.

Hart, J. K. (1971). "Informal Income Opportunities and Urban Employment in Ghana." *Third World Employment: Problems and Strategy*. R. Jolly et al. (eds), Harmondsworth.

———— (1973). "Informal income opportunities and urban employment in Ghana" *Journal of Modern African Studies*, Vol. 11(1) 61-89.

Hanson, Gordon (2003). "What Has Happened to Wages in Mexico Since NAFTA? Implications for Hemispheric Free Trade." NBER Working Paper no. 9563.

Hausman, J. and D. McFadden (1984). "Specification tests for the multinomial logit model." *Econometrica 52* (5) 1219-40.

Heckman, James (1979). "Sample selection bias as a specification error", *Econometrica* Vol. 47(1) 153-61.

Heckman, James and G. Sedlacek (1985). "Heterogeneity, aggregation and market wage functions: An empirical model of self-selection in the labor market", *Journal of Political Economy*, 93, 1077-125.

Heckman, J. and V. J., Hotz (1986). "An Investigation of the Labor Market Earnings of Panamanian Males: Evaluating Sources of Inequality", *Journal of Human Resources*, 21(4), Fall, 509-42.

Huesca Luis y Mario Camberos, (2009), "El mercado laboral mexicano 1992 y 2002: Un análisis contrafactual de los cambios en la informalidad", *Economía Mexicana*, CIDE, Vol. XVIII, Núm. 1, primer semestre, pp. 5-43.

Huesca, Luis (2005). *La Distribución Salarial del Mercado de Trabajo en México: Un Análisis de la Informalidad*, PhD thesis, Departament d' Economia Aplicada, Universitat Autónoma de Barcelona.

_____ (2004). "La Rentabilidad de la Escolaridad en los Hogares Asalariados de México." *Problemas del Desarrollo*, July-September, 35, 138, 125-154.

Instituto Nacional de Estadística, *National Survey of Urban Employment* -ENEU- from 1992 to 2004.

——— (2014). DATA: Cuadro resumen indicadores de ocupación y empleo al cuarto trimestre de 2013, downloaded on Feb/25/14 from: http://inegi.org.mx. INEGI.

International Labor Office (1972). *Employment Incomes and Equality: A Strategy for Increasing Productive Employment in Kenya*. Geneva.

_____ (2001). *World Employment: Life at Work in the Informal Economy*, Geneva.

_____ (2002). *Decent Work and the Informal Economy*: *VI Report, International Labour Conference, 92nd. Session*. Geneva.

Jenkins, Stephen P. and P. Van Kerm (2004). "Accounting for Income Distribution Trends: A Density Function Decomposition Approach" IZA-Discussion Paper Series no. 1141.

Jusidman, Clara (1993). "The Informal Sector in Mexico", *Secretaría del Trabajo y Previsión Social and the Uniter States Department of Labor*, Working Paper No. 2.

Koenker, R. and G. Bassett (1978). "Regresión quantiles." *Econometrica*, 46, 33-50.

La Porta and Shleifer (2014). "Informality and Development", *Journal of Economic Perspectives*, 28(3), 109-126.

Lee, L. F. (1983). "Generalized Econometric Models with Selectivity." *Econometrica*, 51, 507-12.

Leontaridi, R. (1998). "Segmented Labour Markets: Theory and Evidence", *Journal of Economic Surveys*, Vol. 12(1).

Levy, S. (2010). *Buenas intenciones, malos resultados: política social, informalidad y crecimiento económico de México*. Océano.

Magnac, T. (1991). "Segmented or Competitive Labor Markets." *Econometrica*, 59, 165-87.

Maloney, William (2003). "Informality Revisited." *World Development*, 32 (7) 1159–78.

———— (2002). "Distortion and Protection in the Mexican Labor Market." *Working Paper* 138, The World Bank, Center for Research on Economic Development and Policy Reform.

Marcouiller, D. V., R. de Castilla and C. Woodruff (1997). "Formal Measures of the Informal Sector Wage Gap in Mexico, El Salvador, and Peru." *Economic Development and Cultural Change*, 45 (January): 367-92.

Mazumdar, D. (1983). "Segmented Labor Markets in LDC's" *American Economic Review*, 73, 2, 254-59.

McFadden, D. L. (1973). "Conditonal Logit Analysis of Qualitative Choice Behavior." *Frontiers in Econometrics*, Academic Press.

McNabb, R. and P. Ryan. (1990). "Segmented Labour Markets." *Current Issues in Labour Economics*, Sapsford, D. y Z. Tzannatos, Mcmillan Press.

Meza, Liliana (2001). "Wage Inequality and the Gender Wage Gap in Mexico." *Economía Mexicana*, 2, 291-323.

Mincer, Jacobs (1974). *Schooling, Experience and Earnings*, Columbia University Press.

Moreno-Treviño, Jorge O. (2007). "Los salarios del sector formal e informal en México: Análisis de ganancias y pérdidas por formalización" *Ensayos Revista de Economía*, Vol. 26(1), 1-44.

Oaxaca, Ronald (1973). "Male-female differentials in urban labor markets." *International Economic Review*, 14(3), 693-709.

Habitat, U.N. (2006). Case Study: Mexico City, in: Innovative Policies for the Urban Informal Economy. *UN-HABITAT:* Nairobi.

Poterba, J. and K. Rueben (1994). "The Distribution of Public Sector Wage Premia: Evidence Using Quantile regression Methods." NBER Working Paper 4734.

Roubaud, Francois (1995). *La economía informal en México*, Fondo de Cultura Económica.

Saavedra, J. and A. Chong (1999). "Structural reform, institutions and earnings: Evidence from the formal and informal sectors in urban Peru". *Journal of Development Studies 35* (4), 95-116

Sethuraman, S. V. (1997). "Urban Poverty and the Informal Sector: A Critical Assessment of Current Strategies." *Geneva: ILO.*

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, Chapman and Hall, London.

Tannuri-Pianto, M., D. Pianto and O. Arias (2004). "Informal Employment in Bolivia: A Lost Proposition?" Econometric Society 2004 Latin American Meetings 149, Econometric Society.

Taylor, Mark, P. (1996). "Earnings, Independence or Unemployment: Why Become Self-Employed?." *Oxford Bulletin of Economics and Statistics*, 58(2) 253-66.

Tokman, V. E. (1994). "Informalidad y Pobreza: Progreso social y modernización productiva." *El Trimestre Económico*, 241 (January-March): 177-99.

Van Der Sluis J., M. Van Praag and W. Vijverberg (2004). "Entrepreneurship Selection and Performance: A Meta-analysis of the Impact of Education in Less Developed Countries", Tinbergen Institute Discussion Paper, TI 2003-046/3.

Van Kerm, Phillip. (2003). “Adaptive kernel density estimation.” *Stata Journal*, 3(2) 148-56.

Zhang, H. (2004). “Self-selection and Wage differentials in Urban China: A Polychotomous Model with selectivity.” Department of Urban Studies and Planning. (August), Massachusetts Institute of Technology, Cambridge, Massachusetts.

Wu, Z. (2010). Self-selection and Earnings of Migrants: Evidence from Rural China. *Asian Economic Journal*, *24*(1), 23-44.

Appendix

Table A1. Selection variables

Education	From 1 to 21 years
Age	16 to 65
Labor status:	
a) Formal self employed:	Entrepreneurs which does have official registry and pay social security contributions = 1; otherwise = 0
b) Informal self employed:	Entrepreneurs which does not have official registry and does not pay social security contributions = 1; otherwise = 0
c) Formal wage earner:	Workers who does receive social security benefits = 1; otherwise = 0
d) Informal wage earner:	Workers who does not receive social security benefits = 1; otherwise = 0
Working time:	At least 20 hours or more per week
Household head	= 1; otherwise = 0
Tradable sector and non-tradable sector of goods (See Table A2): Variable "prodtype":	
a) Tradable:	Reclasification of activity branch 1 and 4 = 1
b) Non-tradables:	Reclasification of activity branches 2, 3, 5 and 6 = 0.
Family structure: Variable "dmembers"	
Worker in a household with number of members higher than the national mean	= 1; otherwise 0

Source: Author's elaboration based on the Mexican current classification of economic activities (CAE-94).

Table A2. Summary of statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
1992					
Formal self employed	46789	0.028	0.164	0	1
Informal self employed	46789	0.238	0.426	0	1
Formal wage-earner	46789	0.555	0.497	0	1
Informal wage-earner	46789	0.180	0.384	0	1
Monthly wage (USD)	46789	511.385	479.229	102.324	18,479.79
Education	46789	6.661	3.084	1	12
Age	46789	33.755	11.452	16	65
Age2	46789	1270.555	863.513	256	4225
Head	46789	0.670	0.470	0	1
dmembers	46789	0.373	0.484	0	1
Prodtype	46789	0.791	0.407	0	1
dsec1	46789	0.266	0.442	0	1
dsec2	46789	0.109	0.311	0	1
dsec3	46789	0.193	0.395	0	1
dsec4	46789	0.332	0.471	0	1
dsec5	46789	0.086	0.280	0	1
dsec6	46789	0.015	0.121	0	1
Dregion1	46789	0.360	0.480	0	1
Dregion2	46789	0.104	0.305	0	1
Dregion3	46789	0.313	0.464	0	1
dregion4	46789	0.125	0.330	0	1
dregion5	46789	0.037	0.189	0	1
dregion6	46789	0.062	0.241	0	1
2004					
Formal self employed	69311	0.021	0.142	0	1
Informal self employed	69311	0.229	0.420	0	1
Formal wage-earner	69311	0.590	0.492	0	1
Informal wage-earner	69311	0.160	0.367	0	1
Monthly wage (USD)	69311	525.035	485.281	95.415	14342.590
Education	69311	10.349	4.382	1	21

Age	69311	35.141	11.179	16	65
Age2	69311	1359.883	854.544	256	4225
Head	69311	0.701	0.458	0	1
dmembers	69311	0.617	0.486	0	1
prodtype	69311	0.558	0.497	0	1
dsec1	69311	0.261	0.439	0	1
dsec2	69311	0.095	0.293	0	1
dsec3	69311	0.188	0.391	0	1
dsec4	69311	0.101	0.302	0	1
dsec5	69311	0.333	0.471	0	1
dsec6	69311	0.009	0.095	0	1
dregion1	69311	0.353	0.478	0	1
dregion2	69311	0.140	0.347	0	1
dregion3	69311	0.317	0.465	0	1
dregion4	69311	0.062	0.240	0	1
dregion5	69311	0.044	0.205	0	1
dregion6	69311	0.084	0.277	0	1

Source: Authors' elaboration based on ENEU, respective years.

Table A3. Earning equations by OLS of the formal-informal sectors
(Estimates corrected by BFG technique –second stage-).

Variables	1992				2004			
	Wage earner		Self-employed		Wage earner		Self-employed	
	Formal	Informal	Formal	Informal	Formal	informal	Formal	Informal
Education	0.075**	0.059**	0.050***	0.046**	0.110*	0.087*	0.096**	0.107*
Age	0.116*	0.067*	0.063**	0.039*	0.106*	0.077*	0.014*	0.032*
Age2 x 100	-0.123*	-0.064*	-0.051*	-0.043*	-0.133*	-0.065*	-0.020*	-0.036*
Head	0.093*	0.086**	-0.032**	0.053**	0.154*	0.089**	0.078**	0.081*
dmembers	-0.037*	-0.015*	0.040**	0.026*	-0.016*	-0.048*	-0.001**	-0.007*
Prodtype	0.119**	0.049**	0.139	0.051**	0.110**	-0.109**	-0.019	-0.134**
dsec1	0.002*	-0.070**	-0.001***	-0.116**	0.043**	-0.426***	-0.050**	0.001**
dsec2	0.046**	0.031**	0.312	0.098**	0.297**	0.160**	0.149	-0.011**
dsec3	-0.094*	-0.081**	0.009**	-0.075**	-0.197*	-0.068**	-0.034**	0.124**
dsec5	0.008**	0.603**	0.483	0.555***	0.142**	0.103**	-0.018***	0.037**
dregion1	0.765**	0.522***	0.531	0.656***	0.748**	0.688***	0.663**	1.008***
dregion2	0.220**	0.140**	0.243	0.160**	0.478**	0.281**	0.316**	0.565**
dregion3	0.111**	0.167**	0.196***	0.076**	0.313*	0.223**	0.197**	0.308**
dregion4	0.488**	0.303**	0.310	0.350**	0.383**	0.244**	0.282**	0.399**
dregion6	0.669**	0.459***	0.448	0.392***	0.625**	0.263***	0.297**	0.601***
$(\sigma_\eta \rho_1)$	0.387	0.311	0.662	0.010	-0.987**	0.806***	0.237	-0.327
$(\sigma_\eta \rho_2)$	-1.597**	1.73**	-2.588***	0.754**	0.634**	0.995**	0.116	-0.803**
$(\sigma_\eta \rho_3)$	-9.127***	4.962***	5.770***	-6.833***	-4.039***	2.826**	-2.046	-5.342***
$(\sigma_\eta \rho_4)$	-0.613***	-0.059	-0.399	-0.694**	-0.749**	-1.609**	-0.527	-0.530**
Constant	-0.119	0.350	1.121	1.521	0.361	1.678	0.420	1.079
$\bar{R}_2$	0.372	0.265	0.092	0.113	0.396	0.249	0.136	0.163
$F(19, N)$	815.94	166.60	8.04	82.2	1,566	224.6	12.92	167.98
Obs.	25,942	8,414	1,309	11,125	40,893	11,090	1,456	15,872
F-Test[a]: $(\sigma_\eta \rho_{ij} = 0)$	122.03	89.11	21.12	42.36	187.52	90.80	16.79	98.11

[a] Note: Test statistic is approximately $F_{(4,\ N-19)}$ distributed. Standard error at significance levels: *: 1% **: 5% ***: 10%.
Source: Authors' calculations using the data base of ENEU, 1992 and 2004.