I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14278

By Choice or by Force? Exploring the Nature of Informal Employment in Urban Mexico

Robert Duval-Hernández

APRIL 2021

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14278

By Choice or by Force? Exploring the Nature of Informal Employment in Urban Mexico

Robert Duval-Hernández
Open University of Cyprus, CIDE and IZA

APRIL 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

By Choice or by Force? Exploring the Nature of Informal Employment in Urban Mexico*

Using a special module of the 2015 Mexican Labour Force Survey with information on workers' preferences for jobs with social security coverage, I estimate that 80 per cent of informal workers in large urban areas would prefer to work in a job that provides them with such coverage. The estimation of a discrete choice econometric model which distinguishes between wanting a formal job and the probability of getting one shows that schooling increases the chances of being hired in formal employment and of having higher earnings in it. Women with greater responsibilities at home are less likely to want formal employment, and they also face a lower probability of being hired in such jobs. The findings indicate the segmentation of Mexican labour markets and the rationing of formal jobs, together with the existence of workers who voluntarily participate in informal employment. However, the estimated fraction of involuntary informal workers is quite high.

JEL Classification: O17

Keywords: informal employment, labour markets, segmentation, rationing

Corresponding author:
Robert Duval-Hernández
Faculty of Economics and Management
Open University of Cyprus
PO Box 12794
CY-2252 Latsia
Cyprus
E-mail: robert.duval@ouc.ac.cy

* This paper is part of the project 'Transforming informal work and livelihoods' funded by the United Nations University World Institute for Development Economics Research (UNUWIDER). An extended version of this paper has appeared as a WIDER Working Paper #2020/151. I would like to thank Gary S. Fields, Tim Gindling, Kunal Sen, seminar participants at the workshops organised by the project, as well as two anonymous project referees for their valuable comments. The opinions and views here expressed are my own and do not represent those of my employer or of UNU-WIDER.

1. Introduction

A central question in the study of labour markets in the developing world is whether its large share of informal employment is mainly composed of workers who cannot find a better employment option in the formal economy, or whether this sector is formed by workers who voluntarily seek this type of employment given the incentives in the market.

Broadly speaking, traditional theories of dualistic labour markets conceive informal employment as a second-best option for workers who cannot find a formal job, and who cannot afford to remain unemployed while searching for a good employment option (see, for instance, the seminal model of Fields, 1975). In contrast, a different strand of the literature has argued that workers optimally choose informal employment because, given their individual characteristics, they obtain a higher utility/payoff in that type of work (see, for instance, Maloney, 1999). This can arise when workers prefer to be informally employed because they will pay lower taxes, face less regulations, or have more flexible contract arrangements than if they were formally employed.

Despite the above opposing views, the literature recognises that, in practice, informal employment is heterogeneous. It is composed of workers who participate in it voluntarily and those who are there because they cannot find a formal job (Fields, 1990; Maloney, 2004). While there is agreement about the heterogeneous nature of informal employment, little is known about how many informal workers are voluntarily in informal employment and how many simply cannot find a better job elsewhere. Little is also known about the characteristics of these different types of workers.

This gap in our knowledge arises because, in most of the available survey data, workers are not asked about their preferred type of employment and, for this reason, researchers have devoted a substantial amount of energy in testing for the existence of rationing in labour markets through: i) the estimation of complex structural econometric models of sector allocation (e.g., Günther & Launov, 2012; Magnac, 1991); ii) the analysis of sector transitions over the business cycle (e.g., Bosch & Maloney, 2010; Fiess, Fugazza, & Maloney, 2010); and iii) the use of ancillary information about the characteristics of informal employment in order to identify those workers who perform successfully in it and, thus, are more likely to participate in it voluntarily (e.g. Gindling & Newhouse, 2014).

The contribution of this paper is to use a special supplement of the 2015 Mexican Labour Force Survey (Instituto Nacional de Estadística y Geografía [INEGI], 2015a) which directly asks urban workers about their valuation of jobs with social security coverage. Since, in Mexico, having social security coverage in a job is the defining characteristic of formal wage employment, this piece of information gives us a proxy measure of workers' preferences for this type of jobs.

This variable is used to address important questions in the informality literature, such as: What fraction of informal workers would rather be formal employees? What individual characteristics increase the probability of applying to and of being hired in a formal job? and How earnings are related to individual characteristics, after accounting for the non-random selection into each type of employment?

Our analysis indicates that around 80 per cent of the respondents who lack social security coverage would prefer to have a job with such benefits, even if that entailed having to pay the corresponding contributions for them. Although one cannot determine with certainty whether this reflects a preference for social security benefits alone or, more generally, a preference for the entire set of characteristics that accompany a formal job, the figure indicates that a large fraction of the urban informal workers in Mexico are not voluntarily so.

A discrete choice model that distinguishes the worker's decision to apply for a formal job from the formal sector employer's hiring decision is estimated. The results of this model reveal that women living in households with a higher fraction of dependents are less likely to want formal employment, likely because a heavier burden of household duties requires jobs with flexible schedules, such as the informal ones. In fact, this negative effect of dependents on the probability of applying for formal jobs disappears once other adult females (who can help with care activities at home) are present in the household. For males, the opposite effect is observed. Namely, a higher fraction of dependents is positively associated with the probability of males applying to formal jobs. This finding is consistent with a more traditional role of men as the main breadwinners in the household.

The number of years of schooling significantly increases the probability of being hired in a formal job, but workers still enrolled in school face a substantial hiring penalty. Finally, marital status affects the probability of being hired differently for men and women. Other things equal, married men are more likely to be hired in formal employment compared to single males, while the exact opposite occurs for married females. Again, given the unequal division of labour at home in Mexico, this reflects that formal sector employers prefer workers who have fewer household responsibilities.

Selectivity-adjusted earnings equations are estimated for three different types of workers: formal, voluntary informal, and other informal workers. The estimations show a large gender wage premium, and married women who voluntarily work in an informal job face large earnings penalties compared to single females. The returns to education for formal workers are around 15 per cent, and between 4 per cent and 8 per cent for the voluntary informal. The earnings of involuntary informal workers show no statistically significant relation to education.

Overall, the empirical findings confirm the view that informal employment is formed of a heterogeneous group of workers, some of whom participate voluntarily in it while others do so because of a lack of better options. However, contrary to previous suggestions in the literature, the fraction of involuntary informal workers is quite high.

Broadly speaking, the findings of this paper highlight two main factors that limit the number of workers employed in formal jobs. The first is related to household demographics and the division of housework. In particular, women who have a higher burden of work at home are less likely to seek formal employment and are less likely to be hired in formal jobs. The second is related to human capital, as higher levels of schooling increase the chances of being hired in formal jobs and of obtaining higher earnings in them.

While the econometric estimations cannot be used to predict the consequences of policy changes that occur at the aggregate level, they nevertheless indicate some of the dimensions that need to be considered when designing reforms that seek to encourage the growth of better-paying formal employment. Also, this study highlights the usefulness of incorporating direct information about the valuation of different types of jobs in the study of labour markets in developing countries.

The paper is organised as follows. Section 2 presents the data on the valuation of jobs with social security benefits, together with its main descriptive statistics. Section 3 discusses the econometric methodology for the *ceteris paribus* analysis of job allocation and earnings, while the results of these estimations are presented in Section 4. Section 5 concludes.

2. Informal employment and the valuation of social security coverage

In Mexico, having social security coverage is the defining characteristic of formal wage employment (see for instance Levy, 2008 and INEGI, 2014). For this reason, the analysis in this paper equates ‘formal employment’ with having a job that is covered by social security benefits.[1]

In the second quarter of 2015, the Mexican Labour Force Survey was supplemented by a module inquiring about the employment trajectories of workers and their contribution to and valuation of social security protection.[2] The MOTRAL module (after its acronym in Spanish) was applied to a representative sample of workers aged 18 to 54, living in large urban centres who were either employed or had previous labour market experience. This target population represented around 90 per cent of the labour force in large urban centres and 60 per cent of the overall urban labour force in 2015.[3]

This module included the following key question: Do you think it is better to have a job with social security, even if you have to make payments to be eligible for it?.[4] This question is central to the study of informality because most labour surveys contain information on the sector of employment, but they do not collect information on the types of jobs workers value.

---

[1] Alternatively, one could define the formality status for the self-employed based on whether a business is registered with tax authorities and has fixed work premises (INEGI, 2014). A robustness check that considers this criterion of formality for the self-employed is presented in section 4.3.

[2] The supplementary module is the Módulo de Trayectorias Laborales, 2015 (MOTRAL) and its data can be publicly accessed online (see INEGI 2015a). A similar module was also applied in 2012, but in that edition the key variable used in the analysis that follows was not included.

[3] The individuals interviewed in the module also answered the questions in the Labour Force Survey, and the two datasets can therefore be linked, as is done here, in order to have a richer set of variables.

[4] The original question reads: ‘¿Considera que es mejor tener un empleo con seguridad social, aunque tenga que realizar pagos para tener derecho a ella?’.

To the extent that social security coverage is the defining characteristic of formal wage employment in Mexico, the answer to the question on the preference for a covered job can be used as a proxy for the workers' preferred type of job (formal or informal). Linking this variable to information on the *actual* type of job can help us to approximate the fraction of involuntary informal workers. Without this piece of information researchers have no choice but to try to infer through indirect methods what fraction of the informal workforce is so because of a lack of options rather than by choice.

2.1 Descriptive Statistics

Table 1 presents the key sociodemographic characteristics of employed workers in the MOTRAL module depending on whether they have social security coverage in their job and on whether they would prefer to have a job with coverage.[5] The table shows that the age and gender composition is more or less homogeneous across groups except for workers with coverage who do not value their social security benefits (in column 2), who are predominantly male. Workers with coverage are more educated than those without coverage, and within each market segment (with coverage or without coverage) respondents who do not value social security have higher levels of schooling than those who value it. In addition, workers without coverage are more likely to still be enrolled in school. Also, workers without coverage who do not value social security benefits are less likely to be married, have a higher number of dependents at home, and have the lowest earnings of all groups. In contrast, the group that exhibits higher average earnings is the workers with coverage who do not value social security coverage.

---

[5] The Supplementary Material that accompanies this paper presents an additional set of descriptive statistics for the samples used in the estimation of econometric models.

2.2 Voluntary versus involuntary informal employment

In this paper, I assume that all workers employed in a formal (covered) job prefer their formal employment status to an uncovered status, irrespective of their valuation of social security benefits *per se*. This entails assuming that the 8.9 per cent of the employed population who has a covered job but do not value their social security benefits (see column 2 in Table 1), still prefer formal to informal employment, as otherwise they would voluntarily move to an uncovered informal job. In other words, I assume they remain in formal employment because of other job qualities such as higher wages, greater job security, and in general better working conditions.

In the case of informal (uncovered) workers, I assume that those who consider it better to have a job with social security coverage are ‘involuntarily employed’ in their current job, while those who do not value covered jobs are classified as ‘voluntary informal’. This means that I interpret their answer to the question on the valuation of social security benefits, as reflecting an overall preference for formal employment.[6]

Given the above classification of who is a voluntary and who an involuntary informal worker one can estimate, based on the figures reported at the bottom of Table 1, that around 80 per cent of informal workers are involuntarily employed.[7] This high proportion of involuntary informal workers contrasts with the view put forward by a

---

[6] In Duval-Hernández (2020) I discuss at length the different interpretations one can give to the answers to the valuation of coverage question, and based on the descriptive evidence in the module, as well as in other ancillary datasets, I argue in favour of the classification here adopted.

[7] This estimate is obtained from the figures reported in columns 3 and 4 of the table. There, it is reported that 38.3 per cent of the sample are involuntary informal, and 9.3 per cent are voluntary informal. Therefore, of the total informal population (which represent 47.6 per cent of the sample), 80 per cent are involuntary informal (i.e., 38.3/47.6).

strand of the informality literature which considers Mexican informal labour markets to be mainly composed of voluntary workers (see, for instance, Bosch and Maloney, 2010; Maloney, 1999, 2004).

This literature reaches this conclusion mainly based on an analysis of the patterns of sector transitions over the business cycle rather than from the direct measurement of workers' stated preferences, as this paper does. While interesting on their own, sector transitions over the business cycle only provide indirect evidence about the preferred employment status of workers changing jobs and, by construction, do not tell us anything about the preferred jobs of stayers.[8]

Other studies have tried to estimate the proportion of involuntary informal workers in Mexico using structural econometric methods. Two examples are the papers by Duval-Hernandez & Smith (2010) and Alcaraz, Chiquiar & Saucedo (2015). Both use discrete choice models that allow for rationing of formal jobs under a context of partial observability because they lack information on the preferred sector of workers.[9] The estimates of the proportion of involuntary informal workers vary widely between

---

[8] The results also contrast with the argument that many workers find the benefits associated with social security not worth the taxes that have to be paid to obtain them (Levy, 2008). While it is possible that some informal workers do not value such benefits *given their current employment*, the answers in the module here analysed suggests that a large proportion of uncovered workers would prefer to be employed as wage employees in a formal job with better pay and better working conditions.

[9] Other papers have tried to test segmentation by comparing formal and informal wage equations after correcting for self-selection into each sector. However, the methods used to correct for sample selectivity are often based on sector choice models that assume free entry into formal employment. This is problematic as free choice among sectors is precisely the issue these papers try to test (see, for instance, Marcouiller, Ruiz de Castilla and Woodruff, 1997 for the case of Mexico).

these studies, and sometimes even within a given paper, depending on the sample analysed. This highlights the potential danger of trying to infer the proportion of involuntary informal workers using indirect methods.[10]

The next section presents the econometric model used to exploit the information on the preferred type of job, with the goal of identifying the sociodemographic factors associated with applying for and being hired in formal employment.

3. Econometric methodology

As previously emphasised, this paper's main contribution is its analysis of a dataset that contains information about the preferred type of job for a representative sample of urban workers. By having information that distinguishes between the desired and the actual types of job, one can estimate econometric models of sector assignment that disentangle application from hiring decisions for formal jobs.

To set the notation, V_i^a denotes the utility of worker *i* of applying and being employed in a formal job, and V_i^h denotes the corresponding propensity of a formal sector employer of hiring this worker i. Let us assume that such propensities depend on vectors of observable characteristics of the worker, (Z_i, X_i), as well as on a set of unobservables, (u_{ai}, u_{hi}).[11] In particular, we assume that these components are related by the following system of equations:

$$V^a = Z\gamma_a + u_a \tag{1.1}$$

[10] Few studies have analysed direct survey questions on the valuation of formal employment in other Latin-American countries (see for instance Soares, 2004 for Brazil in 1990 and Contreras, Gillmore & Puentes, 2017 for Chile in 2009). Yet, this type of studies remains the exception in the broader informality literature.

[11] In this context, 'unobservable' means characteristics not available to the econometrician.

$$V^h = X\gamma_h + u_h \tag{1.2}$$

where (γ_a, γ_h) is a pair of vectors of unknown parameters.[12] To estimate the parameters from this model, we assume the unobservables (u_a, u_h) follow a standard bivariate normal distribution with (unknown) correlation parameter ρ.

The vectors of individual observable characteristics, (Z, X) need not be the same across equations, and they will only include characteristics of the worker. Ideally, one would like to include characteristics of the various potential employers with whom a worker might be matched, but this information is not available in the household survey here analysed. Another piece of missing information is whether individuals who do not apply for a formal job would be hired in that position if they were to apply.[13]

In practice, for any given worker, there are three possible scenarios that can be distinguished in the data: i) being a formal worker, which occurs with probability $P(V^a > 0, V^h > 0|Z, X)$; ii) being an involuntary informal worker, which occurs with probability $P(V^a > 0, V^h \leq 0|Z, X)$; and iii) being a voluntary informal worker, which occurs with probability $P(V^a \leq 0|Z, X)$. These three scenarios are incorporated into a discrete choice model that captures the joint decisions of workers and formal sector employers.[14] The likelihood function of the discrete choice problem is

[12] In the above equations and the ones that follow, I drop the individual subscript *i*.

[13] It is conceptually possible that voluntary informal workers could be offered a formal job if they were to apply for that job. However, I do not allow for that possibility in the econometric model estimated.

[14] This model is sometimes called a bivariate probit with 'sample selection' (see for instance Van de Ven and Van Praag, 1981).

$$\mathcal{L} = \prod_{i\in\mathcal{V}}[1 - F(Z\gamma_a)]\prod_{i\in\mathcal{J}}G(Z\gamma_a, -X\gamma_h; -\rho)\prod_{i\in\mathcal{F}}G(Z\gamma_a, X\gamma_h; \rho) \tag{2}$$

where $F(\cdot)$ and $G(\cdot.,\cdot;\cdot)$ are the standardised normal and bivariate normal distributions, respectively, $\mathcal{V}$ is the set of voluntary informal workers, $\mathcal{J}$ is the set of involuntary informal workers, and $\mathcal{F}$ is the set of formal workers. This model can be estimated via maximum likelihood. The estimates of the model provide a reduced form *ceteris paribus* answer to the question of which individual characteristics are associated with applying for and being hired in a formal job.[15]

The discrete choice model can also be extended to estimate, in a second stage, selectivity-corrected earnings equations for different types of workers through a switching-regression model that allows for sector allocation based on both the application and hiring decisions modelled in equations (1.1) and (1.2). In this case, it will be assumed that the error terms in the log-earnings equations and the unobservables (u_a, u_h) defined above are jointly distributed with a multivariate normal distribution. The selectivity-adjusted log-earnings functions will be

$$\log y_s = X_s\beta_s + \theta_{sa}\lambda_a(Z, X) + \theta_{sh}\lambda_h(Z, X) + \epsilon_s, \tag{3}$$

where the subscript *s* refers to the three different groups of workers characterised previously, i.e., formal, involuntary informal, and voluntary informal workers. The

[15] The model is a reduced form one because it does not explicitly incorporate wages in the choice decisions equations.

terms $\lambda_a(Z,X)$ and $\lambda_h(Z,X)$ are selectivity correction terms that adjust for the fact that individuals are not randomly assigned across sectors.[16]

Through these selectivity-adjusted earnings equations, one can also obtain a prediction of the counterfactual earnings that an informal worker would obtain if working in a formal job. This exercise can be performed separately for voluntary and involuntary informal workers.

Key demographic characteristics are included in the vector of observables Z that affect the probability of applying for formal employment. In particular, Z includes age (and its square), marital status, as well as the household dependency ratio, and the interaction of this dependency ratio with the number of adult females in the household. As previously mentioned, more dependents will need more hours of care and a higher income to sustain them. Since females typically do most of the housework, this variable is expected to negatively affect their probability of applying for a formal job as these jobs are less flexible in their schedules. However, as the number of adult females increases within a household, the housework load per woman will be smaller, hence the need to interact these two variables. The effects of these demographic variables are estimated separately by gender through the corresponding interactions with a gender dummy variable.

The vector Z also includes the respondents' years of schooling and a dummy indicating whether they are still enrolled in school. Finally, a set of dummy variables at the city level are included to control for varying conditions in local labour markets.

---

[16] In the case of the earnings equation of voluntary informal workers, the term $\theta_{sh}\lambda_h(Z,X)$ is zero. A detailed presentation of the model is included in the Supplementary Material. A more detailed exposition, including the formulas for the standard errors of the parameters, can be found in Tunalı (1986).

The variables included in vector *X*, which enters the hiring equation (1.2), are a subset of Z, and only include variables that would be observable by a formal sector employer, such as age, gender, marital status, the aforementioned schooling variables, and city-level dummies. In other words, the demographic characteristics of the household (i.e. the dependency ratio and its interaction with the number of adult females) are excluded, as these characteristics are typically not observable by employers and, thus are less likely to affect their hiring decisions. Finally, all three log-earnings equations (3) include age (and its square), years of tenure in the current job (and its square), gender and marital status (interacted), the above schooling variables, and city-level dummies.

It is important to note that neither in vector Z nor in X I include information about firm characteristics or the current occupation, as these characteristics occur *after* the sector selection process has taken place, and thus are already an endogenous outcome of the job allocation process.

4. Results

This section presents the results of the estimation of the models described above, beginning with the results pertaining to the discrete choice model.

4.1 Discrete choice model

The results of the parameter estimates of equations (1.1) and (1.2) are included in Table S4 in the Supplementary Material. Instead, Table 2 presents the average partial effects of the model as these are easier to interpret, i.e., I report the average derivatives of a probability of interest (e.g., the probability of applying for a formal job) with respect to one explanatory variable, holding all other observable variables constant.

The first two columns of Table 2 report the result of a standard probit which does not separate the applying from the hiring decisions. The last four columns report the parameter estimates (and their standard errors) of the discrete choice model in equations (1.1) and (1.2). Columns 3 and 4 contain the corresponding partial effects of the ‘apply’ equation (1.1), while the last pair of columns presents the partial effects of the ‘hiring’ decision *conditional* on having applied for a formal job position, i.e. the average partial effects of covariates on the conditional probability:

$$\frac{P(V^a > 0, V^h > 0 | \, Z, X)}{P(V^a > 0 | \, Z, X)}.$$

Analysing first the partial effects corresponding to the decision to apply to a formal job, columns 3 and 4 of Table 2 show a concave relation between the propensity to apply and age for females. For males, however, the partial effect of age is negative among young workers and statistically insignificant otherwise. Married workers are about 4 percentage points more likely to apply for formal jobs, although the effect is only statistically significant for males. While several of the partial effects differ by gender, on average males are equally likely as females to apply to formal jobs. Also, the education variables are not significantly associated with the applying decision.

A marginal increase in the dependency ratio decreases for women the probability of applying for formal employment by almost 5 percentage points, but this negative effect disappears if there is another adult female present in the household, presumably because that other female will help with the care of the dependents. In contrast, the opposite effect is found for males, i.e., there is a positive association between this ratio and the propensity to apply. These findings support the idea that the division of responsibilities at home has an important influence on whether workers want formal jobs, but the effects differ depending on the gender of the worker. More specifically, having more dependents and no extra help leads women to search for more flexible informal jobs, as these will allow them to care for dependents. For men, in contrast, having more dependents increases the propensity to apply for formal jobs.

Regarding the probability of being hired in a formal job, conditional on having applied, this probability presents an inverted-u association with age for females and a u-shaped association for males. Married men are almost 12 percentage points more likely to be hired in formal employment than single males, while the opposite is the case for married females. This hiring penalty for married females likely reflects that formal sector employers perceive them as less attractive employees, a finding that is consistent with recent experimental evidence for Mexico (Arceo-Gomez and Campos-Vazquez, 2014). These findings, together with those pertaining to the applying decision, highlight the importance of the division of labour at the household level in shaping both applying and hiring decisions in the formal sector.

Finally, education is a key factor affecting the probability of being hired. In particular, an extra year of schooling increases this probability by about 3 percentage points, while being enrolled in school decreases this hiring probability by almost 25 percentage points.[17]

Note also that the partial effects of the standard probit model (in the first two columns of Table 2) are a mix of the effects identified by the bivariate probit model, which separates applying from hiring decisions. Without the bivariate probit though, it

[17] Supplementary Material Table S4 reports the correlation parameter between the unobservables (u_a, u_h), which is -0.53 (and statistically significant at the 1 per cent level), indicating that the unobservable factors that lead to a greater probability of applying for formal employment are negatively correlated with the unobservables that affect the probability of being hired.

would not be possible to disentangle which factors affect each of these separate decisions.

4.2 Selectivity-adjusted earnings equations

If the above discrete choice model is complemented with log-earnings equations, one can obtain parameters for the latter which are adjusted for potential sample-selectivity biases. As not all individuals report their earnings or have positive earnings, the previous discrete choice model is re-estimated for a subsample of workers with positive earnings and is used to estimate the selectivity correction terms in equation (3).[18]

The results of these earnings estimations are presented in Table 3 which shows that, consistent with standard Mincerian equations, age has increasing concave returns, with inflection points around ages 37–47. However, this shape is statistically significant only among formal workers.

In general, males obtain substantially higher earnings relative to otherwise comparable single females. The earnings premiums range between 20 per cent for single males in formal employment and 60 per cent for married males in voluntary informal employment.[19] In contrast, among voluntary informal workers, married women display a *ceteris paribus* earnings penalty of more than 30 per cent relative to their single female counterparts. These numbers indicate that demographic factors at the household level affect the job allocation process and the earnings obtained in the market.

---

[18] The parameter estimates of the bivariate probit over this slightly more restricted sample are available upon from request from the author.

[19] These premia are obtained by noting that for a given gender x marital group g, the parameters estimated are $\hat{\beta}_g = \widehat{\ln(y_g)} - \widehat{\ln(y_{sf})}$, where $\widehat{\ln(y_g)}$ and $\widehat{\ln(y_{sf})}$ are the predicted log earnings of group g and single females, respectively. Hence $exp(\hat{\beta}_g) - 1$ approximates the percentage premium for group g relative to single females.

The number of years of tenure at a firm have increasing concave returns. However, this shape is statistically significant only among involuntary informal workers, with an inflection point around 14 years of tenure. An extra year of schooling is associated with an increase in earnings of about 15 per cent in formal jobs and 8 per cent among those who do not want a job with coverage, and has no significant effect on the earnings of involuntary informal workers. Being enrolled in school leads to *ceteris paribus* earnings losses of almost 40 per cent among formal workers only.

Finally, most of the coefficients for the selectivity correction terms are statistically insignificant at the 90 per cent level. The exception occurs for the correction term in the earnings equation of voluntary informal workers. In this case, the unobserved factors that affect the probability of applying for formal jobs are negatively correlated with the unobservables in the earnings equation of these workers.

It is worth mentioning that the fit of the earnings equations is higher for workers employed in their preferred sector, i.e., formal and voluntary informal workers. This indicates that there is a greater degree of unexplained heterogeneity in the earnings of involuntary informal workers.

Finally, using these regression models, counterfactual earnings of informal workers are predicted if they worked in formal employment. These differentials are presented in Appendix Table A1. As it can be seen these differentials are not statistically different to zero.

4.3 Robustness checks

In addition to the baseline specification above, I estimate two additional models to test the robustness of the previous results. In the first model the definition of 'formal employment' is modified to incorporate in this group the formal self-employed, as defined by the Mexican Statistical Agency (INEGI, 2014).

In this alternative sample, the interpretation of the hiring equation (1.2) changes, as it now describes factors that affect hiring by a formal sector employer, as well as the costs and benefits for a firm operating formally (see Ulyssea, 2020 for a discussion of what these costs might be). [20]

The average partial effects of this specification are presented in Table 4.[21] A comparison of these results with those in Table 2, shows that the estimated partial effects are very similar and the differences are mainly in the magnitudes and statistical significance of some effects. Among these differences, one can note that marital status has a weaker effect on the hiring equation in this sample. Also, the effects of the dependency ratio on the application decision for men are stronger, and the years of schooling positively affect the probability of applying for a formal job, while in the previous sample this effect was statistically insignificant. This last finding indicates that more educated workers are more likely to seek establishing a formal business.

---

[20] The second sample is smaller than the first one because I import information from the Labour Force Survey (INEGI, 2015b) to determine who among the self-employed is formal, and in the process of merging both samples, I discard observations with discrepancies at the job characteristics level (e.g., occupation, industry, etc.). Additional descriptive statistics for this sample are found in Table S1 in the Supplementary Material.

[21] The original parameter estimates are presented in the Supplementary material Table S5.

The results corresponding to the earnings equations estimations in this sample are reported in Table 5. The results are qualitatively similar to those of the baseline sample and, for the sake of brevity, I omit further discussion of them.[22]

A second robustness check adds to the baseline specification a regressor measuring whether an adult member in the household has a formal job and examines whether this influences the respondent's decision to apply to a formal job. To save space the results of this second specification are presented in the Supplementary Material Tables S2, S3, and S6.

The results reported in Table S2 of the Supplementary Material show that having a member of the household employed in a formal job *increases* the respondent's probability of applying to formal employment by about 7 percentage points, while the coefficients of other variables are very similar to the ones reported in Table 2. This means that rather than strategically dividing who is formal and who is informal, respondents with a formally employed household member are more likely to seek such type of employment as well. This probably reflects positive assortative mating, whereby individuals with a higher valuation of formal jobs form households together.[23]

In summary, the econometric analysis performed indicates that the division of labour at home across gender lines plays an important role in determining who applies for and who is hired in formal jobs, as well as the earnings therein gained. Also, the

[22] The predicted counterfactual earnings of informal workers if they worked in formal employment are presented in Appendix Table A2. Again, none of these differentials are statistically different to zero.

[23] An interesting extension of this exercise would be to jointly model the sectoral labour supply of couples, rather than treating the employment status of other family members as exogenously given. This exercise is however, beyond the scope of this paper.

levels of schooling are a crucial factor affecting both the probability of being hired in a formal job and the earnings obtained in it.

5. Conclusions

This paper exploits a unique dataset containing information about the preferred type of jobs of workers in large urban centres in Mexico. Comparing this information with the actual jobs they have, it is estimated that almost 80 per cent of informal workers consider it preferable to be employed in a formal job which provides them with social security coverage, even if it entails paying the corresponding taxes for such benefits. This suggests that many of the urban informal workers are in this sector because of a lack of better options.

Our econometric analysis highlights two important factors that affect the workers' access to formal employment, i.e., the division of housework at the household level and the levels of human capital. In particular, the traditional division of labour at home is a likely culprit for limiting the willingness of females to apply for formal wage jobs and the probability of being hired in such jobs. In addition, having a higher level of education plays a significant role in increasing the chances of being hired in a formal job and of earning a higher income from it. While one should not draw direct policy recommendations from these findings, it is important to consider how policy can affect these two dimensions in order to encourage the successful transition of workers into formal employment.

Finally, one methodological point arises from this research. So far, the overwhelming majority of research on labour markets in developing countries is based on variables such as wages, employment status, and so forth. The analysis conducted here shows that there is much to be gained by also considering the stated preferences of workers about potential jobs and their characteristics.

This new piece of information can enrich our understanding of the functioning of labour markets and the welfare of workers. In particular, this type of information can help to solve some unresolved theoretical debates in the literature, where traditional analyses have led to ambiguous conclusions. To exploit this type of information, however, will require a better data collection effort, including carefully worded questions for eliciting workers' preferences for different types of job characteristics.

References

Alcaraz, C., Chiquiar, D., & Saucedo, A. (2015). *Informality and Segmentation in the Mexican Labor Market* (Banco de México Working Papers 2015-25). Mexico City: Mexico. https://doi.org/10.36095/banxico/di.2015.25

Arceo-Gomez, E.O., & Campos-Vazquez, R.M. (2014). Race and Marriage in the Labor Market: A Discrimination Correspondence Study in a Developing Country. *American Economic Review*, *104*(5): 376–80. https://doi.org/10.1257/aer.104.5.376

Bosch, M., & Maloney, W.F. (2010). Comparative Analysis of Labor Market Dynamics Using Markov Processes: An Application to Informality. *Labour Economics*, *17*(4): 621–31. https://doi.org/10.1016/j.labeco.2010.01.005

Contreras, D., Gillmore, R., & Puentes, E. (2017). Self-Employment and Queues for Wage Work: Evidence from Chile. *Journal of International Development*, *24*(4): 473–99. https://doi.org/10.1002/jid.3074

Duval-Hernández, R. (2020). 'By choice or by force? Uncovering the nature of informal employment in urban Mexico' WIDER Working Paper 2020/151. https://doi.org/10.35188/UNU-WIDER/2020/908-2

Duval-Hernández, R., & Smith, R. (2010). 'Counterfactual Wages and Segmentation between Formal and Informal Sectors in Mexico'. Unpublished Manuscript División de Economía, CIDE, Mexico City, Mexico.

Fields, G.S. (1975). Rural-Urban Migration, Urban Unemployment and Underemployment, and Job-Search Activity in LDCs. *Journal of Development Economics*, *2*(2): 165–87. https://doi.org/10.1016/0304-3878(75)90014-0

Fields, G.S. (1990). Labour Market Modelling and the Urban Informal Sector: Theory and Evidence’. In D. Turnham, B. Salome, and A. Schwarz (Eds), *The Informal Sector Revisited*. Paris: Organisation for Economic Co-operation and Development.

Fiess, N.M., Fugazza, M., & Maloney, W.F. (2010). Informal Self-employment and Macroeconomic Fluctuations. *Journal of Development Economics, 91*(2): 211–26. https://doi.org/10.1016/j.jdeveco.2009.09.009

Gindling, T.H., & Newhouse, D. (2014). Self-Employment in the Developing World. *World Development, 56*: 313–31. https://doi.org/10.1016/j.worlddev.2013.03.003

Günther, I., & Launov, A. (2012). Informal Employment in Developing Countries: Opportunity or Last Resort?. *Journal of Development Economics, 97*: 88–98.

INEGI (Instituto Nacional de Estadística y Geografía) (2014). *La informalidad laboral: Encuesta Nacional de Ocupación y Empleo: marco conceptual y metodológico* [Labour Informality. National Survey of Occupation and Employment: conceptual and methodological framework]. México: INEGI.

INEGI (Instituto Nacional de Estadística y Geografía) (2015a). *Módulo de Trayectorias Laborales (MOTRAL) 2015* [Module of Labour Trajectories (MOTRAL)_2015]. Available at: https://www.inegi.org.mx/programas/motral/2015/ (accessed 1 July 2019).

INEGI (Instituto Nacional de Estadística y Geografía) (2015b). *Encuesta Nacional de Ocupación y Empleo, población de 15 años y más de edad* [National Survey of Occupation and Employment, population aged 15 years and more]. Available at: https://www.inegi.org.mx/programas/enoe/15ymas/ (accessed 1 November 2018).

Levy, S. (2008). *Good Intentions, Bad Outcomes: Social Policy, Informality, and Economic Growth in Mexico*. Washington, DC: Brookings Institution Press.

Magnac, T. (1991). Segmented or Competitive Labor Markets. *Econometrica*, *59*(1): 165–87. https://doi.org/10.2307/2938245

Maloney, W.F. (1999). Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico. *World Bank Economic Review*, *13*(2): 275–302. https://doi.org/10.1093/wber/13.2.275

Maloney, W.F. (2004). Informality Revisited. *World Development*, *32*(7): 1159–78. https://doi.org/10.1016/j.worlddev.2004.01.008

Marcouiller, D., Ruiz de Castilla, V., & Woodruff, C. (1997). Formal Measures of the Informal-Sector Wage Gap in Mexico, El Salvador, and Peru. *Economic Development and Cultural Change*, *45*(2): 367–92. https://doi.org/10.1086/452278

Soares, F.V. (2004). *Do Informal Workers Queue for Formal Jobs in Brazil?* (Texto para Discussão #1021), Rio de Janeiro: IPEA.

Tunalı, İ. (1986). A General Structure for Models of Double Selection and an Application to a Joint Migration/Earnings Process with Remigration. *Research in Labor Economics, 8*: 235–82.

Ulyssea, G. (2020). Informality: Causes and Consequences for Development. *Annual Review of Economics*, *12*(1): 525–46. https://doi.org/10.1146/annurev-economics-082119-121914

Van de Ven, W.P.M.M., & Van Praag, B.M.S. (1981). The Demand for Deductibles in Private Health Insurance. A Probit Model with Sample Selection. *Journal of Econometrics*, *17*(2): 229–52. https://doi.org/10.1016/0304-4076(81)90028-2

Table 1. Characteristics of employed population in 2015 MOTRAL module.

	Has social security		Does not have social security	
	Wants social security	Does not want social security	Wants social security	Does not want social security
	[1]	[2]	[3]	[4]
Years of age	37.1	35.4	38.4	37.1
Male (%)	48.2	63.0	48.4	49.7
Years of schooling	11.8	12.4	9.9	10.7
Enrolled in school (%)	3.9	2.9	6.1	6.3
Married (%)	45.1	42.5	44.7	32.7
Household composition				
Dependency ratio	0.61	0.59	0.61	0.72
# Adult females	1.06	1.55	1.06	1.00
Earnings (monthly Mx Pesos)	7,004	9,075	4,320	3,599
Share of employment	43.5	8.9	38.3	9.3
# Obs. (unweighted)	1,906	393	1,759	439

Note: unless otherwise stated, all numbers are averages across the different employment groups. All estimates use sampling weights.

Source: author's calculations based on MOTRAL 2015 (INEGI 2015a) and ENOE 2nd Quarter 2015 (INEGI 2015b).

Table 2. Average partial effects of discrete choice models. Formality status defined by job with social security.

	Probit			Bivariate probit with selectivity					
				Apply			Hire | Apply		
	[1]	[2]		[3]	[4]		[5]	[6]	
Age									
Females at 18 years	0.019	(0.004)	***	0.0083	(0.003)	**	0.0170	(0.004)	***
Females at 35 years	0.0022	(0.001)	***	0.0014	(0.001)	**	0.0014	(0.001)	*
Females at 54 years	-0.017	(0.005)	***	-0.003	(0.003)		-0.015	(0.005)	***
Males at 18 years	-0.017	(0.007)	**	-0.003	(0.001)	***	-0.013	(0.004)	***
Males at 35 years	-0.008	(0.002)	***	-0.0031	(0.002)		-0.0066	(0.002)	***
Males at 54 years	0.006	(0.008)		0.0013	(0.005)		0.0048	(0.006)	
Married									
Females	-0.084	(0.023)	***	0.042	(0.027)		-0.130	(0.013)	***
Males	0.124	(0.016)	***	0.048	(0.022)	**	0.116	(0.022)	***
Male	0.0006	(0.018)		-0.0135	(0.009)		0.0169	(0.013)	
Years of schooling	0.029	(0.003)	***	0.0010	(0.003)		0.031	(0.003)	***
Enrolled in school	-0.22	(0.073)	***	-0.030	(0.035)		-0.24	(0.084)	***
Dependency ratio (DR)									
Females									
DR w/ no adult females	-0.066	(0.027)	**	-0.049	(0.025)	*			
DR w/ 1 adult female	0.021	(0.020)		0.0035	(0.008)				
Males									
DR w/ no adult females	0.063	(0.031)	*	0.025	(0.012)	**			
DR w/ 1 adult female	0.049	(0.046)		0.016	(0.010)				

Note: standard errors robust to clustering at the city level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Inference done with a t-distribution with 32 clusters -1 as degrees of freedom. The variable ‘Adult females’ counts the number of females aged 18+ in the household, excluding the survey respondent in the case of female respondents. City fixed effects included. All estimates use sampling weights.

Source: author’s calculations based on MOTRAL 2015 (INEGI 2015a) and ENOE 2nd Quarter 2015 (INEGI 2015b).

Table 3: Selectivity-adjusted log-earnings OLS equations. Formality status defined by job with social security

	Has social security	Does not have social security	
		Wants social security	Does not want social security
Age	0.0271**	0.0425	0.0516
	(0.0106)	(0.0322)	(0.0419)
Age sq.	-0.000367***	-0.000444	-0.000610
	(0.000119)	(0.000318)	(0.000561)
Single female (omitted)			
Married female	-0.0540	-0.145	-0.384**
	(0.0891)	(0.152)	(0.161)
Single male	0.182**	0.439***	0.401***
	(0.0679)	(0.123)	(0.100)
Married male	0.405***	0.307***	0.472***
	(0.0462)	(0.104)	(0.101)
Tenure	0.0392	0.0245***	0.00984
	(0.0328)	(0.00643)	(0.0371)
Tenure sq.	-0.00144	-0.00104**	-0.0000236
	(0.00133)	(0.000387)	(0.00118)
Years schooling	0.147***	0.0391	0.0792***
	(0.0360)	(0.0385)	(0.0194)
Enrolled in school	-0.467**	-0.130	-0.315
	(0.204)	(0.345)	(0.191)
λ(apply)	-0.569	-0.363	-0.403**
	(0.462)	(0.669)	(0.183)
λ(hire)	0.788	-0.280	
	(0.510)	(0.836)	
Constant	5.756***	6.346***	4.853***
	(0.596)	(1.150)	(0.667)
R^2	0.336	0.158	0.487
N	2091	1535	364

Note: standard errors robust to clustering at the city level and adjusted for generated regressors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Inference done with a t-distribution with 32 clusters -1 as degrees of freedom. Formal workers are those in a job with social security coverage. City fixed effects included. Sample includes only individuals with positive earnings. All estimates use sampling weights.

Source: author's calculations based on MOTRAL 2015 (INEGI 2015a) and ENOE 2nd Quarter 2015 (INEGI 2015b).

Table 4: Average partial effects of discrete choice models. Formality status defined by job with social security or by formal self-employment

	Probit			Bivariate probit with selectivity					
				Apply			Hire | Apply		
	[1]	[2]		[3]	[4]		[5]	[6]	
Age									
Females at 18 years	0.015	(0.004)	***	0.0097	(0.003)	***	0.0120	(0.004)	***
Females at 35 years	0.0054	(0.001)	***	0.0026	(0.001)	***	0.0037	(0.001)	***
Females at 54 years	-0.005	(0.006)	***	-0.001	(0.002)		-0.004	(0.005)	
Males at 18 years	-0.019	(0.008)	**	-0.004	(0.001)	***	-0.015	(0.007)	**
Males at 35 years	-0.003	(0.002)		-0.0021	(0.002)		-0.0018	(0.001)	
Males at 54 years	0.018	(0.009)	*	0.0045	(0.004)		0.0144	(0.007)	*
Married									
Females	-0.042	(0.037)		0.040	(0.022)	*	-0.080	(0.028)	***
Males	0.120	(0.030)	***	0.060	(0.028)	**	0.091	(0.048)	*
Male	0.0047	(0.021)		-0.0152	(0.011)		0.0242	(0.015)	
Years of schooling	0.045	(0.003)	***	0.0069	(0.002)	***	0.043	(0.004)	***
Enrolled in school	-0.19	(0.054)	***	-0.031	(0.025)		-0.20	(0.061)	***
Dependency ratio (DR)									
Females									
DR w/ no adult females	-0.046	(0.014)	***	-0.044	(0.020)	**			
DR w/ 1 adult female	0.032	(0.026)		0.0123	(0.015)				
Males									
DR w/ no adult females	0.085	(0.037)	**	0.083	(0.026)	***			
DR w/ 1 adult female	0.089	(0.051)	*	0.069	(0.028)	**			

Note: standard errors robust to clustering at the city level in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01. Inference done with a t-distribution with 32 clusters -1 as degrees of freedom. The variable 'Adult females' counts the number of females aged 18+ in the household, excluding the survey respondent in the case of female respondents. City fixed effects included. All estimates use sampling weights.

Source: author's calculations based on MOTRAL 2015 (INEGI 2015a) and ENOE 2nd Quarter 2015 (INEGI 2015b).

Table 5: Selectivity-adjusted log-earnings OLS equations. Formality status defined by job with social security or by formal self-employment

	Formal employment	Informal employment	
		Wants social security	Does not want social security
Age	0.000953	0.0726	0.0987
	(0.0168)	(0.0640)	(0.0635)
Age sq.	-0.0000264	-0.000859	-0.00140
	(0.000193)	(0.000766)	(0.000874)
Single female (omitted)			
Married female	0.100	-0.296**	-0.340**
	(0.0882)	(0.110)	(0.135)
Single male	0.172**	0.501***	0.545***
	(0.0636)	(0.156)	(0.131)
Married male	0.471***	0.375***	0.390***
	(0.0731)	(0.0763)	(0.103)
Tenure	0.0372	0.0120	-0.000687
	(0.0354)	(0.00976)	(0.0341)
Tenure sq.	-0.00125	-0.00121**	-0.0000208
	(0.00128)	(0.000495)	(0.000953)
Years schooling	0.157***	0.0401	0.0357**
	(0.0337)	(0.0474)	(0.0159)
Enrolled in school	-0.659	-0.114	-0.397**
	(0.397)	(0.445)	(0.169)
λ(apply)	-0.337	-0.115	-0.339**
	(0.345)	(0.226)	(0.151)
λ(hire)	0.979**	0.187	
	(0.471)	(0.453)	
Constant	5.979***	6.165***	4.767***
	(0.555)	(1.139)	(0.943)
R^2	0.293	0.193	0.472
N	2026	1131	276

Note: standard errors robust to clustering at the city level and adjusted for generated regressors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Inference done with a t-distribution with 32 clusters -1 as degrees of freedom. Formal workers are those in a job with social security coverage and the formal self-employed. City fixed effects included. Sample includes only individuals with positive earnings. All estimates use sampling weights.

Source: author's calculations based on MOTRAL 2015 (INEGI 2015a) and ENOE 2nd Quarter 2015 (INEGI 2015b).

Appendix

Table A1: Predicted log-earnings differential. Formal employment – actual. Formality status defined by job with social security

	Does not have social security	
	Wants social security	Does not want social security
$\widehat{\ln y_f} - \overline{\ln y}$	-.947 (.81)	1.013 (.604)

Note: predictions based on the models estimated in Table 3. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All estimates use sampling weights.

Source: author's calculations based on MOTRAL 2015 (INEGI 2015a) and ENOE 2nd Quarter 2015 (INEGI 2015b).

Table A2: Predicted log-earnings differential. Formal employment – actual. Formality status defined by job with social security or formal self-employment

	Informal employment	
	Wants social security	Does not want social security
$\widehat{\ln y_f} - \overline{\ln y}$	-1.26 (.80)	.462 (0.436)

Note: Predictions based on the models estimated in Table 5. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All estimates use sampling weights.

Source: author's calculations based on MOTRAL 2015 (INEGI 2015a) and ENOE 2nd Quarter 2015 (INEGI 2015b).

Supplementary Material

This supplementary appendix presents in more detail the econometric models estimated in the paper. It also includes additional empirical results not reported in the main text.

S.1 Discrete Choice Model

For an individual i, denote the latent propensity to apply to a formal job by V_i^a, and the propensity of being hired by V_i^h. These propensities depend on characteristics of the workers, some of which are observable to the econometrician. In particular, assume that the following structure holds:

$$V_i^a = Z_i\gamma_a + u_{ia} \tag{S.1}$$
$$V_i^h = X_i\gamma_h + u_{ih}, \tag{S.2}$$

where Z_i and X_i are vectors of observable individual characteristics, and u_{ia} and u_{ih} are random terms capturing other unobservable factors. Assume that the vector of error terms (u_{ia}, u_{ih}) follows a standard bivariate normal law with correlation parameter ρ; i.e. the vector has zero mean and variance-covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \tag{S.3}$$

The above propensities in equations (S.1) and (S.2) are not observable. Instead it is only observed whether they are positive or not, i.e. only the following indicators are observed

$$D_{ij} = \begin{cases} 1 & \text{if } V_i^j > 0 \\ 0 & \text{if } V_i^j \leq 0 \end{cases} \tag{S.4}$$

for $j \in \{a, h\}$.

Denote by $G(\cdot,\cdot;\cdot)$ the standard bivariate normal distribution and by $F(\cdot)$ its univariate counterpart. Also, let $C_a = Z\gamma_a$ and $C_h = X\gamma_h$.

Given the above assumptions and notation, one can characterise the probability that a given individual with explanatory variables (Z_i, X_i) is either a formal, $i \in \mathcal{F}$, an involuntary informal, $i \in \mathcal{I}$, or a voluntary informal, $i \in \mathcal{V}$, worker.[1] These probabilities are given by the following equations:

[1]From now on, the individual subscript i is dropped for simplicity.

Formal worker

$$
\begin{aligned}
P(D_a = 1, D_h = 1) &= P(V^a > 0, V^h > 0) \\
&= P(u_a > -Z\gamma_a, u_h > -X\gamma_h) \\
&= G(C_a, C_h; \rho)
\end{aligned}
\tag{S.5}
$$

Involuntary informal worker

$$
\begin{aligned}
P(D_a = 1, D_h = 0) &= P(V^a > 0, V^h \leq 0) \\
&= P(u_a > -Z\gamma_a, u_h \leq -X\gamma_h) \\
&= G(C_a, -C_h; -\rho)
\end{aligned}
\tag{S.6}
$$

Voluntary informal worker

$$
\begin{aligned}
P(D_a = 0) &= P(V^a \leq 0) \\
&= P(u_a \leq -Z\gamma_a) \\
&= F(-C_a)
\end{aligned}
\tag{S.7}
$$

With these probabilities, the likelihood function of the discrete choice problem can be formulated as

$$
\mathcal{L} = \prod_{D_a=0} [1 - F(C_a)] \cdot \prod_{\substack{D_a=1 \\ D_h=0}} G(C_a, -C_h; -\rho) \cdot \prod_{\substack{D_a=1 \\ D_h=1}} G(C_a, C_h; \rho). \tag{S.8}
$$

This model is sometimes called a bivariate probit with sample selection (see for instance Van de Ven and Van Praag, 1981).

Furthermore, this framework can be extended as to estimate earnings regressions adjusted for sample selectivity. In particular, assume the existence of segment-specific earnings functions

$$
\log y_s = X_s \beta_s + u_s
$$

where y denotes earnings, X_s is a vector of observable characteristics, u is an unobservable residual, and the subscript s denotes the type of worker under consideration, namely: formal, involuntary informal, and voluntary informal. To estimate earnings equations adjusted for sample selectivity bias, assume the joint multivariate normality of the error terms (u_a, u_h) and the three errors terms e_s, for $s \in \{\mathcal{F}, \mathcal{I}, \mathcal{V}\}$.

This is a switching regression model with earnings functions defined for all the employed population. However, for any given individual only one realization of earnings is observed, depending on the market segment at which he or she ends up being employed. More precisely, the segment switching rule is the following:

$$
\log y = \begin{cases} \log y_{\mathcal{V}} & \text{if } D_a = 0 \\ \log y_{\mathcal{I}} & \text{if } D_a = 1 \text{ and } D_h = 0 \\ \log y_{\mathcal{F}} & \text{if } D_a = 1 \text{ and } D_h = 1 \end{cases}
$$

Due to the fact that the segment allocation is not random, when estimating the earnings equations through Ordinary Least Squares in a second stage, one needs to correct for sample selectivity bias. In particular, one needs to estimate

$$\begin{array}{ll} \log y = X_{\mathcal{V}}\beta_{\mathcal{V}} + \rho_a^{\mathcal{V}}\hat{\lambda}_0 + \tilde{u}_{\mathcal{V}} & \text{if } D_a = 0 \\ \log y = X_{\mathcal{I}}\beta_{\mathcal{I}} + \rho_a^{\mathcal{I}}\hat{\lambda}_{a\mathcal{I}} + \rho_h^{\mathcal{I}}\hat{\lambda}_{h\mathcal{I}} + \tilde{u}_{\mathcal{I}} & \text{if } D_a = 1 \text{ and } D_h = 0 \\ \log y = X_{\mathcal{F}}\beta_{\mathcal{F}} + \rho_a^{\mathcal{F}}\hat{\lambda}_{a\mathcal{F}} + \rho_h^{F}\hat{\lambda}_{hF} + \tilde{u}_{\mathcal{F}} & \text{if } D_a = 1 \text{ and } D_h = 1 \end{array} \tag{S.9}$$

The selection-correction terms are given by

$$\begin{array}{ll} \lambda_0 = -\frac{f(C_a)}{F(-C_a)} & \\ \lambda_{a\mathcal{I}} = \frac{f(C_a)F(-C_h^*)}{G(C_a,-C_h;-\rho)} & \lambda_{h\mathcal{I}} = -\frac{f(C_h)F(C_a^*)}{G(C_a,-C_h;-\rho)} \\ \lambda_{a\mathcal{F}} = \frac{f(C_a)F(C_h^*)}{G(C_a,C_h;\rho)} & \lambda_{h\mathcal{F}} = \frac{f(C_h)F(C_a^*)}{G(C_a,C_h;\rho)} \end{array} \tag{S.10}$$

where C_a^* and C_h^* are given by

$$C_a^* = \frac{C_a - \rho C_h}{(1-\rho^2)^{\frac{1}{2}}} \qquad C_h^* = \frac{C_h - \rho C_a}{(1-\rho^2)^{\frac{1}{2}}}.$$

Since the selectivity correction terms λ's are estimated using information arising from the parameters estimates of the discrete choice model (i.e., $\{\hat{\gamma}_a, \hat{\gamma}_h\}$), the standard errors of the regressions (S.9) must be adjusted to account for the presence of generated regressors. The formulas to do this, as well as the derivation of the full model can be found in Tunalı (1986).

S.2 Additional Empirical Results

This section presents additional results not included in the main text.

Table S1: Characteristics of the samples used for estimation of the Discrete Choice Models

	Sample 1			Sample 2		
		Does not have social security (SS)			Informal employment	
	Has SS	Wants SS	Doesn't want SS	Formal	Wants SS	Doesn't want SS
Percentage	52.4	38.3	9.3	59.7	32.1	8.2
Years of age	36.8	38.4	37.1	37.8	37.8	35.6
Male (%)	50.7	48.4	49.7	49.3	44.8	48.3
Years of schooling	11.9	9.9	10.7	12.1	9.1	9.8
Schooling level (%)						
Elementary	10.2	24.2	16.7	10.7	27.6	19.4
Intermediate	56.1	58.9	61.0	54.1	63.0	69.0
Higher	33.6	16.9	22.2	35.3	9.5	11.6
Enrolled in school (%)	3.7	6.1	6.3	3.8	5.3	5.9
Married (%)	44.7	44.7	32.7	46.7	42.2	26.6
Household composition						
Dependency ratio	0.60	0.61	0.72	0.60	0.62	0.71
# Adult females	1.14	1.06	1.00	1.13	1.01	0.96
Social Security others (%)	47.9	33.9	22.4	46.5	34.3	24.6
Tenure	7.35	6.95	7.68	7.78	6.14	6.30
Earnings (monthly Mx. Pesos)	7,344	4,320	3,599	7,363	3,517	2,853
Number of obs.	2,299	1,759	439	2,258	1,278	333

Note: unless otherwise stated, all numbers are averages across the different employment groups. Sample 1 is used in the estimation of the discrete choice model presented in Table 2 in the main manuscript, while Sample 2 is used in the estimation of the discrete choice model presented in Table 4 in the main manuscript. The variable '# of Adult females' counts the number of females aged 18+ in the household, excluding the survey respondent in the case of female respondents. The variable 'Social Security others' takes value 1 if other household members are formally employed, and zero otherwise. Earnings are measured in 2015 Mexican Pesos per month. All estimates use sampling weights.

Table S2: Average partial effects of discrete choice models including formality of household member as regressor. Formality status defined by job with social security

	Probit			Bivariate Probit					
				Apply			Hire | Apply		
Age									
Females at 18 yrs	0.019	(0.004)	***	0.0095	(0.003)	***	0.0187	(0.005)	***
Females at 35 yrs	0.0030	(0.001)	***	0.0021	(0.001)	***	0.0017	(0.001)	*
Females at 54 yrs	-0.016	(0.005)	***	-0.002	(0.003)		-0.015	(0.005)	***
Males at 18 yrs	-0.015	(0.008)	*	-0.003	(0.002)		-0.012	(0.004)	***
Males at 35 yrs	-0.007	(0.002)	***	-0.0025	(0.002)		-0.0063	(0.002)	***
Males at 54 yrs	0.004	(0.008)		0.0007	(0.004)		0.0045	(0.006)	
Married									
Females	-0.107	(0.023)	***	0.025	(0.021)		-0.137	(0.012)	***
Males	0.111	(0.016)	***	0.041	(0.021)	*	0.112	(0.020)	***
Male	0.0230	(0.014)		-0.0022	(0.008)		0.022	(0.012)	*
Yrs of Schooling	0.027	(0.004)	***	-0.00021	(0.004)		0.031	(0.003)	***
Enrolled in School	-0.22	(0.075)	***	-0.024	(0.036)		-0.24	(0.085)	***
Dependency Ratio (DR)									
Females									
DR w/ no Adult Females	-0.045	(0.024)	*	-0.037	(0.020)	*			
DR w/ 1 Adult Female	0.029	(0.022)		0.0058	(0.008)				
Males									
DR w/ no Adult Females	0.109	(0.041)	**	0.050	(0.016)	***			
DR w/ 1 Adult Female	0.082	(0.058)		0.032	(0.016)	**			
Social Security others	0.13	(0.024)	***	0.073	(0.021)	***			

Note: standard errors robust to clustering at the city level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Inference done with a t-distribution with 32 clusters -1 as degrees of freedom. The variable 'Adult females' counts the number of females aged 18+ in the household, excluding the survey respondent in the case of female respondents. The variable 'Social Security others' takes value 1 if other household members are formally employed, and zero otherwise. City fixed effects included. All estimates use sampling weights.

Table S3: Selectivity-adjusted log-earnings OLS equations corresponding to the discrete choice model that includes formality of household members as regressor

	Has social security	Does not have social security	
		Wants social security	Does not want social security
Age	0.0193	0.0377	0.0539
	(0.0133)	(0.0307)	(0.0455)
Age sq.	-0.000253*	-0.000381	-0.000631
	(0.000144)	(0.000297)	(0.000607)
Single female (omitted)			
Married female	0.0273	-0.107	-0.353**
	(0.0603)	(0.158)	(0.157)
Single male	0.206***	0.451***	0.409***
	(0.0526)	(0.124)	(0.0916)
Married male	0.427***	0.327***	0.496***
	(0.0431)	(0.101)	(0.0907)
Tenure	0.0410	0.0248***	0.0112
	(0.0346)	(0.00616)	(0.0378)
Tenure sq.	-0.00152	-0.00106***	-0.0000599
	(0.00141)	(0.000378)	(0.00121)
Years schooling	0.140***	0.0364	0.0825***
	(0.0329)	(0.0397)	(0.0181)
Enrolled in school	-0.409**	-0.109	-0.303
	(0.174)	(0.356)	(0.179)
λ(apply)	0.0224	0.347	-0.269*
	(0.209)	(0.914)	(0.152)
λ(hire)	0.474	-0.399	
	(0.355)	(0.884)	
Constant	6.011***	6.300***	4.982***
	(0.508)	(1.159)	(0.852)
R^2	0.336	0.157	0.481
N	2091	1535	364

Note: standard errors robust to clustering at the city level and adjusted for generated regressors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Inference done with a t-distribution with 32 clusters -1 as degrees of freedom. Formal workers are those in a job with social security coverage. City fixed effects included. Sample includes only individuals with positive earnings. All estimates use sampling weights.

S.2.1 Discrete Choice Parameters

Table S4: Parameter estimates of discrete choice models. Formality status defined by job with social security

	Bivariate Probit with selectivity		
	Apply	Hire	Probit
Females			
Age	0.0658**	0.0756***	0.106***
	(0.0253)	(0.0191)	(0.0251)
Age sq.	-0.000796**	-0.00104***	-0.00142***
	(0.000364)	(0.000277)	(0.000368)
Males			
Age	-0.0647	-0.0642**	-0.0908
	(0.0740)	(0.0289)	(0.0576)
Age sq.	0.000663	0.000702*	0.000987
	(0.000907)	(0.000378)	(0.000744)
Single female (omitted)			
Married female	0.287	-0.390***	-0.236***
	(0.203)	(0.0374)	(0.0660)
Single male	2.488**	2.360***	3.397***
	(1.205)	(0.634)	(1.077)
Married male	2.785**	2.634***	3.738***
	(1.335)	(0.608)	(1.059)
Females			
Dependency ratio	-0.272**		-0.184**
	(0.130)		(0.0736)
Dep. ratio x # Adult females	0.298**		0.241**
	(0.139)		(0.0925)
Males			
Dependency ratio	0.169**		0.173*
	(0.0815)		(0.0855)
Dep. ratio x # Adult females	-0.0684		-0.0366
	(0.0617)		(0.0796)
Years schooling	0.00664	0.0846***	0.0817***
	(0.0205)	(0.00805)	(0.0103)
Enrolled in school	-0.172	-0.626**	-0.637***
	(0.185)	(0.238)	(0.224)
ρ		-0.534***	
		(0.154)	
Constant	0.0815	-1.786***	-2.542***
	(0.517)	(0.277)	(0.340)

Note: standard errors robust to clustering at the city level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The variable ‘Adult females’ counts the number of females aged 18+ in the household, excluding the survey respondent in the case of female respondents. City fixed effects included. Inference done with a t-distribution with 32 clusters -1 as degrees of freedom. All estimates use sampling weights. The number of observations is 4,497.

Table S5: Parameter estimates of discrete choice models. Formality status defined by job with social security or by formal self-employment

	Bivariate Probit with selectivity		
	Apply	Hire	Probit
Females			
Age	0.0853***	0.0448	0.0832**
	(0.0267)	(0.0274)	(0.0325)
Age sq.	-0.000935**	-0.000479	-0.000933*
	(0.000408)	(0.000407)	(0.000493)
Males			
Age	-0.0996	-0.0937*	-0.133
	(0.0878)	(0.0501)	(0.0798)
Age sq.	0.00124	0.00121*	0.00170
	(0.00110)	(0.000667)	(0.00104)
Single female (omitted)			
Married female	0.273	-0.251***	-0.130
	(0.196)	(0.0673)	(0.106)
Single male	3.266**	2.567***	3.870***
	(1.374)	(0.765)	(1.185)
Married male	3.699**	2.766***	4.189***
	(1.552)	(0.674)	(1.129)
Females			
Dependency ratio	-0.255**		-0.185***
	(0.0998)		(0.0574)
Dep. ratio x # Adult females	0.353		0.258**
	(0.218)		(0.111)
Males			
Dependency ratio	0.603**		0.294**
	(0.265)		(0.127)
Dep. ratio x # Adult females	-0.186***		-0.0573
	(0.0665)		(0.0940)
Years schooling	0.0461***	0.112***	0.119***
	(0.0117)	(0.00960)	(0.00801)
Enrolled in school	-0.241	-0.621**	-0.634***
	(0.177)	(0.228)	(0.209)
ρ		-0.288**	
		(0.138)	
Constant	-0.738	-1.681***	-2.684***
	(0.453)	(0.324)	(0.393)

Note: standard errors robust to clustering at the city level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The variable 'Adult females' counts the number of females aged 18+ in the household, excluding the survey respondent in the case of female respondents. City fixed effects included. Inference done with a t-distribution with 32 clusters -1 as degrees of freedom. All estimates use sampling weights. The number of observations is 3,869.

Table S6: Parameter estimates of discrete choice models that include formality of household members as regressor. Formality status defined by job with social security.

	Bivariate Probit with selectivity		
	Apply	Hire	Probit
Females			
Age	0.0672**	0.0713***	0.107***
	(0.0248)	(0.0185)	(0.0259)
Age sq.	-0.000753*	-0.000986***	-0.00141***
	(0.000371)	(0.000264)	(0.000380)
Males			
Age	-0.0503	-0.0587**	-0.0798
	(0.0730)	(0.0275)	(0.0585)
Age sq.	0.000499	0.000643*	0.000845
	(0.000895)	(0.000355)	(0.000752)
Single female (omitted)			
Married female	0.164	-0.389***	-0.306***
	(0.153)	(0.0390)	(0.0640)
Single male	2.308*	2.183***	3.275***
	(1.322)	(0.646)	(1.113)
Married male	2.569*	2.430***	3.585***
	(1.447)	(0.621)	(1.090)
Females			
Dependency ratio	-0.215*		-0.126*
	(0.114)		(0.0677)
Dep. ratio x # Adult females	0.257*		0.209**
	(0.129)		(0.0910)
Males			
Dependency ratio	0.368***		0.307***
	(0.124)		(0.112)
Dep. ratio x # Adult females	-0.157***		-0.0765
	(0.0436)		(0.0790)
Social Security others	0.468***		0.354***
	(0.149)		(0.0685)
Years schooling	-0.00131	0.0811***	0.0774***
	(0.0234)	(0.00650)	(0.0111)
Enrolled in school	-0.140	-0.602**	-0.649**
	(0.196)	(0.239)	(0.238)
ρ		-0.744***	
		(0.090)	
Constant	-0.0946	-1.660***	-2.703***
	(0.414)	(0.270)	(0.329)

Note: standard errors robust to clustering at the city level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The variable 'Adult females' counts the number of females aged 18+ in the household, excluding the survey respondent in the case of female respondents. City fixed effects included. Inference done with a t-distribution with 32 clusters -1 as degrees of freedom. All estimates use sampling weights. The number of observations is 4,497.

Bibliography

Tunalı, I. (1986). A General Structure for Models of Double Selection and an Application to a Joint Migration/Earnings Process with Remigration. *Research in Labor Economics 8*(Part B), 235–282.

Van de Ven, W. P. and B. M. Van Praag (1981). The demand for deductibles in private health insurance. A probit model with sample selection. *Journal of Econometrics 17*(2), 229–252.