Ensayos–Volumen XXVI, núm.1, mayo 2007, pp. 1-44

Los salarios del sector formal e informal en México: análisis de ganancias y pérdidas por formalización

Jorge O. Moreno Treviño[*1]

Resumen

Este documento analiza las diferencias que existen en los salarios entre el sector formal e informal del mercado laboral y estima las ganancias netas desde una situación de formalización para los trabajadores, utilizando los parámetros de los efectos de tratamiento, particularmente los de: Mínimos Cuadrados Ordinarios (OLS), Tratamiento Promedio (ATE), Tratamiento en los Tratados (TT) y Tratamiento en los No Tratados (TUT). En general, para México desde el año 2000 al 2003, hombres y mujeres con mayores niveles de educación poseen una ganancia neta si se cambian hacia el sector formal; mientras por contraparte, trabajadores con menores niveles educativos tienen un efecto negativo en sus salarios, en el caso de que formalicen su situación laboral. Complementando lo anterior, se observa que prevalecen grandes diferencias en salarios de un sector a otro, cuando se estudia los efectos condicionales de los tratamientos; sin embargo, esta brecha se desvanece ligera pero consistentemente en el tiempo. En el caso de los hombres, la mayor medida de diferencia en salarios proviene de estudiar la diferencia en salarios usando la media incondicional y el efecto simple condicional OLS, los cuales implican una brecha de aproximadamente 33.0 por ciento en 2000, la cual se reduce consistentemente a 25.3 por ciento en 2003. En contraste, para el mismo período y género, el ATE muestra una brecha mucho menor que es de 14.3 por ciento para 2000, y reduce su valor a 11.7 por ciento en 2003. Además, el TUT para los trabajadores informales es aún mucho menor, ya que se reduce de 7.5 a 2.6 por ciento en el período analizado. Las estimaciones para mujeres son estadísticamente significativas y cualitativamente similares. En esta investigación, la evidencia sugiere que el ordenamiento de los agentes entre sectores, por la vía de auto-selección, explica una proporción importante de la diferencia entre salarios formales e informales en México.

[*] Department of Economics, The University of Chicago.

[1] I would like to thank Karla G. Mendoza, Francisco J. Arias, Lorenzo Caliendo, and Sergio A. Salas for all their comments during the development of this document. In particular, I would like to thank Prof. Robert M. Townsend for his invaluable guidance and support throughout the development of this project. All mistakes are my responsibility and the usual disclaimer applies.

Email: jmoreno@uchicago.edu.

Abstract

This document analyzes the differences on wages between formal and informal labor market and estimate the net gains from formalization using the set of treatment effects parameters, namely: Ordinary Least Square (OLS), Average Treatment Effect (ATE), Treatment on the Treated (TT), and Treatment on the Untreated (TUT). In general for Mexico from 2000 to 2003 both men and women with higher levels of education have net gains from switching to the formal sector, while workers with lower levels of education have a negative impact on their expected wages from formalizing. Moreover, large differences prevail by studying the conditional treatment effects, yet they reduce in the period studied. For men, the larger effect are obtained from analyzing the unconditional mean differences and the OLS effect which imply a wage gap of around 33.0 per cent in 2000, but steadily reduces to reach 25.3 in 2003. In contrast, for the same period and gender, the ATE of formality reduces this gap estimation to 14.3 percent and steadily decreases to 11.7 percent. Furthermore, the TUT effect of informal workers seems to be even smaller and also decreasing in time, from 7.5 to 2.6. The estimations and evidence for women are significant, and qualitatively similar. The evidence on this paper suggests that sorting through self-selection accounts for explaining a large fraction of the average difference between formal and informal wages.

1. Introduction

Labor markets in Latin America, and particularly in Mexico, present a common observed regularity in their structure which differs from those in developed countries: a high percentage of their active labor force works in activities outside the tax system and (or) without social security coverage. This fraction of the population has been known in the literature as the informal sector.

Several studies regarding the informal sector and carried out for Latin America by the OECD (1999), the World Bank (2004), and Banco de Mexico (2003) present two empirical findings consistent across time: 1) the fraction of the total labor force in the informal sector has remained stable and persistent over time, particularly in the urban areas where it accounts for 40 to 60 percent of the active labor force depending of the definition of informality used; and, 2) there is a pronounced gap between the observed average wages between these two sectors, where formal wages are consistently higher than informal ones for both genders across time, in particular the informal sector salaries are around 60 to 80 percent of the formal ones for the periods analyzed.

While some analysis have been done for understanding the importance of the tax systems and the role of law rigidities in determining the informal sector by focusing on an macroeconomic framework, fewer analysis have been done to understand the nature of the informal sector as a result of individual rational decisions.

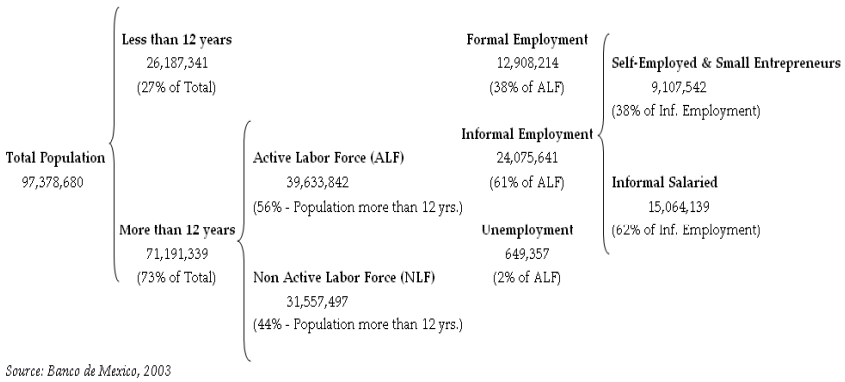
This document presents a proposal for studying the differences in wages between the formal and informal sector by studying the rationality of agents' labor participation on each sector. Using the wage estimations corrected by selection, I calculate alternative treatment effects to measure the gains and lose in wages from formalization. For this purpose, I will study the determinants of aggregate wage distributions for both the formal and informal sectors focusing in the Mexican labor market for the early 2000 years using a model similar to the seminal model studied by McCall (1970) but including a sorting environment similar to that proposed in Willis and Rosen (1979) framework. Hence, the model proposed provides an approach that combines an equilibrium analysis between sectors and the sorting nature of participation decisions with a self selection component in which wages are partially determined by the agent's heterogeneity. By studying the proposed framework, I am able to answer the following questions: 1) how much of the wage inequality is explained by the heterogeneity of the agents who works in each sector, and 2) what are the gains or loses from switching formality in the labor market conditional on the characteristic of the agents.

The present document is organized in five sections including this introduction. The second section presents some literature and a quick review regarding other studies findings for the formal and informal sector. The third section presents the characteristics of the data sets I will use for the analysis of the Mexican labor market and some of the statistical properties in the final data I will use throughout this research. The fourth section presents the econometric model I propose to identify the wage inequality determinants between formal and informal sectors. The fifth section shows the empirical implementation and results of this analysis. The last part of this document concludes the analysis and presents the most important findings.

2. Formality in the Mexican labor market: some facts

Most of the work for analyzing the informal sector in developing countries have been carried out either by local agencies (in the case of Mexico by the National Institute of Statistics, Geography and Informatics, and the Central Bank, (Banco de Mexico - Banxico); or by international organizations, like the World Bank (WB), the International Labour Organization (ILO), and the Inter-American Development Bank (IDB). This section presents the main findings on some of the most representative empirical works carried out for studying the informality on Latin America.

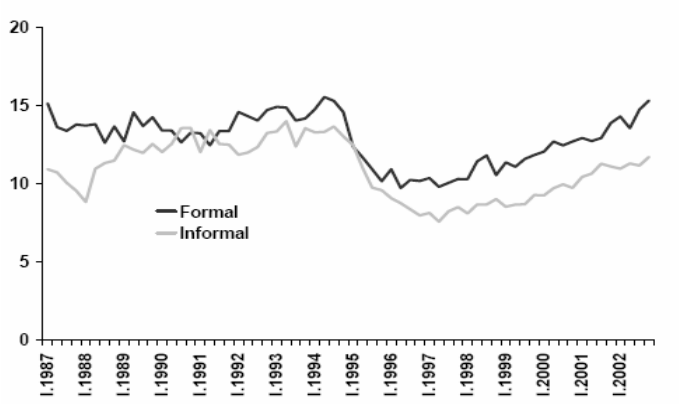
Figure 1
Total population decomposition, Mexico 2000



Source: Banco de Mexico, 2003

The first question to address is to define what the informal sector is, and what the implications are of using this definition on the labor force partition. In general, Banxico (2003) and The World Bank divide the active labor force employment into formal and informal, but distinguishing two types of informality: i) entrepreneurs and self employed; and ii) informal salaried workers. The first group includes the part of the active labor force that produce no agricultural goods and services, which are legal, and with market destiny or scope, through firms which are not in the tax system and are owned by household sector. The second part includes those salaried workers which are not owners of a small firm or beginning a new entrepreneur, yet receive an income per hours worked in a non-formal job. Figure 1 presents a basic scheme of this decomposition and the estimations for Mexico, using data of 2000.

Figure 2
Estimation of median formal-informal labor income, Mexico 1987-2000

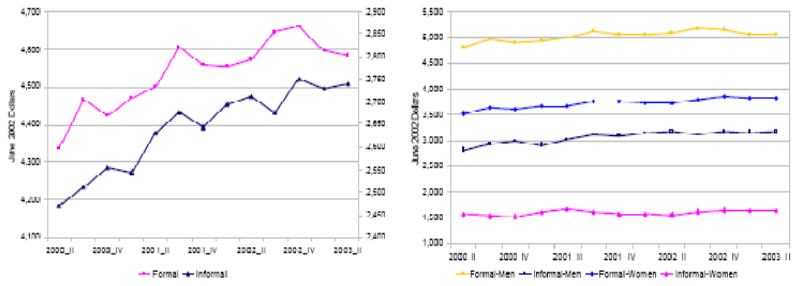


Source: BANXICO 2003, with data from ENEU-INEGI

Maloney´s (1998) working paper is one of the pioneering research work for studying labor market informality in Latin America using microdata for several countries including Brazil, Colombia, El Salvador, and Mexico. Maloney finds that in Latin America, small entrepreneurs who does not pay taxes and do not have social security are indeed a high percentage of the informal sector. In other words, Maloney's hypothesis is that informal sector is not a "residual" of the lack of ability of the formal sector to "create opportunities", but it represents an option for the potential worker to get earnings in a different type of job. In particular, Maloney (1999) logit analysis for rotating panels finds that people are more likely to become an informal worker if: 1) medical benefits covering are provided for the family as a whole by other member, so the worker is not the marginal provider of

this service; 2) administrative costs for social security are relatively high and benefits to worker are of low quality; and 3) high turnover rates within formal employment means that workers not accumulate segniority, and hence, have a lower opportunity cost in terms of segniority benefits.

Figure 3
Estimations on formal-informal wages, Mexico 2000-2003



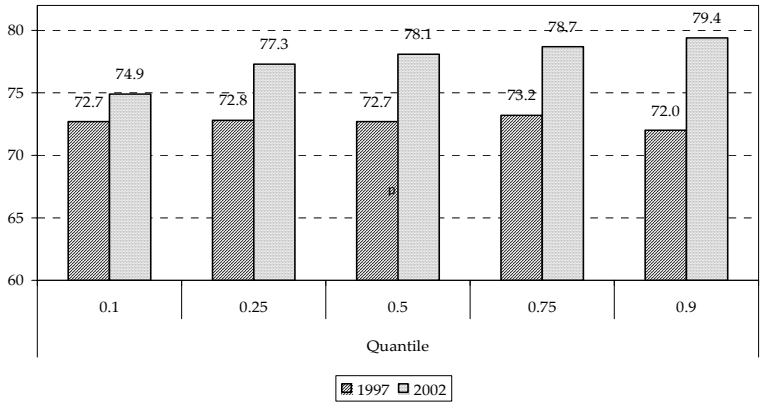
Source: WB staff calculations using ENEU.

In a recent report for Mexico, the World Bank (2004) devoted a small subsection to study the evolution of wage gaps between the formal and informal sectors. They defined `informal worker' as all those unprotected by labor law, but divided this large sector into two specific groups: 1) owners of firms with fewer than 16 employees who do not have social security or medical benefits, identified as "informal self-employed"; and 2) employees in those small firms, identified as "informal salaried workers". During 2000-2002, the report finds that there has been an increase in real wages for both sectors, followed by a flattening of informal and a fall in- formal wages in early 2003. Controlling for relevant characteristics such as basic economic background and education, their results show that the wage gap between informal and formal workers decreased along the conditional earnings (Figure 3 and Figure 4). In this report, they also find a reduction in the wage gap between 1997 and 2002, particularly in the lower tail of the conditional distribution (Figure 5).

The findings of the World Bank report for Mexico in 2002, fits in the history of Maloney for the rest of Latin America developing countries and the findings of Banxico in 2004. Basically, Figure 5 presents the wage distribution by formality category for the second quarter of 2002 in Mexico. This distribution has the skewness characteristic of the labor markets (Rosen and Neal, 1999), but in particular, once we consider the informal sector decomposition, the distribution of wages for the informal workers presents, relative to the formal sector: 1) lower wages for both types of informality; 2)

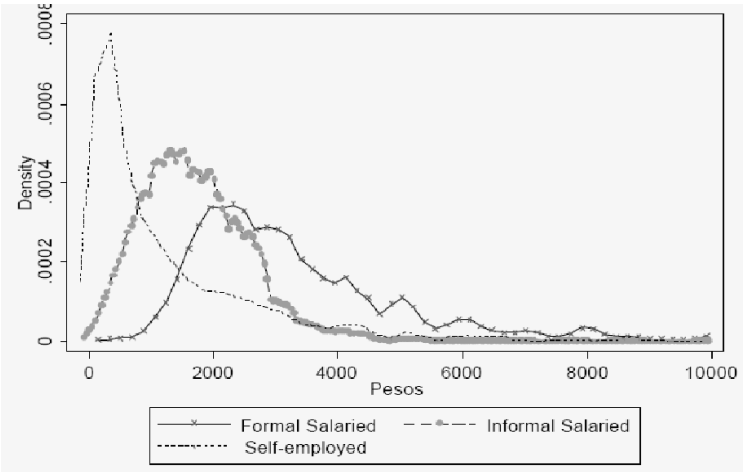
higher variance for both types of informality; and 3) higher skewness in the distribution.

Figure 4
Informal to formal wage ratio by quantile, Mexico 1997-2002



Source: World Bank, 2002. Staff Calculations using the third quarter of ENOE urban population on each year.

Figure 5
Labor income distribution for formal and informal categories, Mexico 2002



Source: World Bank, 2002. Staff Calculations.

The findings presented above are the basis of analysis for this proposal, where the objective is to identify the sources of the skewness of wages distribution and the evolution on wages gap by measuring the self selection and heterogeneity components, and the possibility of having sector specific shocks in productivity. The next section of this proposal presents the characteristics of the datasets I will use in my research, and then, I will describe the econometric model that I will work throughout the research project.

3. Data sets: the Mexican surveys of employment

For purposes of this analysis, I will use the set of surveys known as the National Survey of Employment for Mexico (ENOE). These datasets are collected by the Mexican National Institute of Statistics, Geography, and Informatics (INEGI) and are the basis for official employment statistics. Since 1987 and until 2005, the Surveys of Employment have had the same questionnaire information structure including information at individual level for potential workers (people with 12 or more years old) on variables such as: education, labor participation, hours worked, type of contract and benefit received, some questions about social protection coverage or labor force and other basic variables as gender, marital status and age. Moreover, it is also possible to recover the same information for all the members of the same household. Since its first application on the field in 1987, ENOE have had three different sampling designs: the first design was operative from 1987 to 2000; the second from 2001 to 2004; and the third stage beginning 2005.

The first sampling design divided their data collection into two types of surveys: the National Survey of Urban Employment (Encuesta Nacional de Empleo Urbano, or ENEU), with quarterly data collected since 1987; and the National Survey of Employment (Encuesta Nacional de Empleo, or ENE), with annual data collected from 1996 to 2000. Both ENE and ENEU surveys share the same variables and structure and were statistically representative at urban and national level respectively but with different time collection and sampling structure. On the one hand ENE was a cross sectional survey nationwide collected, which is representative for both rural and urban areas in Mexico. ENEU, on the other hand, focused only on urban areas[2] but with a main advantage over ENE: ENEU is a rotating panel which follows each round of workers for 5 quarters. In each ENEU sample there are five cohorts, each in a different state of completion in the interview cycle: one-fifth of the sample in its first interview, one-fifth in its last (fifth) interview, and three-

[2] This means that ENEU survey was restricted to areas with more than 100,000 inhabitants. The geographical coverage on ENEU has changed over time with the introduction on new urban areas and currently covers the 45 largest urban areas in Mexico.

fifths of the sample in intermediate stages. The ENEU conducts extensive quarterly household interviews in the 16 major metropolitan areas for from 1992 to 2000 and the sample is selected to be geographically and socio-economically representative. Additionally, a household identification variable permits construction of household variables for purposes of having more controls if needed.

In the second sampling stage, from 2001 to 2004 both ENEU and ENE disappear becoming one single nationwide survey: the Quarterly National Survey of Employment (Encuesta Nacional de Empleo Trimestral, or ENET). ENET combines the ENE representativeness with the ENEU rotating panel structure which allows following every individual for at most 5 quarters. Specifically, since the second quarter of 2001, the ENET consists of a rotating panel data representative for the next mutually exclusive aggregate levels: a) National level; b) 48 mayor cities in Mexico; c) 32 Mexican states; d) four levels of urbanization measured by community inhabitants. These multiple representativeness levels allow us to divide the sample for analyzing even labor markets for rural and urban areas.

Finally, since 2005 ENET was the basis for the Encuesta Nacional de Empleo y Ocupaciones (ENOE) by modifying its questionnaires which started to be operational at that year. Using all the ENET row data sets from 2000 to 2003 with information at individual level, I will build measures on wages, education, and sectorial composition, by focusing my analysis in the urban labor markets. From there, I will focus my attention on analyzing wage structure comparing sectors, but differentiating between males and females, given that the variable "number of children", one of the most important variable in determining labor participation, is only available for women

4. Labor market Equilibrium in a two sectors economy

This section presents the model I will use to characterize the wage distribution in both formal and informal sectors, and compare the hypothesis of wage differences under alternative structural measures of switch effects. Initially, I will assume partial equilibrium analysis, trying to integrate the elements from both labor supply and demand, and from there, recover the wage distribution of the workers for both formal and informal sector.

For this purpose I will start by characterizing the possible sources of heterogeneity in the observed wage of each market in two sources: 1) A first source is due to the tax rate imposed on one sector relative to the other; and 2) the self selection of the agents between sectors which may induce a sorting bias on the observed wages for each sector.

4.1. A model of labor demand with taxes

This subsection provides a quick and powerful result on labor demand for two sectors with different tax structure. For simplicity, let us assume for a while that there are two sectors in the economy: F and I. One of them, without loss of generality F, is required to pay a wage tax τ per unit of time hired in labor while the sector I do not require to pay that tax[3]. Accountability of tax is at this point exogenous to the sector and this tax is returned to the agents in the form of lump-sum transfers, so they do not provide additional distortions on the economy.

Assuming there is competition on the labor market, then for each firm the optimal conditions for maximizing profits are equalize the value of the marginal unit hired to the wage paid to workers. In this case, this condition for sector F is given by:

$$P^F L^F_L (K^F, L^F) = (1-\tau) W^F = w^F \quad (1)$$

where w^F is the observed wage on the sector F. Following the same argument, the optimal condition for sector I is given by:

$$P^I L^I_L (K^I, L^I) = W^I = w^I \quad (2)$$

For w^I being the observed wage on the sector I. Hence, on assuming technology on each sector is differentiable, the last worker hired in the sector should provide the same value of marginal productivity for each of these sectors. Then, in any equilibrium allocation must be true the following condition:

$$P^F L^F_L (K^F, L^F) = P^I L^I_L (K^I, L^I) \quad (3)$$

which implies in terms of wages:

$$w^F / w^I = (1-\tau) \; W^F / W^I \quad (4)$$

This simple condition tells us a very powerful history: in any free mobility labor market, the ratio of the observed formal to informal wages may obey to two sort of differences: 1) one associated to differences in the "net wage" ratio paid in each sector, (W^F / W^I); and 2) a second difference associated to the tax rate faced by the formal sector market ($1-\tau$).

[3] The ability to avoid the tax is not endogenously determined in this first approach model. This alternative analysis may be considered in a future research.

This first approach model would suggests that a very important component for explaining the observed wage differentials between formal and informal wages is indeed the tax rate and the economic logic goes as follows: provided the value of the marginal productivity of the last unit of labor is equal on each sector, then the relative wages observed on each sector must be equal to the average tax rate paid on the last marginal unit of labor hired.

Nevertheless, there exists strong evidence suggesting net wages may indeed vary between sectors, particularly if the bundle of skills and the price per unit of skill are different in each sector. This argument strikes the fact that the expected wage, conditional on skills-education on each sector, are biased due to the sorting of agents trough a self selection choice. This is the second brick of my analysis on wage differentials provide a guide to analyze wage evolution, but several question remain to be studied particularly those regarding to the average and marginal agents. From the Roy model perspective, exist the possibility that the average worker at each sector is essentially different from that worker which is on the margin of indecision on being or not formal. In a context of heterogeneous outputs on each sector, for the same skill and abilities characteristics, this lead us to self selection bias on the conditional average wage.

Therefore, to study the fully wage distribution differences between formal and informal sector, and given the lack of pure counterfactual for each agent, I will analyze the switching effect of formality using the engineering elements of Treatment Effects developed in Heckman and Vyvlacil (2006) and studied in detail in the next section.

4.2. The labor supply side: characterizing sorting between sectors

This section presents the model proposal for studying the characteristics of sorting model of labor supply markets. By studying the decision of the marginal agent for being at the formal and informal sectors, we are able to fully characterize the aggregate distribution of wages across sectors and study if the distribution of agents characteristics are different among sectors (sorting effect by self selection) or if the agents consider informal sector to be just an alternative to the formal one which, at the margin, may gives them the same (expected) wage.

The methodology I propose to study then integrates the self-selection nature of micro decisions into the sorting macroeconomics framework, and from the nature of these two complementary approaches identify the aggregate wage distribution which in principle would depend on two type of essential factors: 1) those underlying the essential parameters of the

production function on each sector; and 2) those related to the (given) distribution of heterogeneous essential characteristics of the agents.

Hence, the model can identify decompose the effects of business cycles shocks over the wage distributions and identify how much of the observed changes in wage distributions between and across sectors are due to productivity shocks, and how much are due to sorting of the agents once their expectation on wages on wages are fulfilled.

The model begins with the traditional sequential optimization problem where the agent i maximizes the expected net present value of his consumption in a time-separable utility function. Hence the objective function is of the form:

$$\max \; E\left[\sum_{t=0}^{\infty}\left(\frac{1}{1+\rho}\right)^{t} U(C_{i,t})\right] \qquad (5)$$

Let us assume that, as in the traditional sorting models, the income of the agent comes from his labor wage, and that the utility he earns in consumption is linear in this argument, hence the objective function becomes:

$$\max \; E\left[\sum_{t=0}^{\infty}\left(\frac{1}{1+\rho}\right)^{t} w_{i,t}\right] \qquad (6)$$

provided he decides to supply inelastically one unit of labor per period if accepts the wage offer he has at hand[4].

At the beginning of each period, the agent must decide in which sector j he wants to work. At this point I will assume that there are two sectors A and B which differences will be explored in detail below, this in order to gain the insights about more general sector possibilities[5]. The agent decides in which sector to work by knowing a wage offer which be paid at the end of the period with two components: one known by the worker and which depends on his characteristics, and a second component which is stochastic but for which distribution is known for the agent.

---

[4] In a future research project I will plan to explore the possibility of having both consumption and leisure in the utility function, and so having variable labor supply in a sorting framework.

[5] In this particular case, I will analyze empirically the implication on "formal" and "informal" sector by characterizing the parameters β and the distribution of shocks on each sector.

Hence, the wage the agent obtain $w_{i,j,t}$ will be modeled to have two components: 1) a piece-rate wage conditional on some observed and common known characteristics of the agent $X_{i,t}$ which prices by sector and period $\beta_{j,t}$ are determined by the productivity essentials of the sector; and the second component $u_{i,j,t}$ which represents some random component, in principle different on each sector, reflecting some uncertainty for working on a particular sector. This uncertainty is hence exogenous to the agent, but the distribution of these shocks and its parameters are known for the agent ex ante. Hence the effective wage the agent receives at the end of the period is a linear function of the form:

$$w_{i,j,t} = X_{i,t}\beta_{j,t} + u_{i,j,t} \qquad (7)$$

where $\beta_{j,t}$ is the price rate paid at sector $j \in \{F,I\}$ for each of the observed skill characteristics X of the agent i at period t (which in principle may be time-dependent); and the error term $u_{i,j,t}$ shows the sector-individual specific random term to the period of time t. I will assume that the two terms are non-correlated, so the productivity shocks are independent of the heterogeneous characteristics of the agent. For purposes of my analysis, I will assume that skills are characterized by the years of education of the worker, and a series of dummy variables that may capture the different yearly rates of return to different levels of education.

From equation (7), let me notice one very important feature for this model: there are two different types of wage distributions relevant for our model: 1) that wage distribution which is relevant for the decision of the agent; and 2) the wage distribution for the whole market which result from the sorting and self selection of the agents across sectors.

At an individual level, the wage distribution from which the agent decide where sector to work relies on the distribution of the sector shocks $\{ u_{i,F,t} , u_{i,I,t,}\}$ and the productivity-prices of their heterogeneous abilities $\beta_{j,t}$. At this stage I will assume that the skill prices are known, and at sometime constant, so all the source of randomness on the wage distribution for the agent are result of the sector specific random shocks.

At this stage, and following the McCall original model as in Sargent and Lvunquist (2004), I will assume that both sector productivity shocks are draw from a jointly i.i.d. normally distributed on time, equal for all the agents $i \in \{1,...,I\}$, and these shocks are independent of the agent i heterogeneity $\forall\ i \in \{1,...,I\}$. Hence the shock distribution for each sector is of the form:

$$\begin{pmatrix} u_{F,t} \\ u_{I,t} \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_F^2 & \sigma_{FI} \\ \sigma_{FI} & \sigma_I^2 \end{pmatrix} \right\} \quad (8)$$

Then, using these assumptions, the distribution of wage across sector is indeed a random variable whose joint distribution at each period is characterized as follows:

$$\begin{pmatrix} w_{i,F,t} \\ w_{i,I,t} \end{pmatrix} \sim N \left\{ \begin{pmatrix} X_{i,t}\beta_{F,t} \\ X_{i,t}\beta_{I,t} \end{pmatrix}, \begin{pmatrix} \sigma_F^2 & \sigma_{FI} \\ \sigma_{FI} & \sigma_I^2 \end{pmatrix} \right\} \quad (9)$$

Hence, while the expected wage of the agents relies on their abilities, the variance of the wage will depend on the stochastic term shocks $u_{i,j,t}$.

Given the known bi-normal distributions, and provided the distribution of heterogeneous characteristics of $X_{i,t}$ we are able to find the whole characterization of the value function and how, given these variables, the agent have two optimal policy functions of the "reservation wage type": one is given by the reservation wage of working or not, $w_{i,t}^{o}$, and the second one identify the reservation wage for switching or working in one particular sector, $w_{i,t}^{S}$. In particular, we will observe that the agents decide to work if:

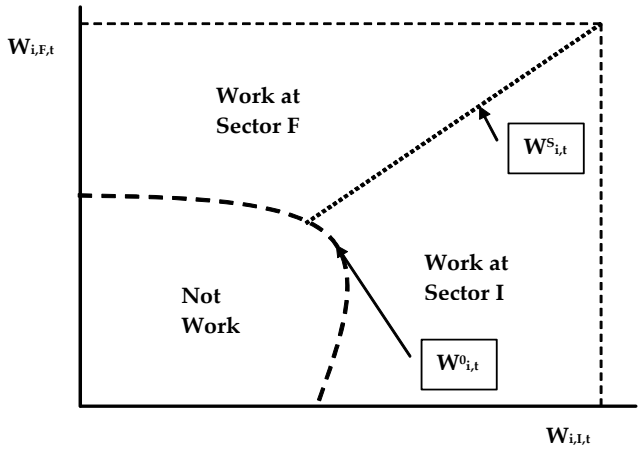
$$min \ \{w_{i,F,t}, w_{i,I,t}\} > w_{i,t}^{0} \quad (10)$$

On the other hand, once the worked has decided to work, the agent will decide to work in a sector (let me use without loss of generality sector F as the basis) or switching to this sector if the following condition occurs:

$$w_{i,F,t} > w_{i,t}^{S} \quad (11)$$

otherwise the agent will stay at the sector originally selected. Graphically, the model expects to find some functional decision forms for the policy function as Figure 6.

Figure 6
Policy functions characterization for an agent in a two sector economy



From the full wage distribution, conditional on observed skill characteristics, we identify if such a wage distribution is then different among sectors, and how much of such a difference is explained only by differences on the characteristics that determine formal sector decisions, and how much are only due to the tax rate applied for that agent. Moreover, these distributions also help us to analyze the determinants of the "size" of the sectors.

At this stage of this research proposal, I will focus on studying the marginal values of switching between sectors given the agent has decided to participate, so I will only study those agents which are employed in the labor market and the analysis of multichoice switching decisions and effects between being out of the labor force are left for a future research.

Then, for a worker which has decided to work, the sequential problem can be reduced to a set of functional equations, which Bellman representation for each given sector is given by a recursive equation showing the value of being employed at that sector at the current wage offer:

$$V_{i,j,t(w_{i,j,t})} = \max\left\{w_{i,j,t} + \beta E_t\left[V_{i,j,t+1}\left(w_{i,j,t+1}\right)\right], \beta E_t\left[V_{i,t}\left(w_{i,A,t+1}, w_{i,B,t+1}\right)\right]\right\} \quad (12)$$

For $j \in \{F, I\}$

Given the agent has decided to participate in the labor force[6], each period he faces the decision of entering or not to a particular sector. In this case, this decision can be modeled using the latent variable approach associated to select "being in the formal" sector. Each period the agent has to decide which sector to work in given he has a wage in hand at each sector. Let us define the value function the agent i may have for being at sector j as $V_{i,j,t}$ In particular, In this document I will analyze the case for two sector such that j∈{F=Formal, I=Informal}.The agent then has the following optimal decision conditional on his characteristics X and so:

$$V_{i,t}(w_{i,F,t}, w_{i,I,t}) = max\{V_{i,F,t}(w_{i,F,t}) \ , V_{i,I,t}(w_{i,I,t})\}$$

Then, provided the value of being at each specific sector, the agent decides which sector to work. Without loss of generality, let me assume $V_{i,j,t}$ is represented by a linear function of some observed characteristics for the econometrician $Z_{i,t}$ and some non-observables for the econometrician $e_{i,j,t}$ as follows:

$$V_{i,j,t} = Z_{i,t}\pi_{j,t} - e_{i,j,t} \qquad (13)$$

I will assume at this stage that the decision the variables that characterize the choice of sector have two orthogonal components so $Y_{i,t} \perp e_{i,j,t}$, and also $\pi_{j,t} \perp e_{i,j,t}$ avoiding the possibility of having random coefficients on the choice equation.

In this case, the sector selection criteria for agent i is given by a latent variable $\Upsilon_{i,t}$ that measures the net benefit of being at the formal (or informal sector). The index variable $\Upsilon_{i,t}$ is given by the following expression:

$$\Upsilon_{i,t} = V_{i,F,t} - V_{i,I,t} = Z_{i,t}\pi_{i,F,t} - Z_i\pi_{i,I,t} - (e_{i,F,t} - e_{i,I,t}) \qquad (14)$$

[6] At this first stage, I will work only with those agents who are already in the labor market, and not developing the possibility of multi-stage self selection were the agents can choose working or not, and then which sector work on. On a later stage of the research I will include this possibility by using a multinomial decision to correct this multilevel self selection possibility.

or grouping terms,

$$\Upsilon_{i,t} = Z_{i,t}(\pi_{i,F,t} - \pi_{i,I,t}) - (e_{i,F,t} - e_{i,I,t}) \quad (15)$$

If we simplify and collect terms, the selection variable is then given in terms of the variable:

$$\Upsilon_{i,t} = Z_{i,t}\Pi_{i,t} - \xi_{i,t} \quad (16)$$

In other words, the agent will decide to work in the formal sector if the utility value from being in the formal sector is higher than its second option, otherwise he will optimally chose to go into the informal sector. Then the selection rule in the Willis and Rosen sense given by the variable $\Upsilon_{i,t}$ where the agent will select being at the formal sector if and only if $\Upsilon_{i,t} > 0$, while he would choose go to the informal sector otherwise. Then, the probability of observing an agent in the formal or informal sector, or propensity score for each agent, is therefore defined in terms of the index variable $\Upsilon_{i,t}$ as follows:

$$P_{i,t}(j = F) = P_{i,t}(V_{i,F,t} - V_{i,I,t} > 0) = P_{i,t}(Z_{i,t}\Pi_{i,t} - \varepsilon_{i,t} > 0) = P_{i,t}(\gamma_{i,t} > 0) \quad (17)$$

$$P_{i,t}(j = i) = P_{i,t}(V_{i,F,t} - V_{i,I,t} \leq 0) = P_{i,t}(Z_{i,t}\Pi_{i,t}\varepsilon_{i,t} \leq 0) = P_{i,t}(\gamma_{i,t} \leq 0) \quad (18)$$

Hence, the observed wages have a self-selection component associated to the decision of being or not part of the formal or informal sector, which can be modeled in the same spirit of Willis and Rosen (1979) using the Roy model framework exposed in Maddala (1983). Assuming the characteristics associated to the participation decision on formal sector $\xi_{i,t}$ is normal distributed, we can simplify using Heckman (1979) error decomposition and the truncation normal in the following reduced form:

$$w_{i,F,t} = X_{i,t}\beta_{F,t} + \sigma_{F,t}\rho_t^F \lambda_{i,t}^F(c) + \varepsilon_{i,F,t} \quad (19)$$

$$w_{i,I,t} = X_{i,t}\beta_{I,t} + \sigma_{I,t}\rho_t^I \lambda_{i,t}^I(c) + \varepsilon_{i,I,t} \quad (20)$$

$$\lambda^F(c) = -\frac{\phi(c)}{\Phi(c)} \quad (21)$$

$$\lambda^I(c) = \frac{\phi(c)}{1 - \Phi(c)} \tag{22}$$

$$c = \frac{Z_{i,t}\Pi_{i,t}}{\sigma_{\xi_{i,t}}} \tag{23}$$

$$\rho_t^j = \frac{Cov(\varepsilon_{i,j,t}, \xi_{i,t})}{\sigma_{\xi_{i,t}}\sigma_{j,t}}, \quad j \in \{F, I\} \tag{24}$$

For purposes of this analysis, I will assume the set of equations satisfy the exclusion restrictions on the variables X_i and Y_i in order to identify the set of relevant parameters. Once the model have included the nature of the selectivity of the agents to participate in each type of the labor market, the possible sources of wage variability associated with the heterogeneity of the agents is key for answering several questions on policy effects from formalization.

4.3. Wage differences and the formalization treatment effects

This section develops the following policy question: what would be the expected benefit from "formalizing"[7] an informal worker? Moreover, is there any wage gain from moving a worker from the formal to the informal sector? In particular, if the benefits are fully characterized by the wage the agent earns in each sector conditional on his education, once we consider the self-selection bias, what would be the impact of switching agents between sectors?

A useful tool for answering this sort of questions is the set of treatment effects corresponding to this experimental simulation. In particular, following Heckman and Vytlacil (2006) we can study at least 4 particular treatment effects, and form these estimations and controlling by self selection, infer the impact on the expected effect from switching across sectors.

Assuming the outcomes (wages in this case) depends on the same set of variables X_t in the two sectors, and that the coefficients on the outcome equation are the same for all the agents (i.e. at this stage I will not permit

[7] By formalization I mean moving a worker, keeping his education and background, from an informal sector to be a formal wage worker.

heterogeneity on the returns of the variables X_t) then the first treatment effect is given by the ordinary least squares difference ΔOLS_i conditional on this set of observables in the two outcomes. This treatment is given by the following expression:

$$\Delta OLS(w_{i,j,t}|X_{i,t}) = (\beta_{F,t} - \beta_{I,t})X_{i,t} \tag{25}$$

Given the nature of the sorting and self selection of the agents across sectors, it is possible that the previous effect does not reflect the average gain for a worker from switching from formal to informal, hence, an unbiased treatment estimators is proposed to be the average treatment effect ΔATE given by the following definition:

$$AUT(w_{i,j,t}\Big|X_{i,t}) = E(w_{i,F,t} - w_{i,I,t}\Big|X_{i,t}) = (\beta_{F,t} - \beta_{I,t})X_{i,t} - (\rho_t^F \lambda_{i,t}^F(c) - \rho_t^I \lambda_{i,t}^I(c)) \tag{26}$$

In this second treatment effect, the control function to address the self selection bias is given by $\rho_t^F \lambda_{i,t}^F(c) - \rho_t^I \lambda_{i,t}^I(c)$ and represents the correction for this average treatment.

Finally, two other treatment effects may be interesting to consider namely: the treatment on the treated ΔTT_i , and the treatment on the untreated ΔTUT_i defined as follows:

$$ATT(w_{i,j,t}\Big|X_{i,t}) = E(w_{i,F,t} - w_{i,I,t}\Big|X_{i,t}, D_i = F) = \Delta ATE(w_{i,j,t}\Big|X_{i,t}) + \sigma_\varepsilon \rho_t^F \lambda_{i,t}^F(c) \tag{27}$$

$$AUT(w_{i,j,t}\Big|X_{i,t}) = E(w_{i,F,t} - w_{i,I,t}\Big|X_{i,t}, D_i = I) = \Delta ATE(w_{i,j,t}\Big|X_{i,t}) + \sigma_\varepsilon \rho_t^I \lambda_{i,t}^I(c) \tag{28}$$

In a environment where free mobility across sector occurs, such as in the model of labor demand studied in the previous section, we would expect that conditional on the agents characteristics the net wages in the two sectors, must be the same, or quite similar, unless that there are sector specific skills and different prices per unit of skill. This result may be challenged by the observed wage differences of the previous studies. The next section presents the definition of formality I will follow using ENET, and the empirical implementation of the models proposed to show the estimations regarding the set of treatment effects for the Mexican labor market.

5. Empirical implementation

To study the wage differences between formal and informal sector in the Mexican labor market, I will use all the information provided by the ENET and from the system of equations (19) to (24) defined in the previous sections I will build and analyze the different unconditional and conditional treatment effects, corrected and not by self-selection. In this case, I used the quarterly databases from 2000 to 2003 to study the evolution of wages and their estimations for the treatment effects already defined, separating each sample between men from women, as this second group includes key information not provided for the first group.

The first important question is the definition of formality and informality I will use for the rest of the document. In particular, using the information of ENET questionnaire and an approach close to the one exposed in Banxico, I will infer and define formality in terms of the labor contractual specifications, which are close related to the payment of labor income taxes. In particular, for purposes of this analysis, I propose the following definition:

Definition: An informal employed worker is a person who satisfies simultaneously the following four conditions:

i) Works for a positive earning compensation for her (his) work at the labor market during a positive number of hours per week.

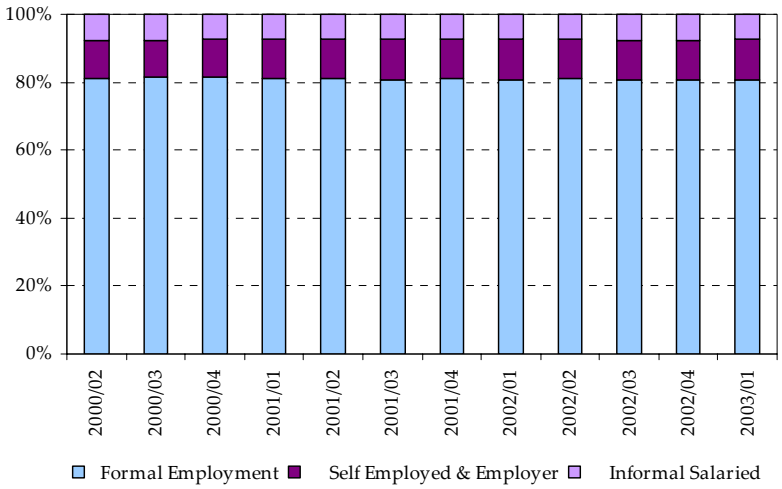
ii) Does not receive at least one of the following benefits mandatory by law: IMSS, ISSSTE, INFONAVIT; SAR, or Private Medical Insurance.

iii) Works in an economic unit (firm) without official registration;

iv) The economic unit (firm) where she (he) works does not have more than 50 employees, including the employer.

This informality definition includes three main informal employment categories: self-employed (one-worker self managed firm), informal employers, and informal salaried worker. The main difference between the definition of formality of this proposal and the previous definitions studied by the World Bank and Banxico is the inclusion of the condition of "official registration of the economic unit" (the third condition in my definition of formality.) While this question is part of the ENET basic questionnaire, the previous definitions of formality ignores it. At this stage, I will follow this alternative definition as one contribution to the discussion over the previous analysis, and I will let the possibilities of studying the robustness of my results for different measures of formality for a future research.

Figure 7
Employment composition by formality, Mexico 2000-2003



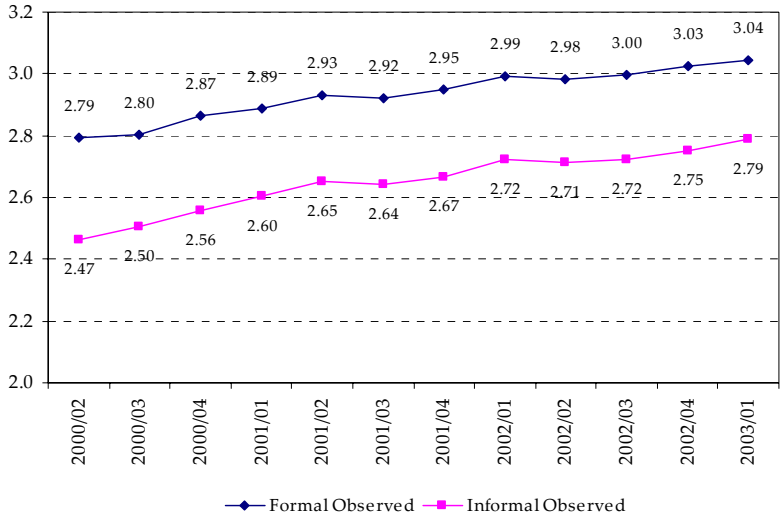
ENET datasets for each quarter provides us information on basic socioeconomic background for more than 120,000 observations of people who potentially participates in the labor market. From there, I used the people who are currently working at the labor market at each quarter I had information, resulting in more than 85,000 observations per cross section quarter. Using this data and the definition proposed, the composition of formality for 2000 to 2003 shows the relative size of the informal sector has been steady and around 20 per cent of the total labor participation. This estimation differs from the size studied in the other analysis due to the alternative definition I am proposing, and the sources of difference are the inclusion of official registration as a filter of informality. Following my definition, the main source of informal employment is given by the "self-employed and employer" category, which represents around than 65 percent of the informal employment. Figure 7 shows the composition and evolution of employment by category for Mexico in the period studied, while Table 1 in the appendix shows more precise estimations by gender.

The second question is the evolution of average wages (labor income per hour) for both sectors. For this purpose I decomposed the databases for studying the wage evolution by gender to use the different information regarding the number of children they have (women has explicitly this question, while information on men is not reported.) ENET databases shows the net of taxes labor income of the workers, so, by using the number of hours worked per week, I can estimate the wage-per-hour for each worker net of taxes, and from there, estimate the log-wage which will be my

"dependent variable". For both men and women, the average log-wages shows a steady but low increase in real terms for both formal and informal markets. Nevertheless, the most important feature of having log-wages is that we can compare the percentage gap between formal and informal sector. This unconditional average wage difference has been reducing for both men and women from 33 percent and 54.2 percent in 2000/02 respectively, to 25.3 percent and 45.3 percent in 2003/01. Figure 8 and Figure 9 present these findings in more detail.

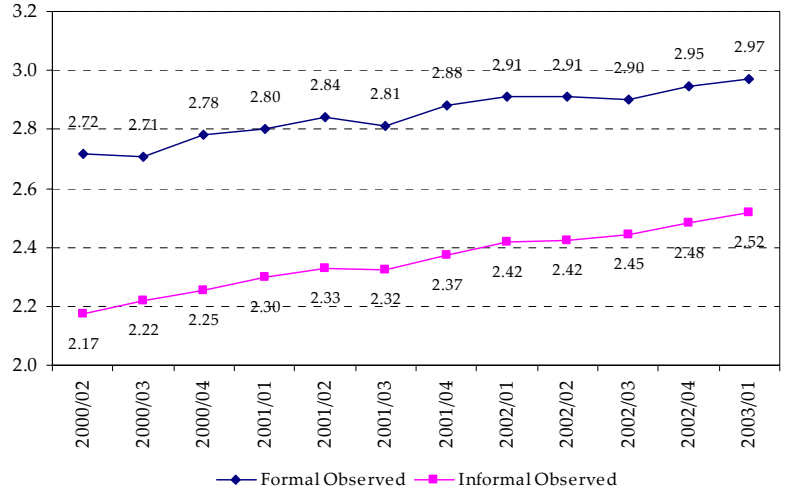
Once the time series of wages are estimated by gender, the next question is what are the sources of these wage gaps, and particularly if such a gap persists once we consider the self selection component of the agents which may bias the unconditional differences and the ordinary least squares differences. These unbiased effect estimations are precisely the treatment effects the next section is focused in gathering.

Figure 8
Average log-wages for men by formality, Mexico 2000-2003



Source: Own estimations using ENET 2000-2003.

Figure 9
Log-wages for women by formality, Mexico 2000-2003



Source: Own estimations using ENET 2000-2003.

5.1. Formal-informal wage system of equations

This section presents a brief analysis of the results on the estimations for the self selection wage equations using the two stage procedure described in Section 4. In a first stage of estimation, I characterize the choice of formality and from there infer the full wage distributions for both formal and informal sectors identifying the wage counterfactual of each worker. In the second stage, I will use these wages distributions to build the different treatment effects from "formalizing" and "un-formalizing" a worker conditional on his or her characteristics.

Following the basic statements of human capital theory, I used as independent "outcome variables" (i.e. independent variables in the wage equations) a set of measures which could persist independently of the sector in which the worker is, and at the same time, which are close related to the level of human capital the worker possesses. Hence, I included the labor experience[8], experience squared, and a set of seven dummies for controlling rates of return for different levels of education: less than basic education, basic education, lower secondary education, high school, technical education, professional, and graduate. Each set of models were estimated by gender to capture the information on number of children in the case of

[8] Experience is traditionally measured as “age” minus “years of education” minus “6”.

women. On the other hand, I used a larger set of variables to model the participation decision choice including those related to human capital already defined in the wage equation and new information among which I included: the marital status, the number of children (for women), the status in the house of the worker (head of the household or not), and one dummy for each state (32 in total) using Mexico City as exclusion base for these dummies.

The set of estimations for the equation of participation for both men and women are included in Table 2a and Table 2b in the appendix to this document, for each of the cross section studied from 2000/02 to 2003/01. All of the participation models were statistically significant using the LRχ^2 criteria.

For both men and women, participation in the formal labor market is consistently negative related to the experience variable but positively correlated to the dummimes for level of education. This would suggest that people with higher level of human capital are less likely to be an informal worker, but experience measured in years correct downward the rates of return of education. In any case, the coefficient associated to this variable is rather small, and the effect of the level of education captured by the dummies is high enough to measure the rates of return by level of education in the wage effect. Also, for both men and women the probability of being in the formal sector increases if the worker is head of the household, which could reflect the fact that people with more responsibilities may prefer being at the formal sector looking for the protection of health and social security system mandatory by law. This last result also reinforces for the case of men in which the probability of being formal increases between 3 to 5 percent if he is married. This hypothesis nevertheless has a counterpart with the positive and significant coefficient that number of children has in the probability of being in the formal sector for women, where each additional children reduces the probability of being at the formal sector between 0.5 and 1 percent. The set of dummies related to the Mexican states identification in the survey shows mixed effects on their signs and statistically significance, but in the majority of the cases are statistically significant and modifies the probability of being formal with respect to Mexico City (Distrito Federal) in some cases for 10 to 11 percent. This heterogeneity in significance and size of the effects of each state may reflect in a broad way the differences in the development of the informal sector in each state relatively to Mexico City.

In order to estimate the formal and informal wage equations, I considered as outcome variables only the set of information regarding the human capital of the worker. I estimated the Mills ratio for each sector using the participation choice from the probit estimated in the first stage, and in the

second stage I estimated the two sets of equation for each gender, and for each quarter studied. The full set of estimations is presented in Tables 3a to 4b in the appendix of this document. It worth to notice that the coefficient in these set of equations does not have any direct interpretation as returns to education as in the traditional Mincer equation, given that the interest of the model is to estimate the expected treatment effects of switching from one sector to the other. Rather, at this estimation stage the properties we are interested in studying for the model are: 1) the statistical significance of the coefficients, 2) the difference in the coefficients which are the unbiased estimation of the difference in the effect of each independent variable from switching, and 3) if the self-selection variable explains some of the variability in wage equation for each sector.

In all of the models, all of the variables considered in each of the wage equations turned to be significant and with the expected sign in experience and experience squared, but as noticed before, some negative and significant signs on basic education level dummies. For this case, the inclusion of the Mills ratio, which turns to be significant in all of the models for both men and women, is the main cause of such changes. In any case, the analysis of the difference in these coefficients is the source of interest for our analysis and will be explained in details below. The fitted properties of models (R-square) for the formal wages are between 0.2989 and 0.3147 for men, and 0.3394 and 0.3705 for women. In the case of informal wage equation the R-square is around 0.0945 and 0.1168 for men, and 0.0876 and 0.1227 for women.

The unbiased estimations of the difference in the coefficients, corrected by self-selection, are interesting in the sense that they show some important effects of being or not a formal worker over the wage of the worker, conditional on the level of education. Figure 10 presents these estimations.

In general, for the 2000/02-2003/01 period, people with higher levels of education have a net gain from switching to the formal sector. Nevertheless, there is another huge impact by levels of education which turns to be the most important effect for purposes of this analysis. In the case of men, workers with less or equal education equivalent to high school would earn significantly more in the informal sector than in the formal sector, while workers with technical education, college, or graduate studies, have a higher return on wage in the formal sector compared to the one they obtain in the informal one. For women, the difference in returns to education by level are negative from switching from informal to formal for basic school and lower secondary, and turns to be positive for any education level higher or equal to high school education. What turns to be large for both men and women is the returns to education of college and graduate studies from switching, which

means that people with higher levels of education would earn even more in the formal sector relative to the informal.

More analysis could be done in this area, but this analysis goes beyond the purposes of this document at this stage. The next section shows the estimations of the different policy treatment effects from switching from informal to formal, and reveals interesting evidence on wage differences for these sectors.

Figure 10
Average conditional wage differences between formal and informal sector, Mexico 2000-2003

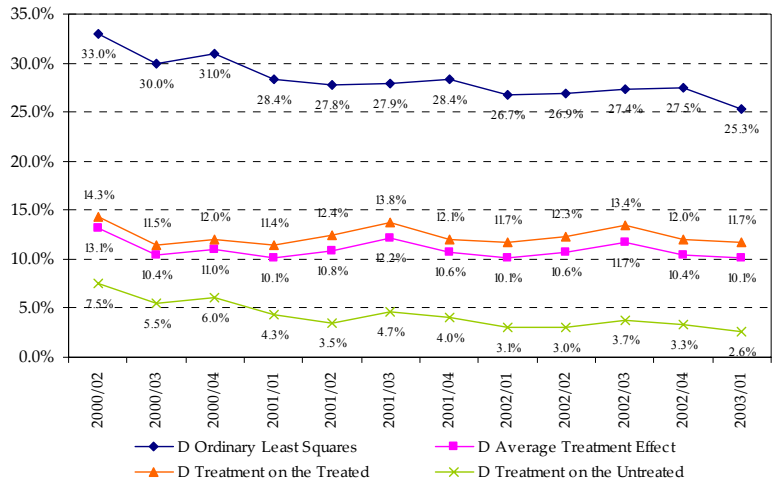
Variable	2000/02	2000/03	2000/04	2001/01	2001/02	2001/03	2001/04	2002/01	2002/02	2002/03	2002/04	2003/01
Men												
Experience	0.34%	0.27%	0.45%	0.21%	0.12%	0.29%	0.13%	0.42%	0.35%	0.13%	0.42%	0.38%
Experience Squared	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%
Basic School	-14.83%	-18.69%	-13.07%	-13.33%	-10.18%	-9.30%	-13.17%	-16.09%	-8.00%	-10.21%	-15.07%	-7.60%
Lower Secondary	-13.13%	-21.46%	-9.38%	-17.64%	-9.86%	-12.28%	-14.39%	-18.17%	-7.77%	-11.77%	-18.24%	-6.39%
High School	-4.51%	-14.63%	-1.06%	-12.25%	-0.57%	-3.60%	-2.80%	-7.00%	3.28%	-2.62%	-9.63%	6.84%
Technical Education	6.26%	-8.37%	2.85%	3.05%	12.23%	6.63%	2.16%	4.13%	13.55%	5.66%	0.08%	16.51%
College	45.05%	-25.67%	42.94%	24.13%	41.16%	41.82%	33.99%	34.05%	53.05%	39.13%	40.27%	57.83%
Gaduate Studies	84.17%	80.22%	6230%	77.62%	76.11%	156.55%	92.52%	95.00%	151.42%	157.98%	89.52%	106.39%
Women												
Experience	0.56%	0.12%	0.83%	0.33%	0.35%	0.55%	0.72%	0.61%	0.50%	0.54%	0.64%	0.635
Experience Squared	-0.01%	0.00%	-0.01%	0.00%	0.00%	0.00%	-0.01%	0.00%	0.00%	-0.01%	-0.01%	0.00%
Basic School	-11.93%	-21.39%	-20.94%	-29.48%	-9.10%	-18.03%	-12.73%	-14.07%	-17.79%	-12.91%	-6.46%	-9.94%
Lower Secondary	-1.62%	-10.87%	-16.83%	-27.07%	-5.25%	-15.73%	-1.15%	-4.24%	-17.42	-2.19%	-0.67%	-3.41%
High School	19.05%	12.77%	5.98%	-11.40%	21.43%	2.22%	19.35%	21.67%	7.41%	18.28%	19.63%	16.41%
Technical Education	33.99%	26.11%	16.83%	4.71%	34.34%	13.15%	31.15%	32.54%	19.05%	36.44%	39.85%	29.64%
College	77.83%	83.80%	58.78%	40.80%	85.37%	52.27%	81.96%	92.37%	44.84%	81.40%	82.96%	77.93%
Graduate Studies	19.01%	0.15%	133.83%	41.34%	55.02%	92.20%	192.34%	66.26%	74.55	234.02%	173.21%	159.31%

Source: Own estimations using ENET 2000/02-2003/01.

5.2. Treatment effects estimations

As it was discussed in Section 4 of this document, the purposes of this analysis is studying four policy treatment parameters for a worker from switching between informal to formal sector. After doing the TSLS estimation and the correction for self selection on each wage equation, I estimate the treatment effects for each quarter and each gender. Figure 11 and Figure 12 below resume the treatment effects for comparing the wages for formal and informal sectors, dividing the analysis between men and women. The punctual estimations of these parameters are included in Table 5a and Table 5b in the appendix of this document.

Figure 11
Effect from switching informal to formal, men 2000-2003



Source: Own estimations using ENET 2000/02-2003/01.

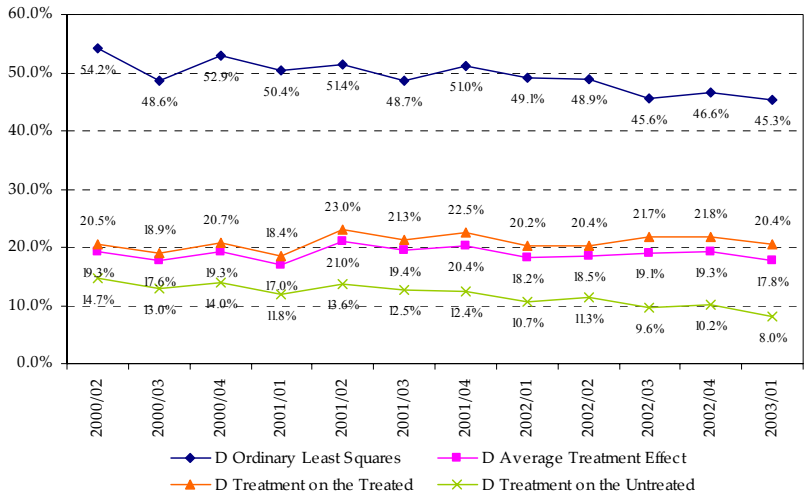
The OLS treatment effects resemble the unconditional average differences of wages observed in the previous section for both men and women. In this case, the OLS wage difference between formal and informal sector has been steadily reduced since 2000 for both men and women. Particularly, for this period the OLS treatment for men has reduced from 33.0 to 25.3 percent and for women has reduced from 54.2 to 45.3 percent. Nevertheless, as I will discuss below, the average wage difference seems to be dramatically different once we consider the self selection component and take into account this source of bias in the other conditional policy treatment effects.

The second policy effect considered is the ATE, defined as the average expected difference of wages once we identify the source of bias in the error terms of wages through the correspondent control function or Mills ratio. For the period studied in this document, ATE estimations are largely different with respect to the unconditional mean of wages and the OLS treatment estimators. For instance, for men in 2000/02 wage formal-informal gap reduces from 33.0 percent in OLS estimation, to 13.1 percent in the ATE difference. This results are qualitative similar to both men and women. Like the OLS and the unconditional measures, ATE also has been reducing for both men and women, from 13.1 and 19.3 percent in 2000/02, to 10.1 and 17.8 percent respectively.

The evidence suggests that, sorting across sectors indeed plays an important role in explaining the difference between wages, and that the observed wage

gap may obey to other economic causes rather than low productivity of the informal sector labor units per se. ATE also shows reductions for the period studied to reach a 11.7 percent for men, 20.4 percent women, for the last quarter studied.

Figure 12
Effect from switching informal to formal, men 2000-2003



Source: Own estimations using ENET 2000/02-2003/01.

The third policy parameter considered is the TT which shows the benefits of being formal, conditional on being a formal worker. For men, this parameter suggest that formal workers would have earned 14.3 percent less if they have switched from formal to informal in 2000/02, but the loss of switching reduces to 11.7 percent by 2003/01. For women, the loss from switching (or gains for being formal compared to be informal) have remained stable and around 20.4 percent.

The fourth policy parameter TUT shows the effect from switching from informal to formal sector, conditional on being informal employee. This effects turns to be the most revealing of the parameters estimated in this analysis. For men, the TUT estimations was of 7.5 percent in 2000/02, which dramatically contrast with the unconditional wage differences and the OLS treatment parameter. Moreover, the TUT effect have been reduced steadily during the period studied to be only 2.6 percent for men, and 8.0 percent for women.

Hence, this exercise shows the importance of self selection on formality choice, and how the treatment effect depends on the question we are

interested in answering. Several new questions arise from these estimations such as the importance of the persistence of formal or informal employment on the wage inequality and the role of uncertainty on wages over the decisions of switching across types of works. In principle, expanding the analysis would help us to identify better the sources of differences between formal and informal wages. This expansion indeed is opportunities of a future research agenda on this topic.

Conclusion

This research proposal analyzed the wage differences between formal and informal urban labor markets in Mexico, for the 2000/02-2003/01 period testing an alternative definition of formality and using the treatment effect approach to calculate the gap between wages. In general, wage differences exists between these two markets, nevertheless, these differences changes dramatically once that choice between sectors and the self selection component is considered as a key issue of these wages.

In this document, four different policy treatment effects were estimated to calculate the wage differences. For men, the largest effects are obtained from analyzing the unconditional mean differences and the OLS treatment effect which gave a wage gap of around 33.0 per cent and steadily reduces in time to reach 25.3 from 2000/02 to 2003/01. In contrast, for the same period the ATE effect of formality reduces this gap estimation to 14.3 percent and steadily decreases to 11.7 percent for the same period studied. Furthermore, the TUT effect of informal workers seems to be even smaller and also decreasing in time, from 7.5 to 2.6 percent. These results are qualitatively similar to women.

The evidence suggests that self selection effects accounts for explaining a very important fraction of the average wage differentials between formal and informal wages. This also suggests that, for the last quarter I have data, the wage from moving from informal to formal sector would increase at the most 2.6 percent for men (8.0 percent for women) for the average informal worker, and maybe this switching is not done because there are other factors not considered in this first approach model such as matching of abilities and intertemporal decisions which may also influence the convergence between these two wages.

As areas of opportunity for this research, further robustness test must be performed to consider: 1) alternative definitions of formality; 2) inclusion of alternative variables in the outcome and choice models; and 3) consider the model for rural (low population) communities. Also, future expansions of the structural model should consider: 1) a multichoice model to consider the

switching decision from unemployment to each of these to alternatives; and 2) multiperiod analysis for studying formality and informality employment persistence.

References

Blunch, Niels-Hugo, Sudharshan Canagarajah and Dhushyanth Raju. 2001. "The Informal Sector Revisted: A Synthesis Across Space and Time." Social Protection Discussion Paper No. 1991. The World Bank.

Cunha, Flavio; James Heckman, and Salvador Navarro. 2005. "Separating Uncertainty from Heterogeneity in Life Cycle Earnings." Oxford Economic Papers, 57: 191-261.

Galli, Rossana, and David Kucera. 2003. "Informal Employment in Latin America: Movements over Bussiness Cycles and the Effects of Worker Rights." Discussion Paper 145/2003. International Institute for Labour Studies.

Heckman, James J. and Edward Vyvlacil. 2006. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." Econometrica, 73(3):669-738.

Heckman, James, J. 1979. "Sample Selection Bias as a Specification Error." Econometrica 47 (1):153-161.

Huesca Reynoso, Luis. 2005. "La Distribucion Salarial del Mercado de Trabajo en México: Un Análisis de la Informalidad." Tesis. Universidad Autonoma de Barcelona.

Juhn Chinhui; Kevin Murphy and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill." The Journal of Political Economics, 101(3): 410-442.

Maddala, G.S. 1983. Limited Dependent and Qualitative Variables in Econometrics. Econometric Society Monographs No. 3. Cambridge University Press.

Maloney, William. 2003. "Informality Revisted." World Bank: Policy Research Working Paper No. 2965. Latin America and the Caribbean Region, Poverty Reduction and Economic Management Division.

--------- .1999. "Informality and the Creation of Firms: Evidence from a Rotative Panel" World Bank Policy Research Working Paper No. 1985.

Latin America and the Caribbean Region, Poverty Reduction and Economic Management Division.

--------- .1998. "Are Labor Market in Developing Countries Dualistic?" World Bank Policy Research Working Paper No. 1941. Latin America and the Caribbean Region, Poverty Reduction and Economic Management Division.

Marcouiller, Douglas; Veronica Ruiz de Castilla, and Christopher Woodruff. 1997. "Formal Measures of the Informal-Sector Gap in Mexico, El Salvador, and Peru." Economic Development and Cultural Change, 45 (2): 367-392.

McCall, John J. 1970. "Economics of Information and Job Search." Quarterly Journal of Economics, 84(1):113-126.

Rosen, Sherwin and Derek Neal. 1999. "Theories of the Distribution of Earnings." Chapter 7, in Handbook in Income Distribution. Springler-Verlag.

Stokey, Nancy L. and Robert E. Lucas, with Edward C. Prescott. 1989. Recursive Methods in Economic Dynamics. Harvard University Press.

Teilhet-Waldorf, Saral and William H. Waldorf. 1983. "Earnings of Self-Employed in an Informal Sector: A Case Study of Bangkok." Economic Development and Cultural Change, 31(3):587-607.

Tijerina Guajardo, Jose A. and Antonio Medellin Ruiz. 1999. "Tamaño del Sector Informal y Esfuerzo Fiscal Internacional." Documento de Investigacion No. 7. Centro de Analisis y Difusion Economica.

Willis, Robert J. and Sherwin Rosen. 1979. "Education and Self-Selection." The Journal of Political Economics, 87(5-2): S7-S36.

World Bank. 2004. "Poverty in Mexico: An Assessment of Conditions, Trends, and Government Strategies." World Bank: Report No. 28612-ME. Latin America and the Caribbean Region, Poverty Reduction and Economic Management Division.

Table 1. Formal and informal sector composition
(All, Men, and Women, 2000-2003)

Variable	2000/02	2000/03	2000/04	2001/01	2001/02	2001/03
All						
Total Employment	100	100	100	100	100	100
Formal Employment	81.18	81.37	81.65	81.25	81.09	80.88
Informal Employment	18.82	18.63	18.35	18.75	18.91	19.12
Self Employed & Employer	11.24	11.11	11.15	11.42	11.49	11.83
Informal Salaried	7.58	7.52	7.21	7.33	7.42	7.29
Men						
Total Employment	63.72	64.30	63.48	64.09	63.79	64.39
Formal Employment	52.69	53.15	52.59	52.89	52.56	52.87
Informal Employment	11.03	11.14	10.89	11.20	11.23	11.52
Self Employed & Informal Employer	7.18	7.28	7.23	7.51	7.40	7.76
Informal Salaried	3.85	3.86	3.66	3.69	3.83	3.76
Women						
Total Employment	36.28	35.70	36.52	35.91	36.21	35.61
Formal Employment	28.49	28.22	29.06	28.37	28.53	28.01
Informal Employment	7.79	7.48	7.46	7.55	7.68	7.60
Self Employed & Informal Employer	4.06	3.83	3.91	3.90	4.09	4.07
Informal Salaried	3.73	3.65	3.55	3.64	3.59	3.53

Source: Own estimations using ENET 2000:02 - 2003:01.

Table 1. Formal and informal sector composition
(All, Men, and Women, 2000-2003)

Variable	2001/04	2002/01	2002/02	2002/03	2002/04	2003/01
All						
Total Employment	100	100	100	100	100	100
Formal Employment	81.09	80.86	80.95	80.53	80.53	80.64
Informal Employment	18.91	19.14	19.05	19.47	19.47	19.36
Self Employed & Employer	11.77	11.71	11.69	11.88	11.88	11.88
Informal Salaried	7.14	7.43	7.36	7.59	7.58	7.48
Men						
Total Employment	63.37	64.03	63.53	68.98	63.80	63.57
Formal Employment	52.16	52.54	52.15	52.53	52.07	52.02
Informal Employment	11.21	11.49	11.38	11.73	11.74	11.55
Self Employed & Informal Employer	7.65	7.71	7.65	7.71	7.75	7.76
Informal Salaried	3.56	3.78	3.73	4.02	3.99	3.79
Women						
Total Employment	36.63	35.97	36.47	35.74	36.20	36.43
Formal Employment	28.93	28.32	28.80	27.99	28.46	28.61
Informal Employment	7.70	7.65	7.67	7.74	7.73	7.81
Self Employed & Informal Employer	4.12	4.00	4.04	4.17	4.14	4.12
Informal Salaried	3.58	3.65	3.63	3.58	3.59	3.69

Source: Own estimations using ENET 2000:02 - 2003:01.

Table 2.a. Probit models analysis for participation in formal sector (Men, 2000-2003)

Variable	2000/02	2000/03	2000/04	2001/01	2001/02	2001/03
a_expe	-0.00191	-0.00224	-0.00194	-0.00206	-0.00189	-0.00200
a_expe2	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
a_jefe(+)	0.02319	0.02295	0.01251	0.01770	0.01627	0.01576
a_hcp(+)	0.03413	0.03742	0.04335	0.03751	0.03797	0.04094
a_dprim(+)	0.08815	0.08917	0.07838	0.08458	0.08160	0.08144
a_dsecun(+)	0.15308	0.15625	0.14711	0.15392	0.15543	0.15380
a_dprepa(+)	0.16358	0.16425	0.15840	0.16453	0.16758	0.16747
a_dpro~s(+)	0.20845	0.20754	0.20571	0.20787	0.21201	0.21191
a_dtec~o(+)	0.14745	0.14972	0.14621	0.15159	0.15204	0.15238
a_dpost(+)	0.14728	0.14759	0.14580	0.14980	0.14972	0.15450
edo1(+)	0.05636	0.06424	0.06154	0.06827	0.07471	0.04771
edo2(+)	0.01382	0.02530	0.01356	0.00474	0.02272	-0.00893
edo3(+)	0.01799	0.00946	-0.01180	-0.02039	0.01119	-0.02330
edo4(+)	-0.01128	0.00390	0.00896	-0.01052	0.02648	-0.00466
edo5(+)	0.01296	0.02788	0.02746	0.00127	0.02749	0.00111
edo6(+)	0.02814	0.03400	0.02268	0.00834	0.03324	0.01373
edo7(+)	-0.03545	-0.02799	-0.01595	-0.02180	0.00762	-0.03025
edo8(+)	0.01732	0.02920	0.03294	0.02484	0.04593	0.02104
edo10(+)	0.01487	0.01639	0.01417	-0.00131	0.02418	-0.00215
edo11(+)	0.00651	0.02611	0.02157	0.00344	0.02555	0.01169
edo12(+)	-0.06397	-0.04493	-0.06011	-0.06566	-0.05091	-0.07563
edo13(+)	0.00672	0.02263	0.01747	0.00249	0.01983	0.00089
edo14(+)	-0.00946	0.01182	0.01258	0.00556	0.02839	-0.00291
edo15(+)	0.00952	0.00926	0.00852	0.00251	0.02910	-0.00243
edo16(+)	-0.02446	-0.03737	-0.03660	-0.05563	-0.03564	-0.08169
edo17(+)	-0.09627	-0.08235	-0.08004	-0.13086	-0.08513	-0.12608
edo18(+)	-0.03346	-0.02704	-0.04006	-0.03748	-0.01019	-0.04310
edo19(+)	-0.00854	0.00182	-0.00809	-0.00814	-0.00107	-0.01595
edo20(+)	-0.07317	-0.05863	-0.07470	-0.09037	-0.06805	-0.10692
edo21(+)	-0.02380	-0.00265	0.00067	-0.03353	-0.00558	-0.02447
edo22(+)	0.05637	0.04646	0.04368	0.02390	0.05387	0.02576
edo23(+)	0.03668	0.05851	0.06110	0.04645	0.06117	0.03265
edo24(+)	0.00650	0.01170	0.00800	-0.01625	0.00627	-0.00267
edo25(+)	-0.02216	-0.02276	-0.02711	-0.04928	-0.00320	-0.02563
edo26(+)	-0.01517	0.00638	0.00976	0.00354	0.02587	-0.02612
edo27(+)	-0.00067	0.01664	-0.01537	-0.01943	0.00833	-0.00048
edo28(+)	0.00220	-0.00312	0.00056	-0.02306	0.00293	-0.00607
edo30(+)	-0.01474	-0.00900	-0.02042	-0.02531	0.00038	-0.02561
edo31(+)	0.01337	0.00794	0.00748	-0.00812	0.01007	0.00380
edo32(+)	-0.02046	-0.00279	0.00430	-0.05948	-0.00039	-0.01990
Num. Obs	90767	93504	94529	91948	92074	90627
LR chi2(40)	9914.72	10271.62	10309.28	9907.39	10153.04	9763.93
Prob > chi2	0.00	0.00	0.00	0.00	0.00	0.00
Pseudo R2	0.1185	0.1191	0.119	0.1163	0.1185	0.1147

(1) Each coefficient shows the change in probability at the mean for a change in the variable.

(+) The coefficient is the change in probability for a discrete dummy variable.

* / Coefficient statistically significant at 95% of confidence.

** / Coefficient statistically significant at 90% of confidence.

Source: Own estimations using ENET 2000:02 - 2003:01.

Table 2.a. Probit models analysis for participation in formal sector (Men, 2000-2003)

Variable	2001/04	2002/01	2002/02	2002/03	2002/04	2003/01
a_expe	-0.00196	-0.00177	-0.00191	-0.00218	-0.00178	-0.00181
a_expe2	0.00000	0.00000	0.00000	0.00001	0.00000	0.00001
a_jefe(+)	0.01577	0.01211	0.01711	0.01514	0.02050	0.02042
a_hcp(+)	0.04085	0.04623	0.03913	0.04530	0.03918	0.03545
a_dprim(+)	0.08269	0.08444	0.08333	0.07823	0.09238	0.08357
a_dsecun(+)	0.15366	0.16069	0.15666	0.15276	0.16592	0.16031
a_dprepa(+)	0.16697	0.17030	0.16992	0.16858	0.17796	0.17411
a_dpro~s(+)	0.21462	0.21915	0.21770	0.21725	0.22855	0.22284
a_dtec~o(+)	0.15026	0.15539	0.15202	0.15037	0.15760	0.15441
a_dpost(+)	0.15204	0.15504	0.15455	0.15810	0.15807	0.15734
edo1(+)	0.06249	0.06368	0.07369	0.07657	0.06198	0.05639
edo2(+)	0.00468	0.00607	0.02738	0.02554	0.02900	0.00275
edo3(+)	0.01295	0.02412	0.03212	0.02449	0.00869	-0.00089
edo4(+)	-0.00165	-0.00165	0.01087	0.01041	0.00484	-0.01760
edo5(+)	0.01218	0.00128	0.02024	0.00621	0.00511	0.00054
edo6(+)	0.00578	0.02936	0.03556	0.03240	0.02656	-0.00169
edo7(+)	-0.00744	-0.00948	0.00367	-0.02489	-0.03102	-0.03305
edo8(+)	0.02194	0.03721	0.04650	0.03952	0.04562	0.01679
edo10(+)	-0.00182	0.00594	0.01637	0.00608	0.00667	-0.00342
edo11(+)	0.01649	0.02072	0.02690	0.02156	0.02082	0.00464
edo12(+)	-0.05649	-0.05049	-0.04541	-0.04663	-0.07475	-0.04844
edo13(+)	0.00008	0.01718	0.02417	0.00756	0.01244	0.01122
edo14(+)	0.00526	0.01923	0.04259	0.02223	0.01176	0.02015
edo15(+)	0.01168	0.01906	0.02676	0.01566	0.02579	0.00742
edo16(+)	-0.03826	-0.03858	-0.03567	-0.05241	-0.02621	-0.03371
edo17(+)	-0.12517	-0.12175	-0.10065	-0.17628	-0.14377	-0.15515
edo18(+)	-0.02913	-0.02794	-0.03114	-0.01945	-0.00911	-0.05047
edo19(+)	-0.00142	-0.01712	0.00289	-0.00545	-0.00411	-0.03044
edo20(+)	-0.06712	-0.05866	-0.01851	-0.03926	-0.06544	-0.07270
edo21(+)	-0.01198	-0.01355	0.00918	-0.03782	-0.02652	-0.04667
edo22(+)	0.03578	0.02939	0.04987	0.02295	0.05298	0.03022
edo23(+)	0.05090	0.05452	0.04521	0.03969	0.05245	0.02834
edo24(+)	0.01731	0.01019	0.02640	0.02172	0.02204	0.00566
edo25(+)	-0.03618	-0.04218	-0.00962	-0.03380	-0.02552	-0.03513
edo26(+)	-0.01408	-0.00800	0.01943	0.01128	0.02071	0.00873
edo27(+)	0.01031	0.00619	0.00461	-0.00285	-0.00492	-0.02797
edo28(+)	-0.00025	-0.00256	0.01306	0.00390	-0.00549	-0.02244
edo30(+)	-0.02356	-0.02311	-0.01393	-0.03772	-0.03016	-0.04375
edo31(+)	-0.00555	-0.00623	0.01396	0.00156	-0.00660	-0.00626
edo32(+)	0.00243	-0.01792	0.00985	-0.02793	-0.01129	-0.01863
Num. Obs	91617	89877	89114	87109	88142	87438
LR chi2(40)	10048.55	9704.62	9774.84	9531.52	10149.6	9549.11
Prob > chi2	0.00	0.00	0.00	0.00	0.00	0.00
Pseudo R2	0.1175	0.1147	0.1167	0.1151	0.1206	0.1152

(1) Each coefficient shows the change in probability at the mean for a change in the variable.

(+) The coefficient is the change in probability for a discrete dummy variable.

* / Coefficient statistically significant at 95% of confidence.

** / Coefficient statistically significant at 90% of confidence.

Source: Own estimations using ENET 2000:02 - 2003:01.

Table 2.b. Probit models analysis for participation in formal sector (Women, 2000-2003)

Variable	2000/02	2000/03	2000/04	2001/01	2001/02	2001/03
a_numhij	-0.00448	-0.00551	-0.00718	-0.00657	-0.00648	-0.00503
a_jefe(+)	0.00383	0.00534	-0.00115	0.01218	0.01172	0.00548
a_expe	-0.00097	-0.00227	-0.00106	-0.00138	-0.00169	-0.00199
a_expe2	-0.00002	0.00000	-0.00002	-0.00001	-0.00001	0.00000
a_dprim(+)	0.10184	0.10727	0.09670	0.09655	0.09450	0.09700
a_dsecun(+)	0.18168	0.17715	0.16372	0.16674	0.16635	0.17449
a_dprepa(+)	0.18872	0.18562	0.17788	0.18437	0.18414	0.19279
a_dpro~s(+)	0.26993	0.26095	0.26227	0.26559	0.26687	0.26288
a_dtec~o(+)	0.23283	0.22524	0.21544	0.21946	0.21923	0.22502
a_dpost(+)	0.16713	0.16225	0.15543	0.16200	0.16451	0.16793
edo1(+)	0.08482	0.08157	0.08155	0.07701	0.08496	0.08379
edo2(+)	0.10393	0.10008	0.08850	0.10534	0.09560	0.10184
edo3(+)	0.01870	0.03937	0.01529	0.02457	0.02849	0.01069
edo4(+)	-0.00411	-0.00253	-0.01843	-0.02379	-0.01327	-0.00851
edo5(+)	0.03293	0.04218	0.03339	0.03081	0.02514	0.03691
edo6(+)	-0.00868	0.01874	0.01124	-0.00487	-0.00378	-0.01990
edo7(+)	-0.06523	-0.05872	-0.04581	-0.05794	-0.04645	-0.05928
edo8(+)	0.11495	0.11688	0.10736	0.11063	0.12197	0.12195
edo10(+)	0.04183	0.02879	0.03401	0.03891	0.02771	0.02497
edo11(+)	0.04382	0.04170	0.04029	0.04272	0.03788	0.05082
edo12(+)	-0.04844	-0.04159	-0.01543	-0.01028	-0.01507	0.00801
edo13(+)	0.07059	0.06676	0.05010	0.05542	0.06133	0.06012
edo14(+)	0.01657	0.02637	0.01785	0.01470	0.01041	0.01145
edo15(+)	0.04023	0.03132	0.02470	0.03616	0.02741	0.02550
edo16(+)	0.00361	-0.00638	-0.00870	0.00883	0.00045	0.01448
edo17(+)	-0.02111	-0.03169	-0.04511	-0.04436	-0.02918	-0.02540
edo18(+)	0.00118	-0.01274	-0.02902	-0.02662	-0.03427	-0.03216
edo19(+)	0.02330	0.02510	0.01657	0.01302	0.00578	-0.00128
edo20(+)	0.01939	0.01170	-0.02861	-0.00859	-0.02801	0.00196
edo21(+)	0.04129	0.03065	0.01825	0.00946	0.01366	0.01237
edo22(+)	0.06497	0.05458	0.03562	0.06078	0.05513	0.04001
edo23(+)	0.04487	0.04726	0.04794	0.04455	0.04942	0.03985
edo24(+)	0.03306	0.02962	0.02095	0.03176	0.01672	0.04209
edo25(+)	-0.03002	-0.02677	-0.03590	-0.03431	-0.05358	-0.03698
edo26(+)	0.03215	0.02943	0.02244	0.02080	0.01016	-0.00246
edo27(+)	-0.02675	-0.01725	-0.03371	-0.04720	-0.04745	-0.03193
edo28(+)	0.07067	0.06014	0.04718	0.05113	0.04648	0.05668
edo30(+)	-0.04751	-0.05205	-0.05506	-0.05736	-0.05240	-0.05381
edo31(+)	0.01540	0.02006	0.02613	0.01559	-0.00445	0.01712
edo32(+)	0.05291	0.05320	0.04828	0.01728	0.03788	0.04880
Num. Obs	51679	51919	54391	51527	52268	50112
LR chi2(40)	11440.72	11006.67	11826.23	11177.14	11387.3	10671.48
Prob > chi2	0.00	0.00	0.00	0.00	0.00	0.00
Pseudo R2	0.2128	0.2064	0.2147	0.211	0.2108	0.2054

(1) Each coefficient shows the change in probability at the mean for a change in the variable.

(+) The coefficient is the change in probability for a discrete dummy variable.

* / Coefficient statistically significant at 95% of confidence.

** / Coefficient statistically significant at 90% of confidence.

Source: Own estimations using ENET 2000:02 - 2003:01.

Table 2.b. Probit models analysis for participation in formal sector (Women, 2000-2003)

Variable	2001/04	2002/01	2002/02	2002/03	2002/04	2003/01	2001/04
a_numhij	-0.00726	-0.00695	-0.01008	-0.00929	-0.00595	-0.00977	-0.00726
a_jefe(+)	0.00814	0.00782	0.01099	0.00430	0.00837	0.01988	0.00814
a_expe	-0.00107	-0.00103	-0.00116	-0.00121	-0.00140	-0.00083	-0.00107
a_expe2	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001
a_dprim(+)	0.08427	0.08717	0.08729	0.09484	0.08310	0.09248	0.08427
a_dsecun(+)	0.16238	0.16607	0.16866	0.17467	0.16511	0.17266	0.16238
a_dprepa(+)	0.18291	0.18568	0.18596	0.19288	0.18628	0.19274	0.18291
a_dpro~s(+)	0.27033	0.27316	0.27122	0.27248	0.27539	0.28145	0.27033
a_dtec~o(+)	0.21286	0.21910	0.21877	0.22228	0.21579	0.22028	0.21286
a_dpost(+)	0.16453	0.16584	0.16506	0.17126	0.16732	0.17016	0.16453
edo1(+)	0.07811	0.05703	0.07383	0.08572	0.06749	0.05817	0.07811
edo2(+)	0.08886	0.08923	0.08593	0.08066	0.09217	0.07580	0.08886
edo3(+)	0.02471	0.02730	0.00531	-0.00182	0.01725	0.00215	0.02471
edo4(+)	-0.04131	-0.00999	-0.00949	-0.02106	-0.01442	-0.01002	-0.04131
edo5(+)	0.01362	0.02609	0.01816	0.02687	0.02738	0.02012	0.01362
edo6(+)	-0.00557	-0.01684	-0.00695	-0.02145	-0.01524	-0.03478	-0.00557
edo7(+)	-0.07646	-0.06912	-0.03736	-0.06038	-0.04516	-0.05000	-0.07646
edo8(+)	0.10392	0.10539	0.10494	0.10681	0.11044	0.09808	0.10392
edo10(+)	0.01189	0.02209	0.02743	0.02812	0.03329	0.01771	0.01189
edo11(+)	0.03745	0.03533	0.04697	0.03543	0.03799	0.02333	0.03745
edo12(+)	-0.01268	-0.03306	-0.03281	-0.05131	-0.03759	-0.04358	-0.01268
edo13(+)	0.02320	0.03078	0.04661	0.04295	0.05275	0.04607	0.02320
edo14(+)	0.00520	0.00951	0.02709	0.02632	0.01422	0.02619	0.00520
edo15(+)	0.01718	0.01669	0.03508	0.03825	0.03057	0.02868	0.01718
edo16(+)	0.02408	0.02015	0.01568	-0.01175	0.00234	-0.01408	0.02408
edo17(+)	-0.05887	-0.02396	-0.04432	-0.08040	-0.06432	-0.10548	-0.05887
edo18(+)	-0.02140	-0.01585	-0.02707	-0.02783	-0.01750	-0.01807	-0.02140
edo19(+)	0.00366	-0.00583	0.00771	-0.00422	0.00356	-0.00293	0.00366
edo20(+)	-0.02883	-0.02450	0.01446	-0.01073	0.01877	-0.02031	-0.02883
edo21(+)	0.01075	0.02038	0.02137	0.00590	0.01018	-0.01501	0.01075
edo22(+)	0.03646	0.05402	0.05430	0.04084	0.06782	0.05147	0.03646
edo23(+)	0.02583	0.04261	0.03390	0.02855	0.03284	0.03220	0.02583
edo24(+)	0.02282	0.02399	0.03446	0.02316	0.03947	0.01389	0.02282
edo25(+)	-0.04844	-0.06015	-0.04319	-0.06989	-0.04890	-0.06896	-0.04844
edo26(+)	-0.01082	-0.01439	-0.02240	-0.00590	0.00850	-0.00091	-0.01082
edo27(+)	-0.05064	-0.06482	-0.05847	-0.05629	-0.06138	-0.04010	-0.05064
edo28(+)	0.02718	0.04176	0.05465	0.04503	0.04633	0.03428	0.02718
edo30(+)	-0.06244	-0.05338	-0.04529	-0.06759	-0.05606	-0.07480	-0.06244
edo31(+)	-0.01427	0.00944	0.00589	-0.00380	0.01362	0.00679	-0.01427
edo32(+)	0.03679	0.05834	0.06000	0.04903	0.05192	0.05820	0.03679
Num. Obs	52955	50489	51146	48441	50006	50101	52955
LR chi2(40)	11300.82	10925.57	10948.9	10423.9	10714.11	10479.35	11300.82
Prob > chi2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pseudo R2	0.2075	0.2091	0.2081	0.2058	0.2065	0.2012	0.2075

(1) Each coefficient shows the change in probability at the mean for a change in the variable.

(+) The coefficient is the change in probability for a discrete dummy variable.

* / Coefficient statistically significant at 95% of confidence.

** / Coefficient statistically significant at 90% of confidence.

Source: Own estimations using ENET 2000:02 - 2003:01.

Table 3.a. TSLS wage equation in formal sector
(Men, 2000-2003)

Variable	2000/02	2000/03	2000/04	2001/01	2001/02	2001/03
a_expe	0.0303	0.0306	0.0307	0.0305	0.0307	0.0309
a_expe2	-0.0003	-0.0003	-0.0003	-0.0003	-0.0004	-0.0004
a_dprim	-0.3020	-0.2869	-0.1903	-0.2026	-0.1728	-0.1496
a_dsecun	-0.4578	-0.4342	-0.2950	-0.3285	-0.2823	-0.2712
a_dprepa	-0.4232	-0.3831	-0.2268	-0.2791	-0.2251	-0.2083
a_dtecnico	-0.4118	-0.3911	-0.2189	-0.2688	-0.2144	-0.2009
a_dprofes	0.0111	0.0534	0.2358	0.1734	0.2316	0.2155
a_dpost	0.3957	0.4103	0.5870	0.4990	0.6032	0.5183
mills_f_h	1.8407	1.7621	1.5268	1.6192	1.4461	1.4914
_cons	3.1929	3.1467	3.0025	3.0993	3.0619	3.0584
Num Obs	75055	77300	78310	75880	75870	74413
F	3766.91	3865.58	3945.48	3716.49	3686.5	3548.83
Prob (F)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R 2	0.3112	0.3104	0.312	0.306	0.3043	0.3003
Adj. R2	0.3111	0.3103	0.3119	0.3059	0.3042	0.3003

* / Coefficient statistically significant at 95% of confidence.
** / Coefficient statistically significant at 90% of confidence.
Source: Own estimations using ENET 2000:02 - 2003:01.

Table 3.b. TSLS wage equation in informal sector
(Men, 2000-2003)

Variable	2000/02	2000/03	2000/04	2001/01	2001/02	2001/03
a_expe	0.0269	0.0279	0.0262	0.0283	0.0295	0.0280
a_expe2	-0.0003	-0.0004	-0.0003	-0.0004	-0.0004	-0.0004
a_dprim	-0.1538	-0.1000	-0.0595	-0.0693	-0.0711	-0.0567
a_dsecun	-0.3265	-0.2196	-0.2013	-0.1521	-0.1837	-0.1484
a_dprepa	-0.3780	-0.2368	-0.2162	-0.1566	-0.2194	-0.1723
a_dtecnico	-0.4743	-0.3073	-0.2473	-0.2993	-0.3367	-0.2671
a_dprofes	-0.4394	-0.2033	-0.1936	-0.0679	-0.1801	-0.2027
a_dpost	-0.4460	-0.3920	-0.0361	-0.2771	-0.1579	-1.0472
mills_i_h	0.9385	0.8051	0.7883	0.6806	0.6829	0.6812
_cons	1.0855	1.2108	1.2856	1.4438	1.5115	1.5095
Num Obs	15712	16204	16219	16068	16204	16214
F	222.88	237.96	233.85	218.74	203.47	204.57
Prob (F)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R 2	0.1133	0.1168	0.1149	0.1092	0.1016	0.102
Adj. R2	0.1128	0.1163	0.1144	0.1087	0.1011	0.1015

* / Coefficient statistically significant at 95% of confidence.
** / Coefficient statistically significant at 90% of confidence.
Source: Own estimations using ENET 2000:02 - 2003:01.

Table 3.a. TSLS wage equation in formal sector
(Men, 2000-2003)

Variable	2001/04	2002/01	2002/02	2002/03	2002/04	2003/01
a_expe	0.0295	0.0299	0.0290	0.0296	0.0289	0.0300
a_expe2	-0.0003	-0.0004	-0.0003	-0.0003	-0.0003	-0.0004
a_dprim	-0.2020	-0.1745	-0.1891	-0.1648	-0.2519	-0.2110
a_dsecun	-0.3326	-0.2849	-0.3634	-0.3114	-0.4261	-0.3838
a_dprepa	-0.2822	-0.2139	-0.3397	-0.2795	-0.4083	-0.3491
a_dtecnico	-0.2536	-0.2051	-0.3077	-0.2366	-0.3768	-0.3076
a_dprofes	0.1592	0.2184	0.0756	0.1375	-0.0031	0.0709
a_dpost	0.4726	0.5614	0.4237	0.4923	0.3601	0.4238
mills_f_h	1.5452	1.3842	1.6531	1.5620	1.5936	1.5095
_cons	3.1622	3.1202	3.2716	3.2141	3.3790	3.3091
Num Obs	75411	73746	73152	71209	71930	71557
F	3846.79	3493.69	3634.44	3536.76	3554.66	3432.38
Prob (F)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R 2	0.3147	0.2989	0.309	0.3089	0.3079	0.3016
Adj. R2	0.3146	0.2989	0.3089	0.3089	0.3078	0.3015

* / Coefficient statistically significant at 95% of confidence.
** / Coefficient statistically significant at 90% of confidence.
Source: Own estimations using ENET 2000:02 - 2003:01.

Table 3.b. TSLS wage equation in informal sector
(Men, 2000-2003)

Variable	2001/04	2002/01	2002/02	2002/03	2002/04	2003/01
a_expe	0.0282	0.0258	0.0255	0.0283	0.0246	0.0262
a_expe2	-0.0004	-0.0004	-0.0003	-0.0004	-0.0003	-0.0004
a_dprim	-0.0703	-0.0136	-0.1091	-0.0627	-0.1013	-0.1351
a_dsecun	-0.1887	-0.1032	-0.2858	-0.1938	-0.2438	-0.3200
a_dprepa	-0.2542	-0.1439	-0.3725	-0.2533	-0.3120	-0.4175
a_dtecnico	-0.2752	-0.2464	-0.4432	-0.2932	-0.3777	-0.4727
a_dprofes	-0.1807	-0.1221	-0.4549	-0.2538	-0.4058	-0.5075
a_dpost	-0.4526	-0.3886	-1.0905	-1.0875	-0.5351	-0.6401
mills_i_h	0.7087	0.6515	0.8638	0.7193	0.7833	0.8215
_cons	1.5164	1.6205	1.4582	1.5667	1.6044	1.6296
Num Obs	16206	16131	15962	15900	16212	15881
F	218.02	195.12	205.08	214.31	214.52	184.08
Prob (F)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R 2	0.1081	0.0982	0.1037	0.1082	0.1065	0.0945
Adj. R2	0.1076	0.0977	0.1032	0.1077	0.106	0.094

* / Coefficient statistically significant at 95% of confidence.
** / Coefficient statistically significant at 90% of confidence.
Source: Own estimations using ENET 2000:02 - 2003:01.

Table 4.a. TSLS wage equation in formal sector (Women, 2000-2003)

Variable	2000/02	2000/03	2000/04	2001/01	2001/02	2001/03
a_expe	0.0358	0.0358	0.0363	0.0340	0.0335	0.0337
a_expe2	-0.0004	-0.0004	-0.0004	-0.0004	-0.0004	-0.0004
a_dprim	-0.2889	-0.3740	-0.3263	-0.3538	-0.2294	-0.2448
a_dsecun	-0.3903	-0.4609	-0.3890	-0.4572	-0.3053	-0.3438
a_dprepa	-0.2651	-0.3453	-0.3004	-0.3904	-0.1929	-0.2751
a_dtecnico	-0.2895	-0.3778	-0.2952	-0.3870	-0.2065	-0.2817
a_dprofes	0.2313	0.1349	0.2178	0.0977	0.2934	0.2257
a_dpost	0.5882	0.4987	0.5700	0.4372	0.6214	0.5646
mills_f_h	1.2666	1.3166	1.2788	1.3720	1.1860	1.2754
_cons	2.8428	2.9190	2.8870	3.0490	2.8819	2.9361
Num Obs	40578	41035	43277	40699	41177	39415
F	2476.49	2387.08	2804.17	2525.26	2447.01	2293.34
Prob (F)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R 2	0.3546	0.3437	0.3684	0.3584	0.3485	0.3437
Adj. R2	0.3544	0.3435	0.3683	0.3582	0.3484	0.3436

* / Coefficient statistically significant at 95% of confidence.

** / Coefficient statistically significant at 90% of confidence.

Source: Own estimations using ENET 2000:02 - 2003:01.

Table 4.b. TSLS wage equation in informal sector (Women, 2000-2003)

Variable	2000/02	2000/03	2000/04	2001/01	2001/02	2001/03
a_expe	0.0302	0.0347	0.0279	0.0307	0.0299	0.0283
a_expe2	-0.0003	-0.0004	-0.0003	-0.0004	-0.0004	-0.0003
a_dprim	-0.1696	-0.1600	-0.1169	-0.0590	-0.1384	-0.0645
a_dsecun	-0.3741	-0.3522	-0.2207	-0.1865	-0.2528	-0.1865
a_dprepa	-0.4557	-0.4730	-0.3603	-0.2764	-0.4071	-0.2974
a_dtecnico	-0.6294	-0.6389	-0.4634	-0.4341	-0.5499	-0.4132
a_dprofes	-0.5470	-0.7031	-0.3700	-0.3103	-0.5604	-0.2970
a_dpost	0.3981	0.4972	-0.7683	0.0238	0.0712	-0.3574
mills_f_h	1.0052	1.0607	0.9053	0.8874	0.9087	0.8234
_cons	0.8900	0.7739	0.9986	1.0038	1.1003	1.1410
Num Obs	11101	10884	11114	10828	11091	10697
F	171.19	169.01	161.64	166.89	138.47	143.1
Prob (F)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R 2	0.122	0.1227	0.1158	0.1219	0.1011	0.1075
Adj. R2	0.1213	0.122	0.1151	0.1212	0.1004	0.1068

* / Coefficient statistically significant at 95% of confidence.

** / Coefficient statistically significant at 90% of confidence.

Source: Own estimations using ENET 2000:02 - 2003:01.

Table 4.a. TSLS wage equation in formal sector
(Women, 2000-2003)

Variable	2001/04	2002/01	2002/02	2002/03	2002/04	2003/01
a_expe	0.0328	0.0315	0.0319	0.0325	0.0307	0.0318
a_expe2	-0.0003	-0.0003	-0.0004	-0.0004	-0.0003	-0.0004
a_dprim	-0.2100	-0.2197	-0.2057	-0.2333	-0.1930	-0.2111
a_dsecun	-0.2972	-0.3153	-0.3180	-0.3266	-0.2950	-0.3090
a_dprepa	-0.2105	-0.2287	-0.2199	-0.2584	-0.2299	-0.2283
a_dtecnico	-0.2001	-0.2314	-0.1937	-0.2287	-0.2003	-0.2087
a_dprofes	0.2829	0.2484	0.2790	0.2226	0.2650	0.2577
a_dpost	0.6100	0.5937	0.6240	0.5315	0.5616	0.6076
mills_f_h	1.2635	1.2474	1.1871	1.1846	1.1802	1.2031
_cons	2.9258	2.9960	2.9541	2.9907	2.9992	3.0294
Num Obs	41819	39748	40392	37944	39325	39356
F	2734.15	2374.82	2485.85	2165.94	2348.37	2428.94
Prob (F)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R 2	0.3705	0.3497	0.3565	0.3394	0.3496	0.3572
Adj. R2	0.3704	0.3496	0.3564	0.3393	0.3495	0.357

* / Coefficient statistically significant at 95% of confidence.
** / Coefficient statistically significant at 90% of confidence.
Source: Own estimations using ENET 2000:02 - 2003:01.

Table 4.b. TSLS wage equation in informal sector
(Women, 2000-2003)

Variable	2001/04	2002/01	2002/02	2002/03	2002/04	2003/01
a_expe	0.0256	0.0254	0.0270	0.0271	0.0244	0.0255
a_expe2	-0.0003	-0.0003	-0.0003	-0.0003	-0.0003	-0.0003
a_dprim	-0.0827	-0.0790	-0.0279	-0.1042	-0.1283	-0.1116
a_dsecun	-0.2857	-0.2729	-0.1439	-0.3047	-0.2883	-0.2750
a_dprepa	-0.4040	-0.4454	-0.2940	-0.4411	-0.4262	-0.3924
a_dtecnico	-0.5116	-0.5569	-0.3842	-0.5931	-0.5988	-0.5052
a_dprofes	-0.5367	-0.6754	-0.1694	-0.5914	-0.5646	-0.5216
a_dpost	-1.3135	-0.0689	-0.1215	-1.8087	-1.1705	-0.9856
mills_f_h	0.9577	0.9981	0.7831	0.9202	0.8901	0.8676
_cons	1.1175	1.1281	1.2617	1.2481	1.3530	1.3929
Num Obs	11136	10741	10754	10497	10681	10745
F	150.49	145.57	136.52	123.82	113.9	123.99
Prob (F)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
R 2	0.1085	0.1088	0.1026	0.0961	0.0876	0.0942
Adj. R2	0.1078	0.1081	0.1019	0.0953	0.0869	0.0934

* / Coefficient statistically significant at 95% of confidence.
** / Coefficient statistically significant at 90% of confidence.
Source: Own estimations using ENET 2000:02 - 2003:01.

Table 5.a. Treatment effects of switching from informal to formal (Men, 2000-2003)

Variable	2000/02	2000/03	2000/04	2001/01	2001/02	2001/03
Mean Logw – Formal Observed	2.7947	2.8038	2.8652	2.8865	2.9311	2.9211
Mean Logw – Informal Observed	2.4650	2.5039	2.5552	2.6024	2.6531	2.6423
Mean Logw – Formal OLS estimate	2.7947	2.8037	2.8652	2.8865	2.9311	2.9210
Mean Logw – Informal OLS estimate	2.4651	2.5040	2.5552	2.6024	2.6531	2.6423
Mean Logw – Formal Corrected	2.7506	2.7613	2.8223	2.8443	2.8884	2.8795
Mean Logw – Informal Corrected	2.6197	2.6569	2.7124	2.7430	2.7802	2.7580
Δ Ordinary Least Squares	0.3295	0.2998	0.3099	0.2841	0.2780	0.2788
Δ Average Treatment Effect	0.1310	0.1044	0.1099	0.1013	0.1082	0.1215
Δ Treatment on the Treated	0.1426	0.1148	0.1202	0.1137	0.1238	0.1378
Δ Treatment on the Untreated	0.0753	0.0547	0.0601	0.0429	0.0354	0.0468

Source: Own estimations using ENET 2000:02 - 2003:01.

Table 5.b. Treatment effects of switching from informal to formal (Women, 2000-2003)

Variable	2000/02	2000/03	2000/04	2001/01	2001/02	2001/03
Mean Logw – Formal Observed	2.7174	2.7050	2.7833	2.8027	2.8439	2.8114
Mean Logw – Informal Observed	2.1749	2.2187	2.2548	2.2991	2.3299	2.3244
Mean Logw – Formal OLS estimate	2.7174	2.7050	2.7833	2.8028	2.8439	2.8114
Mean Logw – Informal OLS estimate	2.1749	2.2186	2.2548	2.2990	2.3299	2.3244
Mean Logw – Formal Corrected	2.6325	2.6303	2.7039	2.7218	2.7638	2.7342
Mean Logw – Informal Corrected	2.4399	2.4540	2.5105	2.5516	2.5537	2.5399
Δ Ordinary Least Squares	0.5425	0.4864	0.5285	0.5038	0.5140	0.4870
Δ Average Treatment Effect	0.1926	0.1763	0.1934	0.1702	0.2101	0.1943
Δ Treatment on the Treated	0.2050	0.1887	0.2071	0.1840	0.2299	0.2130
Δ Treatment on the Untreated	0.1474	0.1299	0.1402	0.1184	0.1364	0.1254

Source: Own estimations using ENET 2000:02 - 2003:01.

Table 5.a. Treatment effects of switching from informal to formal (Men, 2000-2003)

Variable	2001/04	2002/01	2002/02	2002/03	2002/04	2003/01
Mean Logw – Formal Observed	2.9519	2.9903	2.9839	2.9976	3.0271	3.0433
Mean Logw – Informal Observed	2.6681	2.7230	2.7146	2.7237	2.7523	2.7908
Mean Logw – Formal OLS estimate	2.9519	2.9904	2.9839	2.9976	3.0272	3.0434
Mean Logw – Informal OLS estimate	2.6681	2.7231	2.7147	2.7237	2.7522	2.7908
Mean Logw – Formal Corrected	2.9088	2.9479	2.9411	2.9544	2.9827	3.0022
Mean Logw – Informal Corrected	2.8023	2.8465	2.8348	2.8377	2.8786	2.9013
Δ Ordinary Least Squares	0.2838	0.2673	0.2693	0.2740	0.2750	0.2525
Δ Average Treatment Effect	0.1065	0.1014	0.1063	0.1167	0.1041	0.1008
Δ Treatment on the Treated	0.1207	0.1169	0.1229	0.1345	0.1201	0.1175
Δ Treatment on the Untreated	0.0401	0.0308	0.0303	0.0372	0.0332	0.0257

Source: Own estimations using ENET 2000:02 - 2003:01.

Table 5.b. Treatment effects of switching from informal to formal (Women, 2000-2003)

Variable	2001/04	2002/01	2002/02	2002/03	2002/04	2003/01
Mean Logw – Formal Observed	2.8838	2.9109	2.9134	2.9010	2.9477	2.9692
Mean Logw – Informal Observed	2.3736	2.4204	2.4248	2.4454	2.4814	2.5163
Mean Logw – Formal OLS estimate	2.8838	2.9110	2.9135	2.9011	2.9478	2.9692
Mean Logw – Informal OLS estimate	2.3736	2.4203	2.4247	2.4455	2.4817	2.5161
Mean Logw – Formal Corrected	2.8026	2.8294	2.8345	2.8231	2.8700	2.8892
Mean Logw – Informal Corrected	2.5990	2.6476	2.6499	2.6323	2.6772	2.7114
Δ Ordinary Least Squares	0.5102	0.4906	0.4888	0.4555	0.4661	0.4531
Δ Average Treatment Effect	0.2036	0.1818	0.1846	0.1908	0.1928	0.1778
Δ Treatment on the Treated	0.2248	0.2019	0.2036	0.2171	0.2175	0.2045
Δ Treatment on the Untreated	0.1240	0.1072	0.1132	0.0958	0.1020	0.0800

Source: Own estimations using ENET 2000:02 - 2003:01.