Structural Modeling of Female Labour Participation and Occupation Decisions

(First Draft)

R. E. De Hoyos*

University of Cambridge

July 12, 2005

1 Introduction

Increasing female labour participation is an important aspect contributing to the development process in many emerging economies (see: Bourguignon, Fournier and Gurgand, 1998). Identifying the determinants of female labour participation decision improves our understanding of the dynamics of labour supply and its interaction with economic development. The present study aims to obtain the determinants of female labour participation and occupation decisions within a structural, utility-maximizing framework.

Based on the relationships described by the structural model, the paper develops a microeconometric model to obtain the determinants of labour participation and occupation decisions. We depart from a utility maximizing framework where individuals' choice depends on a set of comparisons between *expected* market wages and a subjective *reservation* wage. The agent's choice depends—ultimately—on personal and household characteristics and

*red29@cam.ac.uk

a subjective valuation of leisure. Since participation/occupation decisions are not the outcome of a random process, we model *expected* wages taking selectivity into account; therefore selectivity-adjusted wages and labour particiaption/occupation functions are estimated using a two-step procedure.

The structural models of labour occupation developed in Heckman and Sedlacek (1985) and Heckman and Honore (1990) are based on the Roy (1951) concept of comparative advantage. The papers show that agents' occupation decision are not entirely determined by differences in market wages with personal preferences playing a relevant role in the *selection* process. This findings, together with the fact that the econometrician only observe market wages and not personal preferences, have lead almost all studies to reduce-form estimations where wages are substituted by its determinants. Gong and van Soest (2002) develop a model that explicitly links expected wages with labour participation and occupation decisions. Using data for Mexico the authors find that wages have a positive effect upon women's participation and the occupation choice. The study adds to the literature on methodological and empirical grounds. Our approach is novel in three aspects: (1) the microeconometric framework derives from a structural model with expected wages explicitly determining labour participation/occupation decisions (2) using several cross-sectional data we uncover the dynamics of female-male labour participation dependence and (3) participation/occupation effects of changes in parameters are evaluated using micro-simulation techniques.

The model is estimated using Mexican household data for the years 1994 to 2000. The period under analysis allows us to test changes in female labour participation/occupation wage elasticities during a time of liberalizing reforms. Female labour participation in Mexico increased 13% during the first six years of NAFTA. Using micro-simulation techniques we can estimate how much of that change was brought about by changes in wages or any other determinant (e.g. the dependence of female's decisions on that of their spouse).

The paper is organized in the following way. In the next section we develope the model, stressing the necessary assumptions to identify the parameters in our empirical strategy. Section 3 shows the results of the model using household data from Mexico for years 1994 to 2000. In section 4 we carry out microsimulation exercises to test the impact of an exogenous change in expected wages upon the employment and the occupation structure. Finally conclusions are shown in the last section.

2 The Model

2.1 Theoretical Framework

Following Heckman (1974), we assume that agents' participation decision is determined by the difference between market wages and a reservation wage function (what Heckman calls shadow prices of female time). Heckman develops a model for the binomial choice problem, however it can be easily extended to a multinomial one. Define a reservation wage w^*_{ij}, as the minimum wage required to observe individual i working in occupation j:

$$w^*_{ij} = w^*_j(\boldsymbol{X}_i, \boldsymbol{Z}_i) \qquad \forall\, j \tag{1}$$

Where $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are vectors of personal and household characteristics of individual i, respectively. Instead of having a single reservation wage for each individual, as in Heckman's model, we have $(J-1)$ of them plus an implicit one i.e. as many as the number of choices.[1] The $(J-1)$ reservation wages derived from the assumption that occupations have different characteristics—beyond monetary ones—that have a value for the individual. To allow for different reservation wages for each choice is justified on the basis of differences in observable characteristics (e.g. working conditions)

[1]The implicit one is the shadow price of leisure.

and unobservable ones (e.g. individual preference for a particular occupation) across occupations. The individual, will attache different valuations to each occupation. We can think about institutional rigidities in the labour market where the lack of working hours flexibility can be substituted by occupation choices.[2] The utility valuation given to the different occupation characteristics are captured by vectors $\boldsymbol{X}$ and $\boldsymbol{Z}$. On the other hand, expected market wages, following conventional human capital theory, are defined by the well-known function $\hat{w}_{ij} = w_j(\boldsymbol{X}_i)$.

Since we defined a reservation and an expected market wage for each occupation, agents' choice will be based on a series of pair comparisons between $\hat{w}_j$ and w_j^*. The utility maximizing choice will depend not merely on the level of these two components but on the distance between them, i.e. $(\hat{w}_j - w_j^*)$. In this framework, the conventional reservation wage (i.e. whether an individual decides to work or not) is defined implicitly by the same set of pair comparisons. An individual would participate in the labour market as long as one of the differences $\hat{w}_j - w_j^*$ is positive but her final choice will be the one which maximizes the gap between them. There is an implicit utility function embedded in this maximizing process that can be define as:[3]

$$V_{ij} = V(\hat{w}_{ij}, w_{ij}^*), \qquad \forall\, j \tag{2}$$

A reduced form of (2) takes into account the fact the we do not observe w_j^* and hence we can only include the observable components that determine it $(\boldsymbol{X}, \boldsymbol{Z})$. Assuming that utility, $V(\cdot)$, is a linear function of its arguments, we can define it as:

$$V_{ij} = \lambda \hat{w}_{ij} + \boldsymbol{X}_i \boldsymbol{\gamma}_j' + \boldsymbol{Z}_i \boldsymbol{\gamma}_j + \eta_{ij} \tag{3}$$

[2]See: Deaton and Muellbauer, 1980, pg. 86.

[3]Notice that while V is define by elements w and w^*, reservation wages will depend, in turn, on individual preferences. Therefore utility is ultimately define by monetary income and individual preferences.

Where η_{ij} is a stochastic component.[4] We are implicitly assuming that, controlling for differences in $\boldsymbol{X}$ and $\boldsymbol{Z}$, the marginal utility of monetary income, λ, is constant across individuals and occupations, therefore λ is a scalar parameter. A major problem with equation (3) is that, since $\hat{w}$ is fully determined by $\boldsymbol{X}$ we cannot identify both sets of parameters λ and γ'_j at the same time. Changes in $\boldsymbol{X}$ will have a double and simultaneous effect, on the one hand upon market wages and on the other upon reservation wages. To tackle this problem, as an alternative to (3), we define a less flexible but more parsimonious version of the utility function. Substitute the reduced form version of the expected wage function (i.e. $\boldsymbol{X}_i\hat{\beta}_j$) into (3):

$$V_{ij} = \lambda(\boldsymbol{X}_i\hat{\beta}_j) + \boldsymbol{X}_i\gamma'_j + \boldsymbol{Z}_i\boldsymbol{\gamma}_j + \eta_{ij} \tag{4}$$

Simplifying:

$$V_{ij} = (\lambda + \gamma'_j/\hat{\beta}_j)\boldsymbol{X}_i\hat{\beta}_j + \boldsymbol{Z}_i\boldsymbol{\gamma}_j + \eta_{ij} \tag{5}$$

Define $\delta_j = (\lambda + \gamma'_j/\hat{\beta}_j)$:

$$V_{ij} = \delta_j\hat{w}_{ij} + \boldsymbol{Z}_i\boldsymbol{\gamma}_j + \eta_{ij} \tag{6}$$

Equation (6) allows the marginal utility of fitted wages, $\hat{w}$, to differ across choices capturing the unobservable effects deriving from the reservation wage function (γ'_j) and remunerations to personal characteristics across occupations ($\hat{\beta}_j$). Based in this last specification, individual i chooses alternative j if and only if:

$$V_{ij} > \max_{m \neq j}\{V_{im}\} \;\forall\; j \tag{7}$$

[4] To clarify notation, λ is a scalar, $\boldsymbol{\gamma}'_j$ and $\boldsymbol{\gamma}_j$ are vectors of size K_1 (personal characteristics) and K_2 (household characteristics), respectively.

The framework presented, departs from Heckman (1974) reservation wage concept and arrives to McFadden (1974) utility maximization criteria (equation 7). Unifying both approaches help us understand the individuals' processes that might lie behind the data we observe. This can shed light upon the determinants of agents' participation and occupation decisions. One of the advantages of estimating a structural model is that its results can be used to evaluate out-of-sample response to e.g. income tax changes. Furthermore, our focus solely in participation and occupation decisions rather than endogenising hours of work is based on the institutional rigidities present in Mexico where working hours are not freely chosen.[5]

2.2 Empirical Strategy

This section elaborates on the aspects that we have to take into account to obtain a set of equations suitable for estimation. The advantage of having a structural model behind the estimations is that we can interpret the parameters in a way that is consistent with the theoretical framework. For example, from (3) we know that a change in one of the elements of $\boldsymbol{X}$ will have a double—and possibly opposing—effect upon the probability of participating in the labour market. On the one hand an increase in $\boldsymbol{X}$ will tend to increase the agent's expected wage and this might have a positive effect upon participation. On the other hand, the same increase in $\boldsymbol{X}$ can increase the agent's reservation wage and hence reduce her participation probability. Although we do not observe this second effect, we could estimate a specification like (3) and try to get both effects. However, as we already pointed out, we cannot simultaneously identify the parameters on wages and $\boldsymbol{X}$. Furthermore, even using a parsimonious specification like (6), the interpretation of δ_j is not straightforward. Given the normalization assumption necessary to esti-

[5] As stated by Heckman "Participation (or employment) decisions generally manifest greater responsiveness to wage and income variation than do hours-of-work equations for workers" (Heckman, 1993, pg. 117).

mate the probability of participation and occupation based on the criteria described in (7), allowing the parameters on expected wages to vary across all outcomes will be misleading in terms of our theoretical model.[6] Instead of allowing expected wages to enter (6) as if they were characteristic of the individual, we instead restrict δ_j to be the same for all expected wages (δ). Thus we interpret expected wages as an *attribute* of the occupation rather than a characteristic of the individual.[7]

Before estimating the model, we need one further assumption. The random components of (6), η_{ij}, can follow many distributions, e.g. normal, poisson, extreme value or a combination of various distributions (logit kernel or mixed logit). For simplicity, we assume that η_{ij} are i.i.d. with extreme value distribution. With all our assumptions at hand, the probability that agent i chooses occupation s is defined as:

$$Pr(i = s) = \frac{exp[\delta \hat{w}_{is} + \boldsymbol{Z}_i \boldsymbol{\gamma}_s]}{\sum_{j=1}^{J} exp[\delta \hat{w}_{ij} + \boldsymbol{Z}_i \boldsymbol{\gamma}_j]} \tag{8}$$

Equation 8 combines attributes of the occupations, $\hat{w}$, with characteristics of the individual $\boldsymbol{Z}$. The specification is a combination of a conditional and a multinomial logit with $\hat{w}$ varying across individuals and occupations (see: Maddala, 1983, pg. 44).

[6] Say we normalize by making the parameters of outcome "not active" equal to zero. For every possible occupation, we will have $(J-1)$ expected wages, but the interpretation of all of them would be in terms of the base category (not active). In terms of our structural model, an increase in the expected wage in occupation J does not have an effect upon the probability of participating in occupation J' **relative to being not active**, therefore there is no basis for including all $(J-1)$ expected wages as if they were characteristics of the individuals.

[7] A sufficient assumption to have a single parameter for expected wages across all outcomes is: $\mid \gamma_j' \mid = \mid \hat{\beta}_j \mid$ (see equation 5). This is interpreted as making the marginal market value of characteristics $\boldsymbol{X}$ equal to its subjective valuation (in terms of reservation wage). In other words, every time a personal characteristic increases its market value, the individual will increase—in the same proportion—its subjective valuation.

Define wages, w, as a linear function of formal education, education interacting with a dummy variable for higher education or more, experience, experience squared and a regional dummy variable. These are the elements of $\boldsymbol{X}$. We allow for different parameters across occupation, estimating a separate wage equation for each occupation. Female workers can decide to *sell* their labour endowment in 3 different sectors: informal, manufacturing and other formal or not to *sell* it at all (not active). As we already specify in the last section, workers observed in each sector are not the outcome of a random process (they follow criteria (7)). Therefore the wage equations in each of the three remunerated sectors have to account for selectivity. Indeed, following criteria (7) we can obtain the conditional participation probabilities for each sector and given a parametrization rule, include them in the wage equation to control for selectivity.[8] The problem is that the conditional probabilities, $Pr(\cdot)$, are them selves a function of expected wages, therefore we have the following simultaneous equation model:

$$w_{ij} = w[\boldsymbol{X}_i, Pr(\hat{w}_{ij}, \boldsymbol{Z}_i)] \tag{9}$$

To solve for the simultaneity, we estimate (9) following a two-step procedure.[9] In the first step, we estimate $Pr(\cdot)$ using a reduced form with expected wages being substituted by its determinants $\boldsymbol{X}$: $Pr(\boldsymbol{X}, \boldsymbol{Z})$. In the second step, based on the results by Dubin and McFadden (1984), conditional probabilities, $Pr(\boldsymbol{X}, \boldsymbol{Z})$, are included in the wage equation in the following way (we exclude the individual subindex for clarity):

$$w_s = \boldsymbol{X}\beta_j + \sigma \sum_{j=1}^{J} \theta_j \left(\frac{Pr_j ln(Pr_j)}{1 - Pr_j} + ln(Pr_s) \right) + \varepsilon \quad \forall\, j \neq s \tag{10}$$

[8]For a discussion of the advantages and disadvantages of the different ways to control for selectivity using a multinomial logit, see Bourguignon, Fournier and Gurgand (2004).

[9]We could have estimated (9) using maximum likelihood techniques, however since we are interested in the robustness of our results to different selectivity-parametrization rules we opt for the two-step procedure.

Where σ is the variance of ε conditional on residuals from the wage and selection equations. Since the wages and the conditional probabilities from the selection equation share vector $\boldsymbol{X}$, the identifying variables are contained in $\boldsymbol{Z}$. We define $\boldsymbol{Z}$ bearing in mind that they must affect reservation wages but not market ones. In the case of female labourers, $\boldsymbol{Z}$ includes the number of children in the household; a dummy variable equal to one if the women is not head of the household and the head is active; the last element is the income of all other household members. These three variables are expected to have a significant effect upon female reservation wages without affecting their market remunerations.

Notice that estimation of informal sector "wages" using specification (10) implicitly assumes that this sector is *complete* and therefore it remunerates marginal productivity as the outcome of personal characteristics. Studies by Mercouiller, Ruiz and Woodruff (1997), Maloney (1999) and Gong, van Soest and Villagomez (2000) suggest that, controlling for personal characteristics, the informal sector in Mexico is a desirable destination rather than an inferior forced option. A special feature present in the informal sector that might be of advantage over its formal counterpart is the flexibility in working hours. To account for this *attribute* of each occupation, we include the standard deviation of working hours ($\tilde{h}$) in each sector as a determinant of participation and occupation.

Finally, a second multinomial logit [10] is estimated using $\hat{w}$ as the fitted values of (10) for each sector including $\tilde{h}$ and $\boldsymbol{Z}$ as regressors:

$$Prob(i = s) = \frac{exp[\delta \hat{w}_{is} + \varphi \tilde{h}_s + \boldsymbol{Z}_i \boldsymbol{\gamma}_s]}{\sum_{j=1}^{J} exp[\delta \hat{w}_{ij} + \varphi \tilde{h}_j + \boldsymbol{Z}_i \boldsymbol{\gamma}_j]} \tag{11}$$

[10] In our case a combination of a conditional and a multinomial logit. Maddala (1983) shows that this two models are mathematically the same, hence I will simply refer to it as a multinomial logit (MNL).

3 Data

The model described in section 2 is estimated using Mexican household survey data (ENIGH) for years 1994, 1996, 1998 and 2000. Between 1994 and 1996 the country experienced great macroeconomic turbulence as a result of the Peso crises. In 1994 Mexico embarked in a free trade agreement with Canada and the US. The years between 1996 and 2000 was a time of economic recovery with high rates of growth mainly boosted by manufacturing exports growth. All these changes might have had a significant effect upon various of the elements of the labour participation and occupation model.

In table 1 we show the average value of female participation, formal education achievements and real 1994 wages in each occupation. Given that ENIGH is not a probabilistic survey we account for sampling design taking expansion factors, stratification and clustering into account.[11] From table 1 we can see that female participation increased during the period of analysis with the proportion of non active women falling from 66.88% in 1994 to 62.50% in 2000. This might seem to be a small change but when we consider the total amount of women entering the labour market, the increase is far from being trivial. An increase in participation of 4.47 percentage points of the total working age population represents a total of 1,507,355 new entrants. Of that total amount, 987,349 of the new entrance took place in the urban sector with 627,480 going to manufacturing, 402,842 to other formal sectors and 42,973 **abandoning** the informal sector.[12] Between 1994 and 2000 the average years of schooling of females within working age rose from 7.54 to 8.24. Despite this overall and within other formal and informal sector increase in formal education, the average years of schooling of female workers in the manufacturing sector practically remained unchanged (if a change occurred that was a negative one). The combination of this two facts occurring in

[11] All the statistical analysis carried-on through out the paper corrects for survey design.

[12] This is obviously the net entrance in each sector, so far we cannot identify the sector of origin of the new entrant. We will address this important point in the next section.

the manufacturing sector, i.e. a large increase in female participation and a decreasing or constant average level of eduction, leads us to think that the manufacturing sector was absorbing relatively unskilled female labourers. This could be a sign that the boom in the manufacturing sector after the Peso devaluation and the enactment of NAFTA, made many unskilled Mexican housewives more likely to participate in the labour market. Finally, the last column of table 1 shows that by the year 2000 real female wages were still below those ones observed in 1994, most likely, as a result of the Peso crises.

Table 1: Average Values of Female Participation, Education and Wages

	1994			2000		
Occupation	% Working Age Pop.	Years of Schooling	Real Wage	% Working Age Pop.	Years of Schooling	Real Wage
Not Active	66.88	6.09	-	62.50	6.88	-
Man. Earner	4.17	7.75	2,339	5.99	7.72	2,012
Other Earner	16.01	8.87	3,423	17.18	9.63	2,927
Informal	11.18	4.59	1,024	11.05	6.31	919

Notes:

(1) The working age population consists of women between 13 and 65 years old without a physical impediment to work.

(2) Real wages are expressed in 1994 pesos.

(3) Data source: ENIGH 1994 and 2000 with expansion factors being taken into account.

4 Empirical Estimates of the Structural Model

4.1 Estimation Results

In this section we present the results from the estimation of models (10) and (11) for years 1994, 1996, 1998 and 2000.[13] Given the large amount of results the tables with results from estimation of (10) are placed in Appendix I.[14]

The first thing to notice from the wage results (table 1 in Appendix I) is that, in the manufacturing sector, there is no premium for high-level education (in our case when a worker has complete high school or more) as it is the case in other formal sectors and in the informal one for later years. This result suggests that the manufacturing sector in Mexico demands more unskilled female labourers relative to the skill-mix demanded in other formal sectors. Another important feature is that while returns to schooling fell in other formal sectors between 1996 and 1998 quite the opposite occurred in the manufacturing one. This could have been the result of an increase in labour demand explained by the rapid growth in exports in this sector during those years. The increase demand for female labour in the manufacturing sector after the Peso crises and the enactment of NAFTA is also supported by the significantly positive sign in the dummy variable for Northern states during the recovery years of 1998 to 2000.[15]

Table 2 shows the estimation results of (11) taking "not active" as the base

[13] Equation (10) is estimated using our own Stata command, *svyselmlog*. *svyselmlog* is the survey version of the original *selmlog* command written by M. Gurgand. The command estimates the parameters of the main equation (in this case wages) correcting for selectivity using a multinomial logit taking into account the survey design; several forms of selectivity correction are available. *svyselmlog* is available from: www/econ.cam.ac.uk/phd/red29.

[14] Due to space limitations we do not present the results from the first stage estimation, $Pr(\boldsymbol{X}, \boldsymbol{Z})$, however this are available from the author upon request.

[15] Mostly all foreign "maquiladoras" (or export processing zones) created after 1994 are located in the North of Mexico.

category; therefore all the parameters of the multinomial part of the equation are interpreted as changes in the probability of occupation relative to not being active.[16] The first two rows contain the estimates of two occupation attributes, i.e. expected wages ($\hat{w}$) and the standard deviation of working hours ($\tilde{h}$). Regarding the latter, the results show that female workers (or possible ones) value positively working hours stability, therefore an increase in the variance of working hours in a particular occupation reduces the probability of observing a worker in that sector. As we would expected *a priori*, an increase in $\hat{w}$ increases the likelihood of observing workers in that sector. Following our structural model, a value of δ equal to 0.5 translates into a value of λ (the marginal utility valuation of money income)of 1.5. Given an increase in $\boldsymbol{X}$ with its subsequent positive impact upon $\hat{w}$, the probability of observing a worker participating increases (i.e. $\delta > 0$) because the marginal valuation of money income is greater than the increase in the reservation wage. Let me elaborate in this point. From (4) we can see that an increase in $\boldsymbol{X}$ will increase the reservation wage in γ' units. The same increase in $\boldsymbol{X}$ will increase the agent's utility in $\lambda^{*}\hat{\beta}$ units. Since the agents' decision is based on the difference between the utility valuation of market wage and a reservation wage, a positive participation will be observed when $\lambda > 1$, given that the identifying assumption, $|\boldsymbol{\gamma'}_j| = |\hat{\boldsymbol{\beta}}_j|$, holds.

The rest of table 2 shows the effects that individual's characteristics have upon participation and occupation decisions. We can see that, despite an expected negative sign on the number of children, we find that when ever such a variable is significant it enters the equation with a positive sign. This effect is larger in the informal sector, suggesting that the extra child has a negative income effect prompting the mother into the informal sector. A very interesting pattern arose regarding labour participation dependence of

[16] In fact the parameters in table 2 are the marginal effects upon the latent function (equation 6) determining the probabilities. We do not show the marginal effects upon the probability of observing each outcome because we are more interested in the qualitative effect of the parameters than in their quantitative value.

spouses and daughters with respect the head of household labour status. We allow spouses and daughters to have a different response to the head of the household participation decision. H_s^a and H_d^a are dummy variables taking a value of one when the the individual is a non-head of household women (spouse or daughter respectively) and the head of the household is participating in the labour market. Although all parameters are negative and highly significant suggesting a positive income effect, the dependence of daughters is smaller in the manufacturing sector than in the informal one. More interesting, almost all dependence parameters show a steady negative time trend. Mexican labour market is changing in a way female workers participation decisions are less subject to their husband's or father's labour status. In a traditional society like the Mexican one, the reduction of this parameter can be interpreted as a positive sign in women's ability to pursue their professional development. As a result, their participation is increasing *ceteris paribus.*

Table 2: Structural Model Results

	1994	1996	1998	2000
$\hat{w}$	0.509***	0.703***	0.434***	0.560***
$\tilde{h}$	-0.152***	-0.120***	-0.104***	-0.156***
Man. Earner				
Intercept	-1.792***	-2.077***	-2.056***	-1.620***
Children	-0.038	0.106***	0.00	0.106***
H_s^a	-1.168***	-1.215***	-1.110***	-0.884***
H_d^a	-0.794***	-0.659***	-0.619***	-0.489***
Y_m^0	0.014***	0.019***	0.017***	0.015***
Other Earner				
Intercept	-0.356***	-0.920***	-0.421***	0.212**
Children	0.059***	-0.019	0.051**	-0.01
H_s^a	-1.093***	-1.056***	-1.028***	-0.958***
H_d^a	-1.104***	-1.161***	-1.041***	-1.099***
Y_m^0	0.016***	0.020***	0.019***	0.016***
Informal Sector				
Intercept	-	-	-	-
Children	0.145***	0.115***	0.137***	0.187***
H_s^a	-0.291***	-0.228**	-0.556***	-0.259***
H_d^a	-1.276***	-1.273***	-1.017***	-0.854***
Y_m^0	0.012***	0.017***	0.018***	0.015***
R^2	0.378	0.352	0.313	0.316
N	41476	46504	35368	31696

*,**,***, significant at the 10%, 5% and 1% level respectively

Data source: ENIGH years 1994 to 2000, corrected for survey design

The final individual characteristic that we include in $\boldsymbol{Z}$ is the income of all other household members, Y_m^0. All these exogenous income effects are positive, indicating that leisure is seen an "inferior". Although this result is counter-intuitive, Heckman and Sedlacek (1985) find the same result using data for the US. Moreover, the authors report that such a result is an empirical regularity (Heckman and Sedlacek, pg. 1100).

4.2 Micro-simulation Analysis

How much of the total increase in female labour participation between 1994 and 2000 can we attribute to changes in the parameters of the model? How much of the net entrants in the different occupations reported in section 3 were previously not active? Using micro-simulation techniques we are able to answer these important questions.

Define $\boldsymbol{X}_t^*$ as a vector containing all variables explaining female labour participation and occupation at time t:

$$\boldsymbol{X}_t^* = (\hat{w}_t, \tilde{h}_t, \boldsymbol{Z}_t)$$

Similarly, define Ω_t as a vector containing all estimated parameters for time t (those ones shown in table 2):

$$\Omega_t = (\hat{\delta}_t, \hat{\varphi}_t, \hat{\boldsymbol{\gamma}}_t)$$

Finally introduce a time subindex in the random component of the utility function, η_t. Female labour participation and occupation decisions at time t is hence a function of the three components just defined:

$$Pr(\cdot)_t = Pr(\Omega_t, \boldsymbol{X}_t^*, \eta_t) \tag{12}$$

After estimation of (11), the three components of (12) are observable and many illustrative counterfactual exercises can be undertaken. A natural first step is to ask our selves the participation/occupation impacts of parameter changes during the period of analysis. Such a counterfactual will answer the question: what would the participation and occupation structure of female labour market look like if the parameters in 1994 where equal to those ones observed in 2000? In terms of (12):

$$Pr(\cdot)^{\imath}_{1994} = Pr(\Omega_{2000}, \boldsymbol{X}^{*}_{1994}, \eta_{1994}) \qquad (13)$$

Where $Pr(\cdot)^{\imath}_{1994}$ is a *simulated* probability (since it is not observed) at the *micro* level (one for each individual). In effect what we are doing is "importing" the estimated parameters using data for year 2000 into the dataset for 1994. Once the parameters are in the 1994 database we multiply them by vector $\boldsymbol{X}^{*}$, add the residual terms η and finally allow each women within working age (worker and potential worker) to decide her labour status following criteria (7). The results are shown in table 3.

To understand how to read table 3, take the "Totals" of the table, i.e. the last column and the fifth row. On the far right of the table we will find the amount of female labourers (in thousands) observed in each occupation in 1994. After Ω_{2000} is being "imported" and (13) being computed, the simulated number of labourers in each sector is shown in row "Total Simulated". In the penultimate row of table 3 we show the net entrants in each occupation (it is the difference between the observed and simulated total amount of workers for each occupation). We can see that, had Ω been the only element changing in (12) between 1994 and 2000, then more than 2 million women would have abandoned inactivity. Changes in parameters account for more than the total increase in labour participation (almost a million new entrants in the urban areas as documented in section 3). Net participation in all three remunerated occupations increase as a result of the change in parameters.

The inner rows and columns in table 3 show the labour transitions from the observed 1994 occupation (or inactivity) to the simulated choice, with the diagonal containing workers whose choice didn't change after the simulation. Take the third column ("Manufacturing Earner"), the simulation estimates that 349 thousand of previously inactive women would have chosen to work in the manufacturing sector if the value of the parameters in 1994 were as those ones observed in 2000. The rest of the figures in the same column are read in a similar way; 900 thousand workers did not move from that sector, 45 thousand shifted from other formal sectors to the manufacturing one and finally 33 thousand new entrants were previously in the informal sector.

Table 3: Simulation Results ($\Omega_{1994} \rightarrow \Omega_{2000}$)

	Simulated Occupation				
Observed Occupation	Not Active	Manufacturing Earner	Other Earner	Informal Sector	Total Observed
Not Active	9,948	349	1,496	284	12,080
Man. Earner	0	900	26	1	928
O. Earner	0	45	3,874	39	3,958
Informal Sector	0	33	89	1,476	1,598
Total Simulated	9,948	1,329	5,488	1,801	18,564
Net entrants (NE)	-2,131	401	1,530	202	
New participants/NE	0	0.87	0.98	1.40	

The micro-simulation results show that the changes in parameters between 1994 and 2000 had a positive effect upon female labour participation. All three remunerated occupations would have seen an increase in the amount of female labourers had the parameters in 1994 were equal to those ones observed in 2000. These results suggest a change towards a more progressive labour market, with women being more likely and prompt to participate in it.

5 Summary and Conclusions

We develop a structural model describing female labour participation and occupation decisions. Women decision concerning their participation and occupation in the labour market is done in a simultaneous way since the decision for one of them is embedded in the other. The structural model is used to define and interpret a microeconometric model suitable for estimation. We correct for selectivity in the wage equations parameterizing the conditional probabilities deriving from a mutinomial logit as suggested by Dubin and McFadden (1984). Our model creates an explicit and causal relationship between expected selectivity-adjusted expected wages and partcipation/occupation decisions. Other factors like the labour status of the head of the household and number of children are used as regressors within a multinomial logit framework.

We apply the model to Mexican household data for years 1994, 1996, 1998 and 2000. Our results show that the probability of participation/occupation is positively affected by wages, implying a marginal valuation of monetary income of more than one. We also find out that women attach a positive value to working hours stability. One of the most interesting results was the decrease in the dependence between women's participation and the head of the household labour status. Through out the period of analysis the negative participation effect of women living in a household headed by an active men show to be decreasing.

Using micro-simulation techniques we were able to evaluate the female participation/occupation effects of changes in parameters between 1994 and 2000. The results show that had the change in parameters been the only difference between 1994 and 2000, then more than 2 million women would have abandoned inactivity.

References

[1] Bourguignon, F., Fournier, M. and Gurgand, M. (1998) 'Wage and labour force participation in Taiwan', *Mimeo* DELTA, Paris.

[2] Bourguignon, F., Fournier, M. and Gurgand, M. (2004) 'Selection bias correction based on the multinomial logit model: Monte-Carlo comparisons', *Mimeo* DELTA, Paris (http://www.crest.fr/pageperso/lmi/gurgand/selmlog.htm)

[3] Deaton, A. and Muellbauer, J. (1980) *'Economics and consumer behavior'*, New York: Cambridge University Press.

[4] Dubin, J.A. and McFadden, D. (1984) 'An econometric analysis of residential electric appliance holdings and consumption', *Econometrica*, vol. 52, 2.

[5] Gong, X. and van Soest, A. (2002) 'Wage differentials and mobility in the urban labor market: A panel data analysis for Mexico', *Labour Economics*, vol. 9.

[6] Gong, X., van Soest, A. and Villagomez, E. (2000) 'Mobility in the urban labor market: a panel data analysis for Mexico', *CentER discussion paper*, Tulburg University.

[7] Heckman, J. (1974) 'Shadow prices, market wages and labor supply', *Econometrica*, vol. 42, 4.

[8] Heckman, J. (1990) 'Varieties of selection bias', *The American Economic Review*, vol. 80, 2.

[9] Heckman, J. (1993) 'What has been learned about labor supply in the past twenty years?', *The American Economic Review*, vol. 83, 2.

[10] Heckman, J. and Honore, B.E. (1990) 'The empirical content of the Roy model', *Econometrica*, vol. 58, 5.

[11] Heckman, J. and Sedlacek, G. (1985) 'Heterogeneity, aggregation, and market wage functions: An empirical model of self-selelction in the labor market', *Econometrica*, vol. 93, 6.

[12] Lee, L.F. (1983) 'Generalized econometric models with selectivity', *Econometrica*, vol. 51, pp. 507-512.

[13] Maddala, G. (1983) *'Limited dependent and quantitative variables in econometrics'*, Cambridge: Cambridge University Press.

[14] Maloney, W. (1999) 'Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico', *The World Bank Economic Review*, vol. 13, 2.

[15] McFadden, D (1974), 'The measurement of urban travel demand', *Journal of Public Economics*, vol. 3.

[16] Mercouiller, D., Ruiz, V. and Woodruff, C. (1997) 'Formal measures of the informal sector wage gap in Mexico, El Salvador, and Peru, *Economic Development and Cultural Change*, vol. 45, 2.

[17] Roy, A.D. (1951) 'Some thoughts on the distribution of earnings', *Oxford Economic Papers*, vol. 3.

A I: Selectivity-Corrected Wages

Table 1: Wage functions for the Manufacturing Sector

	1994	1996	1998	2000
Schooling	0.131***	0.090***	0.139***	0.096***
$Schooling * I(Ys > 11)$	0.016	0.050***	-0.004	0.005
Experience	0.014	0.016	0.046***	0.016
$Experience^2$	0	0	-0.000*	0
North	0.065	0.123	0.335***	0.249**
Pr(not active)*	-0.348***	-0.224**	-0.042	-0.23
Pr(other earner)*	-0.127	0.106	0.362*	0.186
Pr(informal)*	0.634**	0.122	0.091	-0.04
Intercept	1.212*	0.158	0.036	-0.124
R^2				
N	491	609	511	428

Notes:

(1) *,**,***, significant at the 10%, 5% and 1% level respectively

(2) Data source: ENIGH 1994, 1996, 1998 and 2000 corrected for survey design

(3) $Pr(\cdot)^*$ are computed accordingly to equation 10

Table 2: Wage functions for Other Earning Sectors

	1994	1996	1998	2000
Schooling	0.148***	0.188***	0.144***	0.130***
$Schooling * I(Ys > 11)$	0.021***	0.001	0.029***	0.023***
Experience	0.020*	0.029***	0.039***	0.040***
$Experience^2$	0	-0.000*	0	-0.000**
North	-0.042	0.043	0.064	0.071
Pr(not active)*	-0.177***	-0.144***	-0.007	-0.069***
Pr(manufacture)*	0.093*	0.079	0.150***	0.107*
Pr(informal)*	0.442***	0.480***	0.236**	0.131*
Intercept	1.168***	0.650**	0.175	0.081
R^2				
N	2219	2400	1955	1855

Notes:

(1) *,**,***, significant at the 10%, 5% and 1% level respectively

(2) Data source: ENIGH 1994, 1996, 1998 and 2000 corrected for survey design

(3) $Pr(\cdot)^*$ are computed accordingly to equation 10

Table 3: Wage functions for the Informal Sector

	1994	1996	1998	2000
Schooling	0.060**	-0.027	0.062**	0.081*
$Schooling * I(Ys > 11)$	0.023	-0.001	0.118***	0.085**
Experience	-0.017	-0.022	0.064***	0.040**
$Experience^2$	0.001**	0.001	0	0
North	-0.118	0.277	-0.347	-0.607
Pr(not active)*	-0.154	-0.403	-0.028	0.053
Pr(manufacture)*	0.136	-0.422	0.696*	0.907**
Pr(other earner)*	0.565**	1.117	-0.935*	-0.71
Intercept	1.262*	0.789	-1.154**	0.181
R^2				
N	794	1072	868	753

Notes:

(1) *,**,***, significant at the 10%, 5% and 1% level respectively

(2) Data source: ENIGH 1994, 1996, 1998 and 2000 corrected for survey design

(3) $Pr(\cdot)^*$ are computed accordingly to equation 10