Gender Gaps in Latin American Labor Markets: Implications from an Estimated Search Model.*

Mauricio Tejada† Claudia Piras‡ Luca Flabbi§ Monserrat Bustelo¶

February 5, 2021

Web Appendix - Not For Publication

Contents

A Model 2

B Data 4

C Likelihood Function 10

D Complete Identification Discussion 12

E Complete Estimation Results 16

F Additional Material on Policy Experiments 41

G Robustness Analysis 47

*The findings, interpretations, and conclusions expressed in this paper are those of the authors and do not necessarily represent the views of the Inter-American Development Bank, its Executive Directors, or the governments they represent.

†Department of Economics, Universidad Alberto Hurtado, Erasmo Escala 1835 Office 211, Santiago, Chile. E-Mail: matejada@uahurtado.cl

‡Inter-American Development Bank, 1300 New York Avenue, N.W., Washington, D.C. 20577. E-mail: claudiapi@iadb.org.

§Department of Economics, University of North Carolina - Chapel Hill, 107 Gardner Hall, CB 3305 Chapel Hill, NC 27599-3305. E-mail: luca.flabbi@unc.edu.

¶Inter-American Development Bank, 1300 New York Avenue, N.W., Washington, D.C. 20577. E-mail: monserratb@iadb.org.

A Model

The equilibrium of the model has a simple structure. Agents have to make two discrete choices. The first concerns labor market participation: either they participate in the labor market looking for a job (state U_i) or they stay out enjoying utility from out-of-labor-market activities (state NP_i). Since agents receive different utility from these activities (z), those receiving relative high utility will stay out, those receiving relative low utility will enter the market. The threshold for staying out or coming in is determined by the indifference point between the two states, i.e. by the specific z_i^* such that:

$$NP_i(z_i^*) = U_i \Leftrightarrow z_i^* = \rho U_i \tag{A.1}$$

All agents with $z_i < z_i^*$ participate in the labor market; all those with $z_i > z_i^*$ stay out.

The second discrete choice the agents have to make concerns the labor market state decision: either they accept a job offer or they reject it and continue searching. Again we can identify a threshold: if the productivity and therefore the wage is high enough, they will accept; if not, they will continue searching for a better offer. As before, the threshold is identified by the indifference point between the two alternatives, i.e. by the specific x_{ij}^* such that:

$$U_i = E_{iF}(x_{iF}^*) \quad \Leftrightarrow \quad x_{iF}^* = (1+\tau)\rho U_i \tag{A.2}$$

$$U_i = E_{iI}(x_{iI}^*) \quad \Leftrightarrow \quad x_{iI}^* = \rho U_i + c \tag{A.3}$$

$$U_i = E_{iS}(x_{iS}^*) \quad \Leftrightarrow \quad x_{iS}^* = \rho U_i \tag{A.4}$$

These threshold have a straightforward economic interpretation. Employee jobs require higher productivity to be acceptable than self-employed job because in the first case the worker has to share with the employer. Moreover, the employer has to pay either payroll contributions or illegality costs and therefore the thresholds are increasing in those parameters.

The optimal decision rules and wages schedules can now be incorporated in the value of unemployment defined in equation (2), leading to the following equilibrium equation:

$$\begin{aligned}\rho U_i &= b_i + \frac{\beta\lambda_{iF}}{\rho+\delta_{iF}}\int_{(1+\tau)\rho U_i}[x-(1+\tau)\rho U_i]\,dG_{iF}(x)\\&+\frac{\beta\lambda_{iI}}{\rho+\delta_{iI}}\int_{\rho U_i+c}[x-c-\rho U_i]\,dG_{iI}(x)\\&+\frac{\lambda_{iS}}{\rho+\delta_{iS}}\int_{\rho U_i}[x-\rho U_i]\,dG_{iS}(x), \quad i=M,W \end{aligned} \tag{A.5}$$

The equation is a function of parameters and of the endogenous value of unemployment U_i. Under mild regularity conditions, it admits a unique solution. Given a solution for U_i, all the optimal decisions described in equations (A.1)–(A.4) are fully characterized.

To close the steady state equilibrium, we have to impose that all inflows and outflows in and from each labor market state are equal. The gender specific hazard rate out of unemployment to a job type j is $h_{ij} = \lambda_{ij}\left[1-G_{ij}(x^*_{ij})\right]$, i.e. the probability of receiving an offer times the probability of accepting the offer. The hazard rate out of employment type j is exogenous and equal to δ_{ij}. By denoting with e_{ij} the proportion of type i agents working in job type j and with u_i the proportion of type i agents searching for a job, the steady state conditions are:

$$\lambda_{iF}\left[1-G_{iF}(x^*_{iF})\right]u_i = \delta_{iF}e_{iF} \tag{A.6}$$

$$\lambda_{iI}\left[1-G_{iI}(x^*_{iI})\right]u_i = \delta_{iI}e_{iI} \tag{A.7}$$

$$\lambda_{iS}\left[1-G_{iF}(x^*_{iS})\right]u_i = \delta_{iS}e_{iS} \tag{A.8}$$

Adding the innocuous normalization that the labor force is measure 1, equations (A.6)–(A.8) produce the following solution:

$$u_i = \frac{\delta_{iF}\delta_{iI}\delta_{iS}}{h_{iF}\delta_{iI}\delta_{iS}+h_{iI}\delta_{iF}\delta_{iS}+h_{iS}\delta_{iF}\delta_{iI}+\delta_{iF}\delta_{iI}\delta_{iS}} \tag{A.9}$$

$$e_{iF} = \frac{h_{iF}\delta_{iI}\delta_{iS}}{h_{iF}\delta_{iI}\delta_{iS}+h_{iI}\delta_{iF}\delta_{iS}+h_{iS}\delta_{iF}\delta_{iI}+\delta_{iF}\delta_{iI}\delta_{iS}} \tag{A.10}$$

$$e_{iI} = \frac{h_{iI}\delta_{iF}\delta_{iS}}{h_{iF}\delta_{iI}\delta_{iS}+h_{iI}\delta_{iF}\delta_{iS}+h_{iS}\delta_{iF}\delta_{iI}+\delta_{iF}\delta_{iI}\delta_{iS}} \tag{A.11}$$

$$e_{iS} = \frac{h_{iS}\delta_{iF}\delta_{iI}}{h_{iF}\delta_{iI}\delta_{iS}+h_{iI}\delta_{iF}\delta_{iS}+h_{iS}\delta_{iF}\delta_{iI}+\delta_{iF}\delta_{iI}\delta_{iS}} \tag{A.12}$$

Finally, by denoting with NP_i the proportion of non-participant in the population, we exploit equation (A.1) to find:

$$NP_i = 1 - Q_i(z_i^*) \tag{A.13}$$

We are now ready to provide the following:

Definition 1 ***Equilibrium Definition.***

Given workers' types $i = W, M$ and employment states' type $j = F, I, S$, the vector of parameters $\{\rho, \lambda_{ij}, \delta_{ij}, b_i, c\}$, and the probability distribution functions $\{Q_i(z), G_{ij}(x)\}$ a ***search model equilibrium*** *in an economy with formal contribution rate τ is a set of values $\{U_i\}$ that:*

1. *solves the equilibrium equations* (A.5)*;*

2. *satisfies the steady state conditions* (A.9)–(A.13)*.*

The model is estimated assuming the data are extracted from a steady state defined following Definition 1. Policy and counterfactual will also be performed comparing different steady state at different parameters values. In these experiments, we will use, among others, a measure representing the total output of the labor market. Specifically, we will use two measures of the aggregated average output: the output per worker (Y^{pw}) and the output per-capita (Y^{pc}). The former divides the total production by mass of workers that are currently in a job, while the latter divides the total production by the overall population, including the non-participant. We anticipate here the definitions of these two metrics. For given gender i we define:

$$\begin{aligned} Y_i^{pw} &= \frac{e_{iF}}{1-u_i}\int_{x_{iF}^*} x dG_{iF}(x) + \frac{e_{iI}}{1-u_i}\int_{x_{iI}^*} x dG_{iI}(x) + \frac{e_{iS}}{1-u_i}\int_{x_{iS}^*} x dG_{iS}(x) \\ Y_i^{pc} &= (1-NP_i)\left(e_{iF}\int_{x_{iF}^*} x dG_{iF}(x) + e_{iI}\int_{x_{iI}^*} x dG_{iI}(x) + e_{iS}\int_{x_{iS}^*} x dG_{iS}(x)\right) \end{aligned}$$

They are straightforward averages over the equilibrium measures and distributions of each labor market state in equilibrium.

B Data

We use data from household surveys and employment surveys from five LAC countries: Argentina, Chile, Colombia, and Mexico. In each country, we use the latest available survey leading to survey dates ranging from the third quarter of 2014 to the last quarter of 2016.

In the case of Argentina, we use the *National Survey of Urban Households* (EAHU) conducted in the third quarter of 2014. It is a representative household survey collected by the *National Institute of Statistics and Census* (INDEC) with a cross-sectional structure and reporting information on education, labor force variables and income. In the case of Chile, we use the *National Socio-Economic Characterization Survey* (CASEN) of 2015. It is conducted between November 2015 and January 2016. It is a cross-sectional household survey representative at a national level and reports information on education, labor force, income, and health status. In the case of Colombia, we use the *Great Integrated Household Survey* (GEIH) of the last quarter of 2016. It is a monthly cross-sectional household survey describing labor force status, the quality of life, income and expenditures. Finally, for Mexico we use the *National Occupation and Employment Survey* (ENOE) of the last quarter of 2016. It is a quarterly cross-sectional employment survey focusing on labor markets status and characteristics.

To build the estimation samples, we extract all the individuals aged between 25 and 55 years old and working in non-agricultural activities. Both restrictions are motivated by ensuring a more homogeneous sample of workers. Labor market careers typically exhibit life-cycle patterns. Our approach is not well equipped to capture them and therefore our age restrictions eliminates some of the major life-cycle dynamics (such as retirement concerns or first-entrants).[1] A shorter age range would have guaranteed more homogeneity but the cost in terms of sample size would have been too large, in particular on some countries. The compromise we reached by considering only 25-55 years old generates an age range similar to the one used in comparable literature.[2] The focus on non-agricultural activities is dictated by the theoretical model. Our proposed search model with bargaining is a good – and commonly used – description of labor markets characterized by a clear division of labor and by work for pay. These characteristics are less predominant in the agricultural sectors of most of the countries under consideration and therefore our theoretical model would have not been a good description of them. Nevertheless, it is important to keep in mind that the share of the labor force working in the agricultural sector in Latin America is relevant. In our sample, as can be seen in Table B.1, this is particularly true for male workers with primary education in all countries, with the share of the labor force working in this sector ranging between 20 and 26% in Colombia and Chile, respectively. For women

[1]Incorporating life-cycle effects in search model of the labor market is notoriously problematic and definitely out of the question with the data at our disposal. Two rare exceptions are Bagger et al. (2014) and Pavan (2011), both of which used long and rich panel data to estimate their models.

[2]For example, Bobba et al. (2017) use 35-55 years old; Meghir et al. (2015) 23-65 years old; Flabbi (2010) 30-55 years old; and Dey and Flinn (2005) 25-54 years old.

with primary education, the share of the agricultural sector drops to a range between 3 and 8% for again Colombia and Chile, respectively. In turn, for secondary education the share of the agricultural sector are considerably lower compared with those of the primary education, being the highest observed in Chile (respectively 9 and 4% for men and women) and Mexico (for men 8%). Finally, as expected, the share of the agricultural sector drops sharply for tertiary education.[3]

We then divide the sample based on the highest level of education completed: primary school or less, secondary school, and tertiary level degree and above. We define four labor market states from the observed data: Unemployed, Formally employed as employee, Informally employed as employee, Self-employed. We also consider the state of no labor market participation. For employed workers we use information about the primary occupation in each sector, formal, informal and self-employment. More than one occupation are not so common in our sample, particularly for primary and secondary education levels. Table B.2 show, the percentage of worker in our sample that have only one occupation, their primary occupation. As can be noticed, at most 3.5 and 5.4% of men and women in primary education, respectively, have more than one occupation (both observed for Argentina), while in secondary education mostly 4% have more than one occupation regardless of gender (again the highest percentages observed for Argentina). For the tertiary education, more occupations are slightly more common, particularly in Argentina and Chile where between 12 and 7% of workers does not have only one occupation.

Following Kanbur (2009) and Levy (2008), we define informal employees as those who are not contributing to the social security system. In most LAC countries, firms are obligated to enroll salaried workers in the social security system and pay contributions which are approximately proportional to wages. Observing this registration in labor market data is considered in the literature a reliable measure of informal employment. Self-employed workers have typically different requirements but they rarely enroll and pay contribution in the system. The overall informal sector is therefore frequently considered the sum of the self-employed and the informal employees (Bobba et al., 2017; Meghir et al., 2015).

When considering women, we also report the presence of young children in the household. We consider two cutoffs based on schooling age: for pre-schoolers we use the cutoff at 5 years of age and for primary and lower-secondary we use the cutoff at 13 age of age. In this way, we are able to identify women with children who are still not old enough to be enrolled in compulsory schooling and women with children who are in the age range typically covered

[3]It is worth to mention that for the case of Argentina, we are not able compute the exact share of the agricultural sector because the survey only covers the urban areas.

by compulsory schooling in the region. Conditioning on the presence of children allows us to capture some of the life-cycle effects that we are forced to ignore given the limitations of our data. We infer the realtionship between children and the adults in our estimation sample in the following way. In the data, we observe the presence and age range of children in the household and the relationship of each household member with the head of household (HH). Crossing this information, we can proxy the child care responsibilities of the women in our sample in the following way. As mentioned, our estimation sample is composed by two sets of adults. The first and by far the largest set is composed by HH and by HH's spouses. In this case, we assume that if a child is the son or daughter of the HH then the HH and the HH's spouse have the main child care responsability of them. The second set is composed by the adult children of HH living at home. We assign childcare responsabilites to these living-at-home adult if in the same household there are grandchildren of the HH.

Finally, our model is constructed to analyze the extensive margin of employment and the determination of hourly wages, leaving out the intensive margin or the determination of hours worked. To have an sense of the relative importance of the contribution of hourly wages, hours worked (the intensive margin) and the probability of being employed the (extensive margin) in the overall wage gap, we make a "fourth-fold" decomposition of the unconditional weekly wage gap in our sample (see for example Daymont and Andrisani, 1984), that is:

$$\begin{aligned} W_M^{UNC} - W_W^{UNC} &= W_M P_M - W_W P_W \\ &= w_M h_M P_M - w_W h_W P_W \\ &= (w_M - w_W)\, h_W P_W + (h_M - h_W)\, w_W P_W + (P_M - P_W)\, w_W h_W \\ &\quad + (w_M - w_W)(h_M - h_W)\, P_W + (w_M - w_W)(P_M - P_W)\, h_W \\ &\quad + (h_M - h_W)(P_M - P_W)\, w_W + (w_M - w_W)(h_M - h_W)(P_M - P_W) \\ &= \Delta w + \Delta h + \Delta P + \Delta I \end{aligned}$$

where the first term Δw is the pure contribution of the hourly wage gap, the second term Δh is the pure contribution of the weekly hours worked gap, the third term ΔP is the pure contribution of the probability of participating and being employed gap, and finally, the last term ΔI is an interaction term accounting for the fact that differences in w, h and P exist simultaneously between men and women. The results are shown in Table B.3. Two comment are worth to mention. First, the hourly wage gap explain between 24 and 36% of the total gap, while the gap in the probability of being employed account between 18 and 33% of the total gap. These two components, which are captured in our model, account for more than

40% of the total gap. Second, the gap in hourly weekly hours is more relevant for worker with tertiary education; it explain between 18 and 36% of the total gap. For workers with less education levels, primary and secondary, the gap in hours explain at most 11%.

Table B.1: Share of the agricultural sector

	Argentina(*)	Chile	Colombia	Mexico
		Men		
Primary	5.1	25.7	20.1	24.6
Secondary	1.9	9.3	3.9	7.7
Tertiary	1.5	3.5	1.6	1.5
		Women		
Primary	0.5	8.1	3.2	2.3
Secondary	0.1	4.0	0.9	0.7
Tertiary	0.2	1.4	0.4	0.2

(*) Survey covering only urban areas.

Table B.2: Percentage of workers with only one job

	Argentina	Chile	Colombia	Mexico
		Men		
Primary	96.5	97.6	98.9	97.0
Secondary	96.1	97.5	99.0	96.3
Tertiary	87.4	93.1	98.0	94.8
		Women		
Primary	94.8	99.1	98.8	99.0
Secondary	96.0	98.5	98.5	98.3
Tertiary	88.8	95.5	98.6	96.8

Table B.3: Wage differential decomposition

	Argentina	Chile	Colombia	Mexico
Gap due to hourly wages: Δw				
Primary	0.24	0.31	0.27	0.28
Secondary	0.29	0.35	0.31	0.31
Tertiary	0.24	0.36	0.34	0.29
Gap due to weekly hours: Δh				
Primary	0.08	0.04	0.07	0.06
Secondary	0.11	0.06	0.09	0.10
Tertiary	0.36	0.18	0.20	0.27
Gap due to the probability of being employed: ΔP				
Primary	0.18	0.23	0.19	0.20
Secondary	0.22	0.26	0.24	0.23
Tertiary	0.22	0.33	0.32	0.27
Gap due to the interactions: ΔI				
Primary	0.50	0.42	0.47	0.46
Secondary	0.38	0.33	0.36	0.37
Tertiary	0.19	0.14	0.14	0.16

C Likelihood Function

We introduce the notation $k = 1, 2, 3...N_i$ to denote an individual observation in the sample.

The probability of observing an individual k non participating in the labor market is $P(z > z^*)$ (see equation A.13). Given the assumption on the distribution of z, $Q(z)$, and the reservation value of the participation decision, $z^* = \rho U_i$, the contribution to the likelihood of the non participation information is:

$$P_i(k \in NP_i) = 1 - Q(\rho U_i) \tag{C.1}$$

To find the contribution of the unemployment duration information to the likelihood we first define the total hazard rate out of unemployment. Because our model features multi-exits to different types of employment, the total hazard rate out of unemployment is comprised of the different hazards from unemployment to each job type: $h_i = h_{iF} + h_{iI} + h_{iS}$. Each hazard is defined as the probability that a match is formed once an individual meets a potential employer or a self-employment opportunity (see equations A.6–A.8).

The hazard rate, conditional on the model, does not exhibit duration dependence. At the same time, the durations observed in the sample are on-going. As a result, the unemployment duration follows a negative exponential distribution with coefficient equal to the hazard rate. Given that the unemployment duration is observed only for individuals who are actively participating in the labor market and are currently unemployed, the actual likelihood contribution of an unemployed individual k is the joint density of participating ($Q(\rho U_i)$), being unemployed (u_i as defined in equation A.9) and observing a duration $t_{i,k}$, leading to:[4]

$$f_{i,u}(t_{i,k}, k \in U_i, k \notin NP_i) = h_i \exp(-h_i t_{i,k}) u_i Q(\rho U_i) \tag{C.2}$$

To derive the contribution of wages and self-employed income to the likelihood function, it is necessary to take into account three features. First, we have information on wages but not on productivity. Second, the observed wages are those related to matches already formed therefore, in terms of the model, they are accepted wages. Third, we only observe data for those individuals who are currently employed or self-employed.

To take into account these data features, we proceed in the following way. In the first step, we map the unconditional wage cumulative distribution from the unconditional productivity

[4] In the particular case of Argentina, where the structure of the duration data is defined as intervals, the contribution of the unemployment duration information uses $\left[1 - e^{-h_j t_s^{(2)}}\right] - \left[1 + e^{-h_j t_s^{(1)}}\right]$, for the interval of durations $t_s^{(2)} - t_s^{(1)}$, instead of the negative exponential density function.

cumulative distribution ($G_{ij}(x)$) using the wage equations (9)–(10) (for the self-employed, productivity and income coincides). In the second step, we construct the truncated version of the distributions taking into account the optimal decisions rules summarized by the reservation values (x^*_{ij}). In the third step, we use the truncated wages distributions, the probability of participating ($Q(\rho U_i)$) and the probability of being employed (e_{ij} as defined in equations A.10–A.12) to compute the joint density of observed wages. In conclusion, the contributions to the likelihood function for agent k in, respectively, formal employment, informal employment and self-employment are:

$$f_{e_{iF}}(w_{i,k}, w_{i,k} \geq w^*_{iF}, k \in E_{iF}, k \notin NP_i) = \frac{\frac{1+\tau}{\beta} g_{iF}\left(\frac{(1+\tau)(w_{i,k}-(1-\beta)\rho U_i)}{\beta}\right)}{1 - G_{iF}((1+\tau)\rho U_i)} e_{iF} Q(\rho U_i) \quad (C.3)$$

$$f_{e_{iI}}(w_{i,k}, w_{i,k} \geq w^*_{iI}, k \in E_{iI}, k \notin NP_i) = \frac{\frac{1}{\beta} g_{iI}\left(\frac{w_{i,k}+\beta c-(1-\beta)\rho U_i}{\beta}\right)}{1 - G_{iI}(\rho U_i + c)} e_{iI} Q(\rho U_i) \quad (C.4)$$

$$f_{e_{iS}}(w_{i,k}, w_{i,k} \geq w^*_{iS}, k \in E_{iS}, k \notin NP_i) = \frac{g_{iI}(w_{i,k})}{1 - G_{iS}(\rho U_i)} e_{iS} Q(\rho U_i) \quad (C.5)$$

We are now ready to proposed the overall loglikelihood function used to identify and estimate the model:

$$\begin{aligned}
\ln L(w_k, t_k, i; \Theta) = & \sum_{i=M,W} \{ N_{NP_i} \ln(1 - Q(\rho U_i)) \\
& + (N_{U_i} + N_{E_{iF}} + N_{E_{iI}} + N_{E_{iS}}) \ln Q(\rho U_i) + N_{U_i} \ln h_i \\
& + N_{U_i} \ln u_i + N_{E_{iF}} \ln e_{iF} + N_{E_{iI}} \ln e_{iI} + N_{E_{iS}} \ln e_{iS} \\
& - h_i \sum_{k \in U_i} t_{i,k} + \sum_{k \in F} \ln \left(\frac{\frac{1+\tau}{\beta} g_{iF}\left(\frac{(1+\tau)(w_{i,k}-(1-\beta)\rho U_i)}{\beta}\right)}{1 - G_{iF}((1+\tau)\rho U_i)} \right) \\
& + \sum_{k \in I} \ln \left(\frac{\frac{1}{\beta} g_{iI}\left(\frac{w_{i,k}+\beta c-(1-\beta)\rho U_i}{\beta}\right)}{1 - G_{iI}(\rho U_i + c)} \right) \\
& + \sum_{k \in S} \ln \left(\frac{g_{iI}(w_{i,k})}{1 - G_{iS}(\rho U_i)} \right) \}
\end{aligned}$$

where $N_{NP_i}, N_{U_i}, N_{E_{iF}}, N_{E_{iI}}, N_{E_{iS}}$ are the sample sizes in each labor market state and Θ is the vector of the primitive parameters of the model.

D Complete Identification Discussion

Since the identification strategy applies in the same way to men and women, in what follows we drop the gender specific index i to reduce notation. Starting with the mobility parameters and taking the first order conditions of the maximization problem of the logarithm of the likelihood function with respect to the hazard rates, we obtain:

$$h_F \quad : \quad \frac{N_U}{h} + \frac{N_U}{u}\partial_{h_F} u + \frac{N_F}{e_F}\partial_{h_F} e_F + \frac{N_I}{e_I}\partial_{h_F} e_I + \frac{N_S}{e_S}\partial_{h_F} e_S - \sum_{k \in U_i} t_k = 0 \tag{D.1}$$

$$h_I \quad : \quad \frac{N_U}{h} + \frac{N_U}{u}\partial_{h_I} u + \frac{N_F}{e_F}\partial_{h_I} e_F + \frac{N_I}{e_I}\partial_{h_I} e_I + \frac{N_S}{e_S}\partial_{h_I} e_S - \sum_{k \in U_i} t_k = 0 \tag{D.2}$$

$$h_S \quad : \quad \frac{N_U}{h} + \frac{N_U}{u}\partial_{h_S} u + \frac{N_F}{e_F}\partial_{h_S} e_F + \frac{N_I}{e_I}\partial_{h_S} e_I + \frac{N_S}{e_S}\partial_{h_S} e_S - \sum_{k \in U_i} t_k = 0 \tag{D.3}$$

and with respect to the arrival rates of termination shocks, we obtain:

$$\delta_F \quad : \quad \frac{N_U}{u}\partial_{\delta_F} u + \frac{N_F}{e_F}\partial_{\delta_F} e_F + \frac{N_I}{e_I}\partial_{\delta_F} e_I + \frac{N_S}{e_S}\partial_{\delta_F} e_S = 0 \tag{D.4}$$

$$\delta_I \quad : \quad \frac{N_U}{u}\partial_{\delta_I} U_i + \frac{N_F}{e_F}\partial_{\delta_I} e_F + \frac{N_I}{e_I}\partial_{\delta_I} e_I + \frac{N_S}{e_S}\partial_{\delta_I} e_S = 0 \tag{D.5}$$

$$\delta_S \quad : \quad \frac{N_U}{u}\partial_{\delta_S} u + \frac{N_F}{e_F}\partial_{\delta_S} e_F + \frac{N_I}{e_I}\partial_{\delta_S} e_I + \frac{N_S}{e_S}\partial_{\delta_S} e_S = 0 \tag{D.6}$$

where $\partial_Y X$ is the partial derivative of the steady state condition X with respect to the parameter Y. Equations (D.1) to (D.6) a system of six nonlinear equations in six unknowns (h_j, δ_j). These parameters are exactly identified if the solution of this system of equations is unique. Given the nonlinearity and issues with empirical identification, we have chosen to follow Bobba et al. (2017) and restrict the set of possible solutions to those that satisfy $\lambda_F = \lambda_I$ and $\delta_F = \delta_I$. The constraints implies that employee jobs share the same arrival and termination rate.

Whit respect to the productivity distributions we assume, as discussed before, that they take a log-normal form. This particular parametrization meets the recoverability condition and belongs to a log location-scale family and therefore the identification of location and the scale of the original distribution should be identified from the location and the scale of the truncated distribution (see Eckstein and van den Berg, 2007, for a detailed discussion). To see this in the context of the distribution of the different types of jobs, we re-parametrize the observed wages distribution for the case of formal jobs in the following way:

$$\frac{\frac{1+\tau}{\beta} g_F\left(\frac{(1+\tau)(w_k-(1-\beta)\rho U)}{\beta}\right)}{1-G_F((1+\tau)\rho U_i)} = \frac{\frac{1}{w_k \sigma_{F,0}}\phi_F\left(\frac{\ln(w_k)-\mu_{F,0}}{\sigma_{F,0}}\right)}{1-\Phi_F\left(\frac{\ln(\rho U_i)-\mu_{F,0}}{\sigma_{F,0}}\right)}$$

where:

$$\mu_{F,0} = (1-\beta)\rho U_i + \frac{\beta}{1+\tau}\mu_F \tag{D.7}$$

$$\sigma_{F,0} = \frac{\beta}{1+\tau}\sigma_F \tag{D.8}$$

that is, $\mu_{F,0}$ and $\sigma_{F,0}$ are the mean (location) and standard deviation (scale) of the observed wages distribution, respectively, and μ_F and σ_F are the mean (location) and standard deviation (scale) of the productivity distribution. From (D.7) and (D.8) it follows immediately that if ρU_i, β and τ are known, then μ_F and σ_F are uniquely identified from the data on wages in the formal sector. The parameters β and τ are set at 0.5 for all countries and at the level of the payroll contributions in each country, respectively. While theoretical identification of β is assured by the model's implications and by the distributional assumptions, its empirical identification is challenging without demand side information[5] and that is why we simply calibrate the parameter to the value of symmetric Nash bargaining. This is definitely a restriction in our context since it force us to the set the same Nash bargaining parameter for men and women. Previous literature has shown that differences in β by gender are likely to be present and they are often interpreted as capturing discrimination or gender-specific attitudes toward negotiation.[6] Even if we have to impose the restriction, it is worth remembering that the presence of endogenous and gender-specific outside options (U_i) still allows the wages to capture differences in bargaining power between men and women. Since the outside option enters directly in the wage equations, a lower outside option for a given gender in a given schooling group translates into lower wages at same productivity compared with the other gender.[7]

Using the same re-parametrization for the observed wages distribution for the case of

[5]For a formal discussion, see Flinn (2006). For an implementation using demand-side information, see Cahuc et al. (2006).

[6]See for example, Bartolucci (2013). Eckstein and Wolpin (1999) and Borowczyk-Martins et al. (2017) are examples of a similar strategy applied to racial gaps instead of gender gaps.

[7]See equations 9 and 10.

informal jobs we have:

$$\frac{\frac{1}{\beta}g_I\left(\frac{w_k+\beta c-(1-\beta)\rho U_i}{\beta}\right)}{1-G_I\left(\rho U_i+c\right)}=\frac{\frac{1}{w_k\sigma_{I,0}}\phi_I\left(\frac{\ln(w_k)-\mu_{I,0}}{\sigma_{I,0}}\right)}{1-\Phi_I\left(\frac{\ln(\rho U_i)-\mu_{I,0}}{\sigma_{I,0}}\right)}$$

where:

$$\mu_{I,0} = (1-\beta)\rho U_i + \beta(\mu_I - c) \tag{D.9}$$

$$\sigma_{I,0} = \beta\sigma_I \tag{D.10}$$

In this case, μ_I and σ_I are uniquely identified from the data if ρU_i, β and c are known, which means that the cost of informality has to be set using additional sources of information in order to be able to identify the productivity distribution in the informal sector. To fix the parameter c, we use the ratio between the cost of informality and the average wage in the formal sector estimated by Bobba et al. (2017) for the case of Mexico and we use that ratio to set this parameter across countries. Finally, the re-parametrization of observed wages distribution for the case of self-employed workers gives:

$$\frac{g_I(w_k)}{1-G_S(\rho U_i)}=\frac{\frac{1}{w_k\sigma_{S,0}}\phi_S\left(\frac{\ln(w_k)-\mu_{S,0}}{\sigma_{S,0}}\right)}{1-\Phi_S\left(\frac{\ln(\rho U_i)-\mu_{S,0}}{\sigma_{S,0}}\right)}$$

where:

$$\mu_{S,0} = \mu_S \tag{D.11}$$

$$\sigma_{S,0} = \sigma_S \tag{D.12}$$

Given that there is no bargaining involved in self-employment, the identification of the location and the scale of the productivity distribution in equations (D.11) and (D.12) is identified one to one from their counterparts in the observed wages distribution provided that ρU_i is known.

To estimate ρU_i, Flinn and Heckman (1982) show that the minimum observed wage is a strongly consistent non parametric estimator of the reservation wage. This estimator is typically used in the literature. However, because an implication of our model is that $w_F(x_F^*) = w_I(x_I^*) = x_I^* = \rho U_i$, the Flinn and Heckman (1982) estimator requires that $\min w_F^o = \min w_I^o = \min w_S^o = \rho U_i$ but nothing guarantees that these equalities hold in the data. Instead, we attempt to estimate ρU_i jointly with all the other parameters maximizing the likelihood function. The problem that arises in this case is that ρU_i determines the reservation productivities, which in turn are the truncation parameters in the accepted wage distributions in all types of job, and changing this parameter in the maximization process of the likelihood function changes its support and violates one of the regularity conditions of the estimation method. To avoid this problem and because it is likely that wages are measured with error (particularly in self-employment), we introduce measurement error in the estimation.

We assumed that the measurement error ϵ is multiplicative, and therefore the observed wage can be expressed as $w^o = w \times \epsilon$. The assumptions we make about the measurement error are threefold: (1) the measurement error is gender specific; (2) we use a log-normal distribution for the measurement error: $v(\epsilon) = \frac{1}{\epsilon\sigma_\epsilon}\phi\left(\frac{\ln\epsilon-\mu_\epsilon}{\sigma_\epsilon}\right)$, where $\phi(\cdot)$ is the standard normal density function, $i = M, W$; and finally (3) we assume that the conditional expectation of the observed wages is equal to the true wages, that is $E[w^o|w] = w$, which implies that $E[\epsilon|w] = 1$. All these assumptions together imply that the parameters μ_ϵ and σ_ϵ satisfy $\sigma_\epsilon = \sqrt{-2\mu_\epsilon}$, and therefore only one parameter of the measurement error has to be estimated. Using the measurement error, the implied density functions of observed wages that should be used in the contributions of wages in all types of jobs to the likelihood function are:

$$f^o_{e_F}(w^o_k) = \int_{\rho U_i} \frac{1}{w} v\left(\frac{w^o_k}{w}\right) f_{e_F}(w, w \geq \rho U_i, k \in F, k \notin NP_i)dw \tag{D.13}$$

$$f^o_{e_I}(w^o_k) = \int_{\rho U_i} \frac{1}{w} v\left(\frac{w^o_k}{w}\right) f_{e_I}(w, w \geq w^*_I, k \in I, k \notin NP_i)dw \tag{D.14}$$

$$f^o_{e_S}(w^o_k) = \int_{\rho U_i} \frac{1}{w} v\left(\frac{w^o_k}{w}\right) f_{e_S}(w, w \geq w^*_S, k \in S, k \notin NP_i)dw \tag{D.15}$$

Finally, to identify the parameter γ in $Q(z)$, the assumed distribution is required to be invertible with respect to its parameter, and the negative exponential distribution meets this requirement. The first order condition of the maximum likelihood estimation gives the following estimator for this parameter:

$$\gamma = \frac{\ln\left(\frac{N}{N_{NP_i}}\right)}{\rho U_i}$$

where N is the total number of individuals and N_{NP_i} is the number of individuals who are not participating in the labor market. To analyze the influence of the presence of kids in the household on the participation rates (in particular in the γ parameter), we divided those non participating individuals into three groups. First those that have kids 5 years old or younger in the household ($k5$), second, those that have kids between 5 and 13 years old ($k13$), and third the remaining non participants ($other$). It can be shown that if $\Pr[NP_i \cap k5] + \Pr[NP_i \cap k13] + \Pr[NP_i \cap other] = \Pr[NP_i]$, the estimator of the parameter γ by group is:

$$\gamma_\kappa = \frac{\ln\left(\frac{N_\kappa}{N_{\kappa,NP_i}}\right)}{\rho U_i}$$

where N_κ is the total number of individuals in the group κ and N_{κ,NP_i} is the number of individuals who are not participating in the group κ.

E Complete Estimation Results

Tables E.1, E.6, E.11, and E.16 report the complete set of descriptive statistics for each country, gender and education group.

Tables E.2, E.7, E.12, and E.17 report the estimated structural parameters of the model for each country, gender and education group.

Tables E.3, E.8, E.13, and E.18, report the implications for the labor market dynamics and the distribution across labor market states, while tables E.4, E.9, E.14, and E.19, report the implications for wages and productivity.

As mentioned in the main text, we perform various policy experiments. Tables E.5, E.10, E.15, and E.20, report the impact of the policy experiments on a variety of labor market outcomes together with the same outcomes reported at benchmark.

Table E.1: Argentina - Descriptive Statistics

Labor Market States	N	Prop.	$\bar{t}_u$	$\bar{w}$	σ_w	N	Prop.	$\bar{t}_u$	$\bar{w}$	σ_w
	Men					Women				
Education Group: Primary										
Unemployed	400	0.05	2.78	-	-	311	0.04	3.33	-	-
Formal Emp.	2594	0.34	-	4.49	2.14	1070	0.14	-	3.78	1.75
Informal Emp.	1773	0.24	-	2.48	1.33	1584	0.21	-	2.60	1.56
Self-Emp.	2030	0.27	-	3.00	2.27	726	0.10	-	2.37	2.18
Non Part.	737	0.10	-	-	-	3946	0.52	-	-	-
$K \leq 5$						1750	0.44			
$5 < K \leq 13$						1091	0.28			
Education Group: Secondary										
Unemployed	190	0.04	3.02	-	-	219	0.05	3.58	-	-
Formal Emp.	2460	0.54	-	5.10	2.36	1426	0.30	-	4.66	2.19
Informal Emp.	665	0.14	-	2.84	1.65	712	0.15	-	2.78	1.78
Self-Emp.	1043	0.23	-	3.52	2.77	565	0.12	-	3.16	3.21
Non Part.	229	0.05	-	-	-	1837	0.39	-	-	-
$K \leq 5$						772	0.42			
$5 < K \leq 13$						485	0.26			
Education Group: Tertiary										
Unemployed	140	0.03	3.29	-	-	252	0.04	3.63	-	-
Formal Emp.	2555	0.59	-	6.73	3.35	3455	0.53	-	6.64	3.03
Informal Emp.	374	0.09	-	4.17	2.96	640	0.10	-	3.89	2.77
Self-Emp.	914	0.21	-	5.21	4.36	812	0.12	-	5.23	4.77
Non Part.	335	0.08	-	-	-	1344	0.21	-	-	-
$K \leq 5$						506	0.38			
$5 < K \leq 13$						292	0.22			

Note: Wage distributions are trimmed at the top and bottom 1 percentile by gender, education group and type of job, and are reported in US Dollars of December 2016 (Exchange Rate = 15.8620 Argentinian Pesos/US). A worker is categorized as informal if he/she reports not having benefits of social security. K means proportion of women with the presence of kids in the household with respect to non participating women. Unemployment durations ($\bar{t}_u$) are only observed in time intervals.

Table E.2: Argentina - Estimated Parameters

	Primary		Secondary		Tertiary	
	Men	Women	Men	Women	Men	Women
ρU	0.2010	0.1482	1.7532	1.4020	1.8743	1.6045
	(0.0452)	(0.0772)	(0.0548)	(0.0518)	(0.0776)	(0.0725)
λ_F	0.1291	0.1270	0.2148	0.1824	0.2090	0.2009
	(0.0064)	(0.0047)	(0.0113)	(0.0051)	(0.0104)	(0.0055)
λ_S	0.0991	0.0492	0.1435	0.1192	0.0857	0.0498
	(0.0158)	(0.0060)	(0.0188)	(0.0800)	(0.0041)	(0.0021)
δ_F	0.0235	0.0298	0.0166	0.0286	0.0115	0.0147
	(0.0010)	(0.0011)	(0.0009)	(0.0008)	(0.0006)	(0.0004)
δ_S	0.0194	0.0212	0.0106	0.0056	0.0100	0.0115
	(0.0012)	(0.0026)	(0.0012)	(0.0011)	(0.0003)	(0.0005)
μ_F	2.5652	2.3973	2.5337	2.4788	2.8458	2.8579
	(0.0120)	(0.0214)	(0.0123)	(0.0133)	(0.0123)	(0.0104)
σ_F	0.0055	0.0056	0.0023	0.0044	0.0015	0.0012
	(0.0014)	(0.0093)	(0.0015)	(0.0014)	(0.0007)	(0.0007)
μ_I	1.6267	1.6492	0.2906	0.7026	-0.8272	-0.7052
	(0.0107)	(0.0222)	(0.0491)	(0.0215)	(0.1035)	(0.0833)
σ_I	0.2555	0.3702	0.8894	0.8819	1.6085	1.6250
	(0.0235)	(0.0189)	(0.0484)	(0.0360)	(0.0765)	(0.0628)
μ_S	0.9628	0.6249	0.3672	-1.1564	1.1741	1.0537
	(0.1716)	(0.0316)	(0.2615)	(0.7305)	(0.0767)	(0.1031)
σ_S	0.5374	0.7032	0.8134	1.2797	0.7675	0.8914
	(0.0575)	(0.0279)	(0.0769)	(0.1621)	(0.0412)	(0.0511)
σ_{ME}	0.4533	0.4495	0.4626	0.4834	0.4778	0.4574
	(0.0066)	(0.0106)	(0.0057)	(0.0086)	(0.0060)	(0.0057)
γ	11.5653	4.4566	1.7096	0.6789	1.3640	0.9826
γ_{k5}	-	3.6063	-	0.5685	-	0.8184
γ_{k13}	-	4.7796	-	0.7131	-	1.0216
γ_{other}	-	5.3355	-	0.7786	-	1.0859
b	-16.2900	-12.0563	-14.1630	-10.3558	-22.8976	-21.3658
c	0.4717	0.4717	0.5350	0.5350	0.4710	0.4710
$LogLikelihood$	-21279	-11291	-13751	-9427	-13581	-17417
N	7534	7637	4587	4759	4318	6503

Note: Bootstrap standard errors (based on 100 replications) in parenthesis. Non estimated parameters: $\beta = 0.5$, $\tau = 0.48$ and $\rho = 0.062$.

Table E.3: Argentina - Labor Market Dynamics and States

	Primary			Secondary			Tertiary		
	M	W	W/M	M	W	W/M	M	W	W/M
h_u									
Data	-	-	-	-	-	-	-	-	-
Model	0.357	0.303	0.849	0.331	0.292	0.880	0.304	0.276	0.906
$h_{u \to e_F}$									
Model	0.129	0.127	0.984	0.215	0.183	0.850	0.208	0.201	0.965
$h_{u \to e_I}$									
Model	0.129	0.127	0.984	0.059	0.095	1.616	0.031	0.038	1.227
$h_{u \to e_S}$									
Model	0.099	0.049	0.497	0.058	0.014	0.249	0.065	0.037	0.565
u									
Data	0.053	0.041	0.767	0.041	0.046	1.111	0.032	0.039	1.195
Model	0.058	0.084	1.444	0.044	0.075	1.732	0.035	0.049	1.389
e_F									
Data	0.344	0.140	0.407	0.536	0.300	0.559	0.592	0.531	0.898
Model	0.321	0.360	1.119	0.563	0.481	0.854	0.640	0.668	1.043
e_I									
Data	0.235	0.207	0.881	0.145	0.150	1.032	0.087	0.098	1.136
Model	0.321	0.360	1.119	0.154	0.249	1.622	0.095	0.126	1.327
e_S									
Data	0.269	0.095	0.353	0.227	0.119	0.522	0.212	0.125	0.590
Model	0.299	0.196	0.657	0.239	0.194	0.811	0.229	0.157	0.686
np									
Data	0.098	0.517	5.282	0.050	0.386	7.732	0.078	0.207	2.664
Model	0.098	0.517	5.282	0.050	0.386	7.732	0.078	0.207	2.664

Table E.4: Argentina - Productivity and Wages

	Primary			Secondary			Tertiary		
	M	W	W/M	M	W	W/M	M	W	W/M
$E[x_F]$									
Model	13.004	10.994	0.845	12.601	11.927	0.947	17.218	17.425	1.012
$SD(x_F)$									
Model	0.072	0.061	0.847	0.035	0.052	1.502	0.026	0.016	0.614
$E[x_I]$									
Model	5.256	5.572	1.060	1.986	2.979	1.500	1.595	1.850	1.160
$SD[x_I]$									
Model	1.365	2.136	1.565	2.181	3.231	1.482	5.608	6.678	1.191
$E[x_S]$									
Model	3.026	2.392	0.791	2.009	0.714	0.355	4.351	4.267	0.981
$SD[x_S]$									
Model	1.751	1.913	1.093	1.946	1.453	0.747	3.889	4.702	1.209
Y_W									
Model	7.192	7.020	0.976	9.027	8.152	0.903	13.448	13.884	1.032
Y_C									
Model	6.109	3.106	0.508	8.203	4.628	0.564	11.968	10.477	0.875
$E[w\|e_F]$									
Data	4.492	3.783	0.842	5.095	4.662	0.915	6.728	6.642	0.987
Model	4.523	3.768	0.833	5.161	4.761	0.922	6.749	6.700	0.993
$SD[w\|e_F]$									
Data	2.140	1.749	0.817	2.361	2.189	0.927	3.354	3.035	0.905
Model	2.169	1.773	0.818	2.541	2.448	0.963	3.443	3.230	0.938
$E[w\|e_I]$									
Data	2.477	2.597	1.048	2.845	2.783	0.978	4.167	3.892	0.934
Model	2.504	2.641	1.055	2.853	2.779	0.974	4.843	4.364	0.901
$SD[w\|e_I]$									
Data	1.329	1.559	1.173	1.645	1.782	1.083	2.957	2.774	0.938
Model	1.420	1.741	1.227	2.430	2.344	0.964	14.675	6.817	0.465
$E[w\|e_S]$									
Data	2.997	2.365	0.789	3.520	3.156	0.897	5.207	5.228	1.004
Model	3.028	2.421	0.800	3.526	3.196	0.906	5.246	5.492	1.047
$SD[w\|e_S$									
Data	2.269	2.184	0.962	2.771	3.206	1.157	4.360	4.770	1.094
Model	2.477	2.334	0.942	3.053	3.515	1.151	4.979	6.201	1.245

Note: $E[x]$ is the average productivity, $SD(x)$ is the standard deviation of productivity, Y_W is the output per worker, Y_C is the output per capita, $E[w|e]$ is the average wage conditional on the employment status e, and finally $SD[w|e]$ is the standard deviation of wages conditioning in the employment status e.

Table E.5: Argentina - Policy Experiments

	Benchmark			Policy Exp. 1		Policy Exp. 2	
	M	W	W/M	W	W/M	W	W/M
Primary							
u	0.058	0.084	1.444	0.084	1.444	0.085	1.449
e_F	0.321	0.360	1.119	0.360	1.119	0.361	1.123
e_I	0.321	0.360	1.119	0.360	1.119	0.361	1.123
e_S	0.299	0.196	0.657	0.196	0.657	0.194	0.648
np	0.098	0.517	5.282	0.422	4.312	0.091	0.930
h_u	0.357	0.303	0.849	0.303	0.849	0.302	0.847
Y_W	7.192	7.020	0.976	7.020	0.976	7.749	1.077
Y_C	6.109	3.106	0.508	3.716	0.608	6.448	1.055
$E[w\|e_F]$	4.494	3.788	0.843	3.788	0.843	4.354	0.969
$E[w\|e_I]$	1.876	1.957	1.043	1.957	1.043	2.340	1.247
$E[w\|e_S]$	1.123	0.882	0.786	0.882	0.786	1.171	1.043
Res. W.	0.201	0.148	0.736	0.148	0.736	0.538	2.674
Secondary							
u	0.044	0.075	1.732	0.075	1.732	0.080	1.836
e_F	0.563	0.481	0.854	0.481	0.854	0.510	0.905
e_I	0.154	0.249	1.622	0.249	1.622	0.250	1.623
e_S	0.239	0.194	0.811	0.194	0.811	0.161	0.671
np	0.050	0.386	7.732	0.297	5.947	0.285	5.708
h_u	0.331	0.292	0.880	0.292	0.880	0.283	0.854
Y_W	9.027	8.152	0.903	8.152	0.903	9.371	1.038
Y_C	8.203	4.628	0.564	5.299	0.646	6.165	0.752
$E[w\|e_F]$	5.133	4.731	0.922	4.731	0.922	5.357	1.044
$E[w\|e_I]$	2.382	2.299	0.965	2.299	0.965	2.705	1.136
$E[w\|e_S]$	2.077	1.782	0.858	1.782	0.858	2.221	1.069
Res. W.	1.753	1.402	0.800	1.402	0.800	1.849	1.055
Tertiary							
u	0.035	0.049	1.389	0.049	1.389	0.051	1.440
e_F	0.640	0.668	1.043	0.668	1.043	0.692	1.081
e_I	0.095	0.126	1.327	0.126	1.327	0.112	1.176
e_S	0.229	0.157	0.686	0.157	0.686	0.146	0.635
np	0.078	0.207	2.664	0.150	1.931	0.101	1.301
h_u	0.304	0.276	0.906	0.276	0.906	0.266	0.875
Y_W	13.448	13.884	1.032	13.884	1.032	15.733	1.170
Y_C	11.968	10.477	0.875	11.228	0.938	13.430	1.122
$E[w\|e_F]$	6.753	6.571	0.973	6.571	0.973	7.513	1.112
$E[w\|e_I]$	3.533	3.358	0.950	3.358	0.950	4.234	1.198
$E[w\|e_S]$	2.729	2.624	0.961	2.624	0.961	3.307	1.212
Res. W.	1.873	1.604	0.857	1.604	0.857	2.334	1.246

Table E.5: Argentina - Policy Experiments – continued from previous page

	Policy Exp. 3		Policy Exp. 4 ($\tau = 0$)		Policy Exp. 4 ($c = 0$)	
	W	W/M	W	W/M	W	W/M
			Primary			
u	0.084	1.444	0.084	1.444	0.084	1.445
e_F	0.360	1.119	0.360	1.119	0.360	1.120
e_I	0.360	1.119	0.360	1.119	0.360	1.120
e_S	0.196	0.657	0.196	0.657	0.196	0.656
np	0.506	5.175	0.480	6.151	0.333	9.308
h_u	0.303	0.849	0.303	0.849	0.303	0.849
Y_W	7.020	0.976	7.020	0.976	7.023	0.976
Y_C	3.174	0.520	3.342	0.535	4.290	0.657
$E[w\|e_F]$	4.191	0.933	5.579	0.844	3.838	0.846
$E[w\|e_I]$	2.162	1.152	2.868	1.047	2.006	1.045
$E[w\|e_S]$	0.972	0.865	1.279	0.788	0.933	0.800
Res. W.	0.153	0.759	0.165	0.746	0.247	0.856
			Secondary			
u	0.077	1.765	0.081	1.713	0.073	1.741
e_F	0.490	0.870	0.514	0.844	0.463	0.858
e_I	0.247	1.605	0.238	1.745	0.293	1.479
e_S	0.186	0.778	0.167	0.804	0.172	0.777
np	0.369	7.392	0.325	11.572	0.362	7.962
h_u	0.288	0.869	0.279	0.901	0.311	0.891
Y_W	8.287	0.918	8.637	0.898	7.896	0.906
Y_C	4.827	0.588	5.357	0.601	4.671	0.586
$E[w\|e_F]$	5.198	1.013	6.790	0.925	4.778	0.926
$E[w\|e_I]$	2.535	1.064	3.350	0.960	2.162	0.983
$E[w\|e_S]$	1.973	0.950	2.634	0.871	1.882	0.885
Res. W.	1.468	0.838	1.654	0.792	1.496	0.828
			Tertiary			
u	0.049	1.401	0.050	1.383	0.048	1.380
e_F	0.674	1.052	0.689	1.038	0.651	1.035
e_I	0.122	1.291	0.114	1.333	0.150	1.321
e_S	0.155	0.674	0.147	0.685	0.152	0.679
np	0.190	2.446	0.149	3.166	0.200	2.667
h_u	0.273	0.898	0.267	0.910	0.284	0.913
Y_W	13.994	1.041	14.289	1.029	13.562	1.025
Y_C	10.779	0.901	11.549	0.906	10.336	0.875
$E[w\|e_F]$	7.237	1.072	9.508	0.977	6.589	0.974
$E[w\|e_I]$	3.752	1.062	5.137	0.957	3.025	0.943
$E[w\|e_S]$	2.902	1.063	3.863	0.976	2.655	0.965
Res. W.	1.691	0.903	1.939	0.866	1.639	0.863

Table E.6: Chile - Descriptive Statistics

Labor Market States	N	Prop.	$\bar{t}_u$	$\bar{w}$	σ_w	N	Prop.	$\bar{t}_u$	$\bar{w}$	σ_w
	Men					Women				
Education Group: Primary										
Unemployed	873	0.07	2.55	-	-	776	0.05	2.09	-	-
Formal Emp.	5807	0.46	-	2.68	1.11	2703	0.17	-	2.13	0.68
Informal Emp.	865	0.07	-	2.31	1.12	403	0.03	-	2.00	1.38
Self-Emp.	3073	0.25	-	2.63	2.02	1871	0.12	-	2.33	2.29
Non Part.	1882	0.15	-	-	-	10176	0.64	-	-	-
$K \leq 5$						3201	0.31			
$5 < K \leq 13$						2710	0.27			
Education Group: Secondary										
Unemployed	1002	0.07	2.89	-	-	980	0.05	2.67	-	-
Formal Emp.	9995	0.65	-	3.26	1.58	7052	0.39	-	2.57	1.04
Informal Emp.	715	0.05	-	2.80	1.71	531	0.03	-	2.37	1.56
Self-Emp.	2717	0.18	-	3.46	3.11	2203	0.12	-	2.84	2.76
Non Part.	892	0.06	-	-	-	7504	0.41	-	-	-
$K \leq 5$						3067	0.41			
$5 < K \leq 13$						2071	0.28			
Education Group: Tertiary										
Unemployed	778	0.06	3.35	-	-	802	0.05	2.93	-	-
Formal Emp.	8510	0.66	-	7.31	5.92	9246	0.60	-	5.50	3.73
Informal Emp.	446	0.03	-	5.73	5.46	497	0.03	-	4.98	3.79
Self-Emp.	1966	0.15	-	8.09	9.04	1442	0.09	-	6.20	6.67
Non Part.	1278	0.10	-	-	-	3401	0.22	-	-	-
$K \leq 5$						1314	0.39			
$5 < K \leq 13$						769	0.23			

Note: Wage distributions are trimmed at the top and bottom 1 percentile by gender, education group and type of job, and are reported in US Dollars of December 2016 (Exchange Rate = 667.17 Chilean Pesos/US). A worker is categorized as informal if he/she reports not having benefits of social security. K means proportion of women with the presence of kids in the household with respect to non participating women.

Table E.7: Chile - Estimated Parameters

	Primary		Secondary		Tertiary	
	Men	Women	Men	Women	Men	Women
ρU	1.1619	0.1351	1.6532	0.9071	3.2588	2.1330
	(0.0422)	(0.0091)	(0.0347)	(0.0486)	(0.5782)	(0.0835)
λ_F	0.2184	0.1394	0.2759	0.2430	0.2085	0.2460
	(0.0137)	(0.0105)	(0.0205)	(0.0234)	(0.0172)	(0.0167)
λ_S	0.2083	0.2016	0.4518	0.2619	0.1850	0.1993
	(0.0176)	(0.0099)	(0.1680)	(0.0156)	(0.0362)	(0.0209)
δ_F	0.0330	0.0697	0.0277	0.0349	0.0191	0.0213
	(0.0021)	(0.0052)	(0.0021)	(0.0039)	(0.0016)	(0.0014)
δ_S	0.0398	0.0836	0.0186	0.0449	0.0313	0.0454
	(0.0020)	(0.0041)	(0.0053)	(0.0033)	(0.0043)	(0.0039)
μ_F	1.6253	1.5930	1.7619	1.6358	2.5841	2.3593
	(0.0119)	(0.0071)	(0.0092)	(0.0105)	(0.0720)	(0.0127)
σ_F	0.0029	0.0829	0.0050	0.0042	0.1405	0.0109
	(0.0014)	(0.0071)	(0.0035)	(0.0011)	(0.2957)	(0.0027)
μ_I	-1.0825	1.3222	-1.2456	-1.6818	-1.1494	-2.3260
	(0.0936)	(0.0214)	(0.1011)	(0.4031)	(0.7627)	(0.1911)
σ_I	1.4107	0.4296	1.3244	1.5077	1.5277	2.0542
	(0.0661)	(0.0308)	(0.0612)	(0.2120)	(0.3560)	(0.1038)
μ_S	0.4615	0.5272	-0.9611	-0.4041	1.0008	0.4947
	(0.0866)	(0.0194)	(0.5700)	(0.1616)	(0.2676)	(0.2191)
σ_S	0.7044	0.8061	1.2033	1.2337	0.9903	1.1606
	(0.0326)	(0.0174)	(0.1232)	(0.0861)	(0.1027)	(0.0751)
σ_{ME}	0.3943	0.2839	0.4271	0.3714	0.6751	0.5976
	(0.0045)	(0.0055)	(0.0030)	(0.0049)	(0.1280)	(0.0037)
γ	1.6295	3.3172	1.7200	0.9809	0.7113	0.7077
γ_{k5}	-	3.0759	-	0.8302	-	0.6117
γ_{k13}	-	3.5540	-	1.0149	-	0.7252
γ_{other}	-	3.3424	-	1.1237	-	0.7782
b	-5.2218	-7.1410	-5.2652	-6.1237	-12.5334	-12.7475
c	0.2809	0.2809	0.3425	0.3425	0.5119	0.5119
$LogLikelihood$	-28044	-15330	-38209	-26514	-42153	-38439
N	12500	15929	15321	18270	12978	15388

Note: Bootstrap standard errors (based on 100 replications) in parenthesis. Non estimated parameters: $\beta = 0.5$, $\tau = 0.20$ and $\rho = 0.067$.

Table E.8: Chile - Labor Market Dynamics and States

	Primary			Secondary			Tertiary		
	M	W	W/M	M	W	W/M	M	W	W/M
h_u									
Data	0.391	0.479	1.225	0.346	0.375	1.082	0.299	0.341	1.142
Model	0.392	0.480	1.226	0.346	0.373	1.078	0.299	0.341	1.142
$h_{u \to e_F}$									
Model	0.218	0.139	0.638	0.275	0.243	0.882	0.209	0.246	1.179
$h_{u \to e_I}$									
Model	0.033	0.139	4.192	0.020	0.025	1.268	0.011	0.013	1.216
$h_{u \to e_S}$									
Model	0.140	0.201	1.438	0.051	0.105	2.061	0.079	0.082	1.036
u									
Data	0.070	0.049	0.698	0.065	0.054	0.820	0.060	0.052	0.869
Model	0.082	0.135	1.640	0.069	0.091	1.306	0.066	0.067	1.006
e_F									
Data	0.465	0.170	0.365	0.652	0.386	0.592	0.656	0.601	0.916
Model	0.545	0.270	0.495	0.692	0.632	0.912	0.727	0.771	1.060
e_I									
Data	0.069	0.025	0.366	0.047	0.029	0.623	0.034	0.032	0.940
Model	0.083	0.270	3.252	0.050	0.065	1.310	0.038	0.042	1.094
e_S									
Data	0.246	0.117	0.478	0.177	0.121	0.680	0.151	0.094	0.619
Model	0.289	0.325	1.124	0.188	0.213	1.129	0.168	0.120	0.716
np									
Data	0.151	0.639	4.243	0.058	0.411	7.055	0.098	0.221	2.244
Model	0.151	0.639	4.243	0.058	0.411	7.055	0.098	0.221	2.244

Table E.9: Chile - Productivity and Wages

	Primary			Secondary			Tertiary		
	M	W	W/M	M	W	W/M	M	W	W/M
$E[x_F]$									
Model	5.081	4.936	0.971	5.824	5.134	0.881	13.383	10.585	0.791
$SD(x_F)$									
Model	0.015	0.410	27.215	0.030	0.021	0.719	1.887	0.115	0.061
$E[x_I]$									
Model	0.916	4.115	4.492	0.692	0.580	0.838	1.018	0.806	0.792
$SD[x_I]$									
Model	2.304	1.852	0.804	1.512	1.711	1.131	3.111	6.593	2.120
$E[x_S]$									
Model	2.039	2.345	1.150	0.797	1.429	1.793	4.441	3.217	0.724
$SD[x_S]$									
Model	1.629	2.243	1.377	1.431	2.706	1.891	5.735	5.426	0.946
Y_W									
Model	4.209	3.706	0.881	5.269	4.502	0.854	12.272	10.028	0.817
Y_C									
Model	3.281	1.158	0.353	4.618	2.412	0.522	10.328	7.289	0.706
$E[w\|e_F]$									
Data	2.676	2.126	0.794	3.262	2.566	0.787	7.312	5.501	0.752
Model	2.714	2.142	0.789	3.269	2.603	0.796	7.229	5.493	0.760
$SD[w\|e_F]$									
Data	1.107	0.679	0.613	1.577	1.039	0.659	5.921	3.730	0.630
Model	1.118	0.663	0.593	1.475	1.003	0.680	5.664	3.596	0.635
$E[w\|e_I]$									
Data	2.315	2.004	0.866	2.798	2.372	0.848	5.730	4.983	0.870
Model	2.419	1.969	0.814	2.824	1.956	0.692	5.797	5.426	0.936
$SD[w\|e_I]$									
Data	1.122	1.381	1.232	1.707	1.560	0.914	5.458	3.787	0.694
Model	2.737	1.088	0.398	2.545	2.179	0.856	6.522	8.316	1.275
$E[w\|e_S]$									
Data	2.632	2.328	0.885	3.457	2.842	0.822	8.091	6.199	0.766
Model	2.655	2.363	0.890	3.420	2.916	0.853	8.127	6.534	0.804
$SD[w\|e_S$									
Data	2.020	2.289	1.133	3.110	2.764	0.889	9.040	6.670	0.738
Model	2.143	2.491	1.163	3.307	3.827	1.157	10.397	9.254	0.890

Note: $E[x]$ is the average productivity, $SD(x)$ is the standard deviation of productivity, Y_W is the output per worker, Y_C is the output per capita, $E[w|e]$ is the average wage conditional on the employment status e, and finally $SD[w|e]$ is the standard deviation of wages conditioning in the employment status e.

Table E.10: Chile - Policy Experiments

	Benchmark			Policy Exp. 1		Policy Exp. 2	
	M	W	W/M	W	W/M	W	W/M
				Primary			
u	0.082	0.135	1.640	0.135	1.640	0.135	1.646
e_F	0.545	0.270	0.495	0.270	0.495	0.271	0.497
e_I	0.083	0.270	3.252	0.270	3.252	0.271	3.264
e_S	0.289	0.325	1.124	0.325	1.124	0.323	1.115
np	0.151	0.639	4.243	0.571	3.789	0.303	2.016
h_u	0.392	0.480	1.226	0.480	1.226	0.478	1.220
Y_W	4.209	3.706	0.881	3.706	0.881	4.093	0.973
Y_C	3.281	1.158	0.353	1.377	0.420	2.465	0.751
$E[w\|e_F]$	2.698	2.124	0.787	2.124	0.787	2.442	0.905
$E[w\|e_I]$	2.209	1.782	0.807	1.782	0.807	2.066	0.935
$E[w\|e_S]$	1.685	1.045	0.620	1.045	0.620	1.266	0.752
Res. W.	1.161	0.135	0.116	0.135	0.116	0.359	0.310
				Secondary			
u	0.069	0.091	1.306	0.091	1.306	0.093	1.333
e_F	0.692	0.632	0.912	0.632	0.912	0.645	0.931
e_I	0.050	0.065	1.310	0.065	1.310	0.061	1.225
e_S	0.188	0.213	1.129	0.213	1.129	0.202	1.070
np	0.058	0.411	7.055	0.322	5.529	0.322	5.535
h_u	0.346	0.373	1.078	0.373	1.078	0.364	1.050
Y_W	5.269	4.502	0.854	4.502	0.854	5.015	0.952
Y_C	4.618	2.412	0.522	2.776	0.601	3.085	0.668
$E[w\|e_F]$	3.253	2.593	0.797	2.593	0.797	2.930	0.901
$E[w\|e_I]$	2.612	1.849	0.708	1.849	0.708	2.136	0.818
$E[w\|e_S]$	2.280	1.695	0.743	1.695	0.743	1.986	0.871
Res. W.	1.653	0.907	0.549	0.907	0.549	1.154	0.698
				Tertiary			
u	0.066	0.067	1.006	0.067	1.006	0.068	1.017
e_F	0.727	0.771	1.060	0.771	1.060	0.780	1.072
e_I	0.038	0.042	1.094	0.042	1.094	0.039	1.026
e_S	0.168	0.120	0.716	0.120	0.716	0.114	0.676
np	0.098	0.221	2.244	0.159	1.612	0.150	1.523
h_u	0.299	0.341	1.142	0.341	1.142	0.335	1.121
Y_W	12.272	10.028	0.817	10.028	0.817	11.099	0.904
Y_C	10.328	7.289	0.706	7.871	0.762	8.797	0.852
$E[w\|e_F]$	7.206	5.477	0.760	5.477	0.760	6.192	0.859
$E[w\|e_I]$	5.370	5.238	0.975	5.238	0.975	6.061	1.129
$E[w\|e_S]$	5.051	3.760	0.744	3.760	0.744	4.377	0.867
Res. W.	3.259	2.133	0.655	2.133	0.655	2.681	0.823

Table E.10: Chile - Policy Experiments – continued from previous page

	Policy Exp. 3		Policy Exp. 4 ($\tau = 0$)		Policy Exp. 4 ($c = 0$)	
	W	W/M	W	W/M	W	W/M
			Primary			
u	0.135	1.640	0.135	1.611	0.135	1.670
e_F	0.270	0.495	0.270	0.486	0.270	0.504
e_I	0.270	3.252	0.270	3.371	0.270	2.667
e_S	0.325	1.124	0.325	1.158	0.325	1.153
np	0.636	4.226	0.630	4.705	0.557	3.782
h_u	0.480	1.226	0.480	1.252	0.480	1.204
Y_W	3.706	0.881	3.706	0.869	3.708	0.893
Y_C	1.166	0.355	1.186	0.350	1.421	0.437
$E[w\|e_F]$	2.233	0.828	2.537	0.804	2.145	0.793
$E[w\|e_I]$	1.873	0.848	2.127	0.806	1.803	0.899
$E[w\|e_S]$	1.097	0.651	1.243	0.629	1.067	0.629
Res. W.	0.136	0.117	0.139	0.113	0.177	0.150
			Secondary			
u	0.091	1.312	0.092	1.289	0.089	1.292
e_F	0.634	0.916	0.642	0.901	0.617	0.902
e_I	0.064	1.293	0.062	1.358	0.089	1.407
e_S	0.210	1.116	0.204	1.198	0.205	1.114
np	0.403	6.924	0.385	8.618	0.405	7.099
h_u	0.371	1.072	0.366	1.083	0.382	1.087
Y_W	4.520	0.858	4.563	0.849	4.417	0.845
Y_C	2.452	0.531	2.549	0.535	2.395	0.522
$E[w\|e_F]$	2.715	0.834	3.054	0.800	2.600	0.798
$E[w\|e_I]$	1.948	0.746	2.227	0.701	1.586	0.664
$E[w\|e_S]$	1.785	0.783	2.036	0.736	1.713	0.747
Res. W.	0.926	0.560	0.974	0.539	0.921	0.553
			Tertiary			
u	0.067	1.009	0.068	1.002	0.066	1.004
e_F	0.774	1.063	0.780	1.056	0.764	1.059
e_I	0.041	1.076	0.039	1.113	0.051	1.109
e_S	0.118	0.704	0.113	0.715	0.119	0.714
np	0.213	2.162	0.194	2.363	0.219	2.242
h_u	0.339	1.136	0.335	1.146	0.344	1.143
Y_W	10.054	0.819	10.119	0.815	9.956	0.816
Y_C	7.383	0.715	7.608	0.716	7.260	0.706
$E[w\|e_F]$	5.735	0.796	6.453	0.764	5.483	0.760
$E[w\|e_I]$	5.547	1.033	6.422	0.990	4.661	0.935
$E[w\|e_S]$	3.966	0.785	4.547	0.752	3.775	0.746
Res. W.	2.186	0.671	2.320	0.660	2.145	0.656

Table E.11: Colombia - Descriptive Statistics

Labor Market	N	Prop.	$\bar{t}_u$	$\bar{w}$	σ_w	N	Prop.	$\bar{t}_u$	$\bar{w}$	σ_w
States	Men					Women				
Education Group: Primary										
Unemployed	607	0.06	3.14	-	-	828	0.07	4.56	-	-
Formal Emp.	1784	0.18	-	1.31	0.41	669	0.06	-	1.17	0.23
Informal Emp.	1311	0.13	-	1.08	0.39	935	0.08	-	0.87	0.36
Self-Emp.	5487	0.55	-	1.12	0.66	4199	0.35	-	0.80	0.57
Non Part.	758	0.08	-	-	-	5429	0.45	-	-	-
$K \leq 5$						1870	0.34			
$5 < K \leq 13$						1552	0.29			
Education Group: Secondary										
Unemployed	577	0.06	4.05	-	-	984	0.09	5.22	-	-
Formal Emp.	3656	0.41	-	1.45	0.54	2246	0.21	-	1.31	0.38
Informal Emp.	819	0.09	-	1.13	0.41	932	0.09	-	0.98	0.35
Self-Emp.	3496	0.39	-	1.40	0.91	3084	0.29	-	1.07	0.84
Non Part.	408	0.05	-	-	-	3335	0.32	-	-	-
$K \leq 5$						1272	0.38			
$5 < K \leq 13$						970	0.29			
Education Group: Tertiary										
Unemployed	840	0.09	5.33	-	-	1611	0.12	6.02	-	-
Formal Emp.	4551	0.50	-	3.06	2.24	5885	0.44	-	2.77	1.94
Informal Emp.	422	0.05	-	1.41	0.79	562	0.04	-	1.28	0.68
Self-Emp.	2775	0.30	-	2.99	2.73	3027	0.23	-	2.60	2.34
Non Part.	583	0.06	-	-	-	2167	0.16	-	-	-
$K \leq 5$						893	0.41			
$5 < K \leq 13$						516	0.24			

Note: Wage distributions are trimmed at the top and bottom 1 percentile by gender, education group and type of job, and are reported in US Dollars of December 2016 (Exchange Rate = 3009.86 Colombian Pesos/US). A worker is categorized as informal if he/she reports not having benefits of social security. K means proportion of women with the presence of kids in the household with respect to non participating women.

Table E.12: Colombia - Estimated Parameters

	Primary		Secondary		Tertiary	
	Men	Women	Men	Women	Men	Women
ρU	0.0950	0.0216	0.7977	0.3285	0.9019	0.8454
	(0.0042)	(0.1220)	(0.0139)	(0.0414)	(0.0210)	(0.0253)
λ_F	0.0746	0.0379	0.1443	0.0757	0.0997	0.0875
	(0.0016)	(0.0146)	(0.0111)	(0.0059)	(0.0035)	(0.0028)
λ_S	0.1727	0.1439	0.4299	0.2744	0.1105	0.0833
	(0.0040)	(0.0335)	(0.1365)	(0.0323)	(0.0063)	(0.0022)
δ_F	0.0291	0.0392	0.0228	0.0457	0.0183	0.0240
	(0.0001)	(0.0156)	(0.0018)	(0.0041)	(0.0006)	(0.0008)
δ_S	0.0190	0.0284	0.0116	0.0158	0.0240	0.0374
	(0.0001)	(0.0066)	(0.0018)	(0.0018)	(0.0010)	(0.0007)
μ_F	1.1613	1.1684	1.0158	1.1223	1.7155	1.8122
	(0.0072)	(0.0277)	(0.0094)	(0.0091)	(0.0223)	(0.0118)
σ_F	0.2402	0.0045	0.0019	0.0006	0.6252	0.0167
	(0.0084)	(0.0070)	(0.0010)	(0.0007)	(0.0280)	(0.0053)
μ_I	0.7369	0.5950	-0.5970	0.5507	-1.3506	-1.3141
	(0.0109)	(0.0449)	(0.0383)	(0.0338)	(0.1015)	(0.1093)
σ_I	0.3455	0.0082	0.7279	0.2083	1.1012	1.0514
	(0.0107)	(0.0586)	(0.0356)	(0.1033)	(0.0676)	(0.0689)
μ_S	-0.0266	-0.3949	-1.1005	-2.5203	0.4301	0.5815
	(0.0083)	(0.0295)	(0.3177)	(0.2928)	(0.0717)	(0.0331)
σ_S	0.5487	0.6566	0.8905	1.6580	0.9237	0.7444
	(0.0055)	(0.0833)	(0.0752)	(0.1859)	(0.0357)	(0.0215)
σ_{ME}	0.1521	0.3836	0.3441	0.3379	0.4046	0.6196
	(0.0064)	(0.0733)	(0.0043)	(0.0318)	(0.0185)	(0.0042)
γ	27.1017	36.9147	3.8723	3.5150	3.0554	2.1419
γ_{k5}	-	34.5620	-	3.1040	-	1.8271
γ_{k13}	-	39.1202	-	3.8469	-	2.2715
γ_{other}	-	37.3036	-	3.6484	-	2.3540
b	-4.7300	-2.4144	-1.6264	-1.9234	-5.2874	-3.4863
c	0.1371	0.1371	0.1520	0.1520	0.2139	0.2139
$LogLikelihood$	-17037	-12564	-17264	-16544	-25763	-33577
N	9947	12060	8956	10581	9171	13252

Note: Bootstrap standard errors (based on 100 replications) in parenthesis. Non estimated parameters: $\beta = 0.5$, $\tau = 0.31$ and $\rho = 0.053$.

Table E.13: Colombia - Labor Market Dynamics and States

	Primary			Secondary			Tertiary		
	M	W	W/M	M	W	W/M	M	W	W/M
h_u									
Data	0.318	0.219	0.690	0.247	0.192	0.776	0.188	0.166	0.886
Model	0.322	0.220	0.683	0.247	0.206	0.834	0.188	0.166	0.886
$h_{u \to e_F}$									
Model	0.075	0.039	0.519	0.144	0.076	0.527	0.099	0.087	0.884
$h_{u \to e_I}$									
Model	0.075	0.039	0.519	0.033	0.076	2.325	0.009	0.008	0.912
$h_{u \to e_S}$									
Model	0.173	0.142	0.824	0.071	0.054	0.769	0.079	0.070	0.885
u									
Data	0.061	0.069	1.125	0.064	0.093	1.443	0.092	0.122	1.327
Model	0.066	0.125	1.899	0.068	0.129	1.913	0.098	0.145	1.486
e_F									
Data	0.179	0.055	0.309	0.408	0.212	0.520	0.496	0.444	0.895
Model	0.168	0.121	0.718	0.427	0.214	0.501	0.530	0.531	1.002
e_I									
Data	0.132	0.078	0.588	0.091	0.088	0.963	0.046	0.042	0.922
Model	0.168	0.121	0.718	0.097	0.214	2.210	0.049	0.051	1.034
e_S									
Data	0.552	0.348	0.631	0.390	0.291	0.747	0.303	0.228	0.755
Model	0.597	0.633	1.060	0.409	0.443	1.083	0.323	0.273	0.845
np									
Data	0.076	0.450	5.907	0.046	0.315	6.919	0.064	0.164	2.572
Model	0.076	0.450	5.907	0.046	0.315	6.919	0.064	0.164	2.572

Table E.14: Colombia - Productivity and Wages

	Primary			Secondary			Tertiary		
	M	W	W/M	M	W	W/M	M	W	W/M
$E[x_F]$									
Model	3.288	3.208	0.976	2.762	3.072	1.112	6.759	6.125	0.906
$SD(x_F)$									
Model	0.801	0.012	0.015	0.005	0.002	0.367	4.674	0.103	0.022
$E[x_I]$									
Model	2.218	1.807	0.815	0.717	1.773	2.473	0.475	0.467	0.983
$SD[x_I]$									
Model	0.790	0.014	0.017	0.601	0.373	0.621	0.730	0.664	0.910
$E[x_S]$									
Model	1.132	0.836	0.738	0.503	0.318	0.633	2.355	2.360	1.002
$SD[x_S]$									
Model	0.671	0.613	0.914	0.548	1.216	2.218	2.734	2.030	0.743
Y_W									
Model	1.716	1.298	0.756	2.042	1.821	0.892	5.204	4.778	0.918
Y_C									
Model	1.481	0.624	0.422	1.817	1.086	0.597	4.396	3.416	0.777
$E[w\|e_F]$									
Data	1.306	1.169	0.895	1.448	1.305	0.902	3.055	2.775	0.908
Model	1.300	1.243	0.956	1.458	1.347	0.924	3.049	2.767	0.907
$SD[w\|e_F]$									
Data	0.411	0.228	0.554	0.544	0.378	0.695	2.245	1.941	0.865
Model	0.371	0.481	1.294	0.519	0.471	0.908	2.337	1.899	0.812
$E[w\|e_I]$									
Data	1.082	0.870	0.804	1.127	0.976	0.866	1.411	1.282	0.908
Model	1.093	0.852	0.780	1.101	0.980	0.891	1.392	1.243	0.893
$SD[w\|e_I]$									
Data	0.386	0.359	0.928	0.407	0.352	0.866	0.793	0.683	0.861
Model	0.434	0.345	0.795	0.534	0.388	0.726	1.134	0.967	0.853
$E[w\|e_S]$									
Data	1.122	0.805	0.717	1.398	1.067	0.763	2.985	2.599	0.871
Model	1.130	0.849	0.751	1.398	1.235	0.884	3.055	2.699	0.883
$SD[w\|e_S$									
Data	0.658	0.572	0.870	0.912	0.845	0.926	2.734	2.338	0.855
Model	0.703	0.752	1.070	0.971	1.959	2.017	3.640	2.934	0.806

Note: $E[x]$ is the average productivity, $SD(x)$ is the standard deviation of productivity, Y_W is the output per worker, Y_C is the output per capita, $E[w|e]$ is the average wage conditional on the employment status e, and finally $SD[w|e]$ is the standard deviation of wages conditioning in the employment status e.

Table E.15: Colombia - Policy Experiments

	Benchmark			Policy Exp. 1		Policy Exp. 2	
	M	W	W/M	W	W/M	W	W/M
Primary							
u	0.066	0.125	1.899	0.125	1.899	0.125	1.900
e_F	0.168	0.121	0.718	0.121	0.718	0.121	0.718
e_I	0.168	0.121	0.718	0.121	0.718	0.121	0.718
e_S	0.597	0.633	1.060	0.633	1.060	0.633	1.060
np	0.076	0.450	5.907	0.369	4.836	0.046	0.599
h_u	0.322	0.220	0.683	0.220	0.683	0.220	0.682
Y_W	1.716	1.298	0.756	1.298	0.756	1.428	0.832
Y_C	1.481	0.624	0.422	0.717	0.484	1.193	0.805
$E[w\|e_F]$	1.302	1.238	0.951	1.238	0.951	1.399	1.074
$E[w\|e_I]$	0.894	0.703	0.787	0.703	0.787	0.811	0.907
$E[w\|e_S]$	0.480	0.332	0.693	0.332	0.693	0.403	0.841
Res. W.	0.095	0.027	0.285	0.027	0.285	0.105	1.102
Secondary							
u	0.068	0.129	1.913	0.129	1.913	0.136	2.009
e_F	0.427	0.214	0.501	0.214	0.501	0.225	0.526
e_I	0.097	0.214	2.210	0.214	2.210	0.225	2.320
e_S	0.409	0.443	1.083	0.443	1.083	0.415	1.016
np	0.046	0.315	6.919	0.238	5.232	0.222	4.878
h_u	0.247	0.206	0.834	0.206	0.834	0.200	0.810
Y_W	2.042	1.821	0.892	1.821	0.892	2.080	1.019
Y_C	1.817	1.086	0.597	1.208	0.665	1.398	0.769
$E[w\|e_F]$	1.453	1.313	0.904	1.313	0.904	1.478	1.017
$E[w\|e_I]$	0.992	0.841	0.847	0.841	0.847	0.958	0.966
$E[w\|e_S]$	0.935	0.649	0.694	0.649	0.694	0.781	0.835
Res. W.	0.797	0.329	0.412	0.329	0.412	0.428	0.537
Tertiary							
u	0.098	0.145	1.486	0.145	1.486	0.147	1.506
e_F	0.530	0.531	1.002	0.531	1.002	0.538	1.015
e_I	0.049	0.051	1.034	0.051	1.034	0.043	0.863
e_S	0.323	0.273	0.845	0.273	0.845	0.272	0.843
np	0.064	0.164	2.572	0.111	1.738	0.103	1.628
h_u	0.188	0.166	0.886	0.166	0.886	0.164	0.872
Y_W	5.204	4.778	0.918	4.778	0.918	5.308	1.020
Y_C	4.396	3.416	0.777	3.632	0.826	4.058	0.923
$E[w\|e_F]$	3.046	2.760	0.906	2.760	0.906	3.101	1.018
$E[w\|e_I]$	1.257	1.166	0.927	1.166	0.927	1.365	1.086
$E[w\|e_S]$	1.619	1.448	0.894	1.448	0.894	1.664	1.028
Res. W.	0.902	0.845	0.937	0.845	0.937	1.059	1.174

Table E.15: Colombia - Policy Experiments – continued from previous page

	Policy Exp. 3		Policy Exp. 4 ($\tau = 0$)		Policy Exp. 4 ($c = 0$)	
	W	W/M	W	W/M	W	W/M
			Primary			
u	0.125	1.899	0.125	1.899	0.125	1.899
e_F	0.121	0.718	0.121	0.718	0.121	0.718
e_I	0.121	0.718	0.121	0.718	0.121	0.718
e_S	0.633	1.060	0.633	1.060	0.633	1.060
np	0.448	5.879	0.443	6.322	0.347	6.656
h_u	0.220	0.683	0.220	0.683	0.220	0.683
Y_W	1.298	0.756	1.298	0.756	1.298	0.756
Y_C	0.627	0.423	0.633	0.424	0.742	0.488
$E[w\|e_F]$	1.332	1.023	1.618	0.956	1.242	0.949
$E[w\|e_I]$	0.756	0.846	0.917	0.792	0.708	0.785
$E[w\|e_S]$	0.357	0.744	0.432	0.702	0.337	0.692
Res. W.	0.027	0.286	0.028	0.281	0.036	0.329
			Secondary			
u	0.130	1.923	0.131	1.808	0.132	2.003
e_F	0.215	0.504	0.218	0.474	0.219	0.525
e_I	0.215	2.221	0.218	2.404	0.219	1.756
e_S	0.440	1.077	0.434	1.149	0.429	1.097
np	0.310	6.811	0.299	8.735	0.292	6.672
h_u	0.205	0.831	0.204	0.876	0.203	0.794
Y_W	1.829	0.896	1.850	0.865	1.863	0.924
Y_C	1.098	0.604	1.127	0.588	1.145	0.636
$E[w\|e_F]$	1.403	0.966	1.677	0.923	1.324	0.908
$E[w\|e_I]$	0.895	0.902	1.058	0.842	0.852	0.911
$E[w\|e_S]$	0.693	0.741	0.825	0.696	0.680	0.719
Res. W.	0.333	0.418	0.344	0.395	0.350	0.434
			Tertiary			
u	0.146	1.491	0.147	1.482	0.142	1.481
e_F	0.533	1.005	0.537	0.995	0.520	0.999
e_I	0.050	1.009	0.047	1.034	0.071	1.062
e_S	0.272	0.841	0.269	0.852	0.267	0.843
np	0.157	2.476	0.143	2.790	0.162	2.585
h_u	0.165	0.881	0.163	0.888	0.169	0.888
Y_W	4.794	0.921	4.835	0.916	4.687	0.915
Y_C	3.450	0.785	3.533	0.784	3.370	0.776
$E[w\|e_F]$	2.948	0.968	3.516	0.908	2.763	0.906
$E[w\|e_I]$	1.241	0.987	1.471	0.923	1.055	0.923
$E[w\|e_S]$	1.542	0.952	1.824	0.886	1.453	0.894
Res. W.	0.863	0.957	0.908	0.934	0.851	0.938

Table E.16: Mexico - Descriptive Statistics

Labor Market States	N	Prop.	$\bar{t}_u$	$\bar{w}$	σ_w	N	Prop.	$\bar{t}_u$	$\bar{w}$	σ_w
	Men					Women				
Education Group: Primary										
Unemployed	328	0.03	1.24	-	-	182	0.01	1.50	-	-
Formal Emp.	2412	0.24	-	1.42	0.59	1063	0.07	-	1.14	0.44
Informal Emp.	3480	0.35	-	1.22	0.52	1177	0.08	-	1.04	0.63
Self-Emp.	2415	0.24	-	1.67	1.14	2248	0.15	-	1.18	1.04
Non Part.	1413	0.14	-	-	-	10430	0.69	-	-	-
$K \leq 5$						3727	0.36			
$5 < K \leq 13$						2902	0.28			
Education Group: Secondary										
Unemployed	1076	0.04	1.95	-	-	713	0.02	1.87	-	-
Formal Emp.	11929	0.46	-	1.59	0.75	6235	0.19	-	1.39	0.69
Informal Emp.	6401	0.25	-	1.29	0.66	2991	0.09	-	1.15	0.67
Self-Emp.	4770	0.18	-	1.99	1.58	4001	0.12	-	1.67	1.63
Non Part.	1832	0.07	-	-	-	18215	0.57	-	-	-
$K \leq 5$						7809	0.43			
$5 < K \leq 13$						5532	0.30			
Education Group: Tertiary										
Unemployed	782	0.06	2.73	-	-	647	0.04	2.61	-	-
Formal Emp.	7078	0.57	-	3.02	1.85	7227	0.42	-	2.86	1.63
Informal Emp.	1389	0.11	-	2.09	1.57	1380	0.08	-	2.02	1.48
Self-Emp.	1897	0.15	-	3.17	2.90	1474	0.09	-	2.64	2.62
Non Part.	1239	0.10	-	-	-	6358	0.37	-	-	-
$K \leq 5$						2115	0.33			
$5 < K \leq 13$						1545	0.24			

Note: Wage distributions are trimmed at the top and bottom 1 percentile by gender, education group and type of job, and are reported in US Dollars of December 2016 (Exchange Rate = 20.52 Mexican Pesos/US). A worker is categorized as informal if he/she reports not having access to health care. K means proportion of women with the presence of kids in the household with respect to non participating women.

Table E.17: Mexico - Estimated Parameters

	Primary		Secondary		Tertiary	
	Men	Women	Men	Women	Men	Women
ρU	0.0769	0.0866	0.9945	0.6806	1.4058	1.1647
	(0.0316)	(0.0068)	(0.0149)	(0.0092)	(0.0572)	(0.0267)
λ_F	0.2605	0.1790	0.2613	0.2914	0.2164	0.2748
	(0.0162)	(0.0161)	(0.0128)	(0.0177)	(0.0116)	(0.0172)
λ_S	0.2825	0.3073	0.3035	0.5869	0.1752	0.4198
	(0.0120)	(0.0233)	(0.0415)	(0.0893)	(0.0160)	(0.1428)
δ_F	0.0290	0.0291	0.0236	0.0336	0.0239	0.0246
	(0.0018)	(0.0026)	(0.0012)	(0.0020)	(0.0013)	(0.0015)
δ_S	0.0384	0.0248	0.0248	0.0179	0.0443	0.0243
	(0.0016)	(0.0019)	(0.0023)	(0.0026)	(0.0035)	(0.0066)
μ_F	1.2965	1.0563	1.0639	1.0282	1.8190	1.8075
	(0.0200)	(0.0286)	(0.0087)	(0.0086)	(0.0122)	(0.0092)
σ_F	0.1133	0.1178	0.0036	0.0190	0.0138	0.0228
	(0.1065)	(0.1153)	(0.0013)	(0.0041)	(0.1028)	(0.0093)
μ_I	0.9051	0.6911	0.1909	-0.1791	-0.3006	-0.6903
	(0.0149)	(0.0275)	(0.0109)	(0.0102)	(0.0930)	(0.0502)
σ_I	0.1614	0.3504	0.4402	0.7646	0.9142	1.1595
	(0.0824)	(0.0569)	(0.0229)	(0.0185)	(0.0698)	(0.0409)
μ_S	0.3910	-0.1133	-0.3025	-1.6260	0.5568	-1.2779
	(0.0286)	(0.0350)	(0.1823)	(0.2935)	(0.1360)	(0.5792)
σ_S	0.5207	0.7612	0.8393	1.3077	0.7454	1.2796
	(0.0463)	(0.0386)	(0.0541)	(0.0748)	(0.0620)	(0.1307)
σ_{ME}	0.3720	0.3206	0.4321	0.4432	0.5736	0.5552
	(0.1398)	(0.1523)	(0.0025)	(0.0040)	(0.0151)	(0.0042)
γ	25.5112	4.2740	2.6677	0.8351	1.6376	0.8487
γ_{k5}	-	3.7243	-	0.6902	-	0.7739
γ_{k13}	-	4.6410	-	0.8890	-	0.8623
γ_{other}	-	4.5131	-	0.9857	-	0.8958
b	-13.7364	-9.0289	-3.4647	-4.5475	-6.6889	-8.2235
c	0.1495	0.1495	0.1669	0.1669	0.2116	0.2116
$Likelihood$	-18023	-9219	-53030	-30738	-31751	-28936
$LRTest$	194.6602	5.8042	184.2963	644.9959	0.0004	76.1075
N	10048	15100	26008	32155	12385	17086

Note: Bootstrap standard errors (based on 100 replications) in parenthesis. Non estimated parameters: $\beta = 0.5$, $\tau = 0.33$ and $\rho = 0.056$.

Table E.18: Mexico - Labor Market Dynamics and States

	Primary			Secondary			Tertiary		
	M	W	W/M	M	W	W/M	M	W	W/M
h_u									
Data	0.804	0.665	0.827	0.512	0.535	1.047	0.366	0.383	1.045
Model	0.804	0.665	0.827	0.512	0.536	1.048	0.366	0.383	1.045
$h_{u \to e_F}$									
Model	0.261	0.179	0.687	0.261	0.284	1.089	0.216	0.275	1.271
$h_{u \to e_I}$									
Model	0.261	0.179	0.687	0.140	0.140	0.999	0.042	0.053	1.242
$h_{u \to e_S}$									
Model	0.283	0.307	1.086	0.111	0.112	1.011	0.108	0.055	0.513
u									
Data	0.033	0.012	0.369	0.041	0.022	0.536	0.063	0.038	0.600
Model	0.038	0.039	1.026	0.045	0.051	1.149	0.070	0.060	0.860
e_F									
Data	0.240	0.070	0.293	0.459	0.194	0.423	0.571	0.423	0.740
Model	0.341	0.240	0.703	0.493	0.443	0.899	0.635	0.673	1.060
e_I									
Data	0.346	0.078	0.225	0.246	0.093	0.378	0.112	0.081	0.720
Model	0.341	0.240	0.703	0.265	0.219	0.825	0.125	0.129	1.036
e_S									
Data	0.240	0.149	0.619	0.183	0.124	0.678	0.153	0.086	0.563
Model	0.280	0.481	1.721	0.197	0.287	1.455	0.170	0.137	0.807
np									
Data	0.141	0.691	4.912	0.070	0.566	8.042	0.100	0.372	3.720
Model	0.141	0.691	4.912	0.070	0.566	8.042	0.100	0.372	3.720

Table E.19: Mexico - Productivity and Wages

	Primary			Secondary			Tertiary		
	M	W	W/M	M	W	W/M	M	W	W/M
$E[x_F]$									
Model	3.680	2.896	0.787	2.898	2.798	0.966	6.166	6.097	0.989
$SD(x_F)$									
Model	0.417	0.342	0.821	0.009	0.006	0.657	0.037	0.139	3.701
$E[x_I]$									
Model	2.505	2.122	0.847	1.333	1.120	0.840	1.125	0.982	0.873
$SD[x_I]$									
Model	0.406	0.767	1.888	0.617	1.003	1.626	1.285	1.654	1.287
$E[x_S]$									
Model	1.693	1.193	0.705	1.052	0.506	0.481	2.306	0.631	0.274
$SD[x_S]$									
Model	0.945	1.057	1.119	1.063	1.019	0.959	1.985	1.285	0.647
Y_W									
Model	2.686	1.850	0.689	2.391	2.234	0.934	5.196	5.194	1.000
Y_C									
Model	2.220	0.550	0.248	2.124	0.919	0.433	4.348	3.065	0.705
$E[w\|e_F]$									
Data	1.424	1.136	0.798	1.589	1.389	0.874	3.022	2.859	0.946
Model	1.430	1.138	0.796	1.594	1.385	0.869	3.046	2.895	0.950
$SD[w\|e_F]$									
Data	0.588	0.437	0.744	0.748	0.690	0.922	1.852	1.630	0.881
Model	0.576	0.404	0.701	0.723	0.644	0.891	1.922	1.736	0.903
$E[w\|e_I]$									
Data	1.216	1.040	0.855	1.288	1.148	0.891	2.091	2.020	0.966
Model	1.226	1.030	0.840	1.295	1.140	0.880	2.085	2.008	0.963
$SD[w\|e_I]$									
Data	0.517	0.628	1.216	0.663	0.672	1.013	1.574	1.483	0.942
Model	0.526	0.538	1.024	0.668	0.833	1.246	1.734	1.926	1.111
$E[w\|e_S]$									
Data	1.672	1.175	0.703	1.988	1.674	0.842	3.171	2.636	0.831
Model	1.688	1.189	0.705	1.963	1.735	0.884	3.133	2.655	0.847
$SD[w\|e_S$									
Data	1.137	1.039	0.914	1.575	1.634	1.037	2.902	2.620	0.903
Model	1.226	1.208	0.985	1.612	1.949	1.209	3.013	2.989	0.992

Note: $E[x]$ is the average productivity, $SD(x)$ is the standard deviation of productivity, Y_W is the output per worker, Y_C is the output per capita, $E[w|e]$ is the average wage conditional on the employment status e, and finally $SD[w|e]$ is the standard deviation of wages conditioning in the employment status e.

Table E.20: Mexico - Policy Experiments

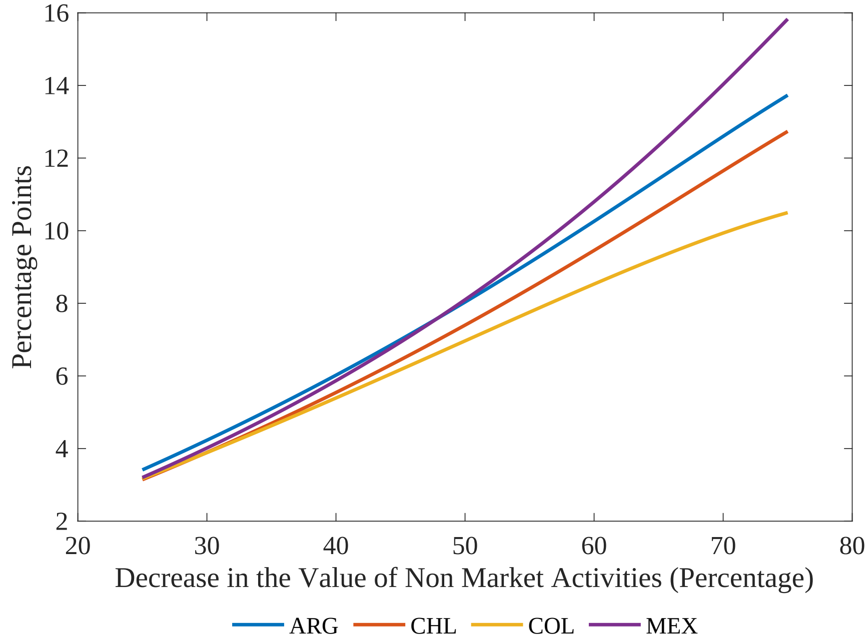
	Benchmark			Policy Exp. 1		Policy Exp. 2	
	M	W	W/M	W	W/M	W	W/M
Primary							
u	0.038	0.039	1.026	0.039	1.026	0.039	1.033
e_F	0.341	0.240	0.703	0.240	0.703	0.242	0.708
e_I	0.341	0.240	0.703	0.240	0.703	0.242	0.708
e_S	0.280	0.481	1.721	0.481	1.721	0.478	1.708
np	0.141	0.691	4.912	0.623	4.428	0.387	2.750
h_u	0.804	0.665	0.827	0.665	0.827	0.661	0.822
Y_W	2.686	1.850	0.689	1.850	0.689	2.049	0.763
Y_C	2.220	0.550	0.248	0.671	0.302	1.207	0.544
$E[w\|e_F]$	1.422	1.132	0.796	1.132	0.796	1.309	0.920
$E[w\|e_I]$	0.980	0.841	0.858	0.841	0.858	0.989	1.009
$E[w\|e_S]$	0.675	0.492	0.729	0.492	0.729	0.611	0.906
Res. W.	0.077	0.087	1.129	0.087	1.129	0.222	2.900
Secondary							
u	0.045	0.051	1.149	0.051	1.149	0.052	1.175
e_F	0.493	0.443	0.899	0.443	0.899	0.453	0.920
e_I	0.265	0.219	0.825	0.219	0.825	0.221	0.833
e_S	0.197	0.287	1.455	0.287	1.455	0.273	1.385
np	0.070	0.566	8.042	0.475	6.750	0.507	7.197
h_u	0.512	0.536	1.048	0.536	1.048	0.526	1.029
Y_W	2.391	2.234	0.934	2.234	0.934	2.467	1.032
Y_C	2.124	0.919	0.433	1.112	0.523	1.153	0.543
$E[w\|e_F]$	1.587	1.391	0.877	1.391	0.877	1.563	0.985
$E[w\|e_I]$	1.153	1.000	0.867	1.000	0.867	1.121	0.972
$E[w\|e_S]$	1.247	0.988	0.792	0.988	0.792	1.124	0.902
Res. W.	0.994	0.679	0.683	0.679	0.683	0.812	0.816
Tertiary							
u	0.070	0.060	0.860	0.060	0.860	0.062	0.888
e_F	0.635	0.673	1.060	0.673	1.060	0.695	1.095
e_I	0.125	0.129	1.036	0.129	1.036	0.121	0.969
e_S	0.170	0.137	0.807	0.137	0.807	0.122	0.714
np	0.100	0.372	3.720	0.299	2.985	0.293	2.928
h_u	0.366	0.383	1.045	0.383	1.045	0.370	1.010
Y_W	5.196	5.194	1.000	5.194	1.000	5.826	1.121
Y_C	4.348	3.065	0.705	3.424	0.787	3.863	0.888
$E[w\|e_F]$	3.021	2.874	0.951	2.874	0.951	3.245	1.074
$E[w\|e_I]$	1.831	1.763	0.963	1.763	0.963	2.040	1.114
$E[w\|e_S]$	1.901	1.599	0.841	1.599	0.841	1.888	0.993
Res. W.	1.406	1.165	0.828	1.165	0.828	1.447	1.029

Table E.20: Mexico - Policy Experiments – continued from previous page

	Policy Exp. 3		Policy Exp. 4 ($\tau = 0$)		Policy Exp. 4 ($c = 0$)	
	W	W/M	W	W/M	W	W/M
			Primary			
u	0.039	1.026	0.039	1.026	0.039	1.027
e_F	0.240	0.703	0.240	0.703	0.240	0.704
e_I	0.240	0.703	0.240	0.703	0.240	0.704
e_S	0.481	1.721	0.481	1.721	0.481	1.720
np	0.687	4.885	0.678	5.549	0.630	9.717
h_u	0.665	0.827	0.665	0.827	0.665	0.827
Y_W	1.851	0.689	1.851	0.689	1.852	0.690
Y_C	0.557	0.251	0.573	0.253	0.659	0.273
$E[w\|e_F]$	1.220	0.858	1.493	0.794	1.143	0.795
$E[w\|e_I]$	0.906	0.924	1.107	0.856	0.852	0.856
$E[w\|e_S]$	0.529	0.784	0.643	0.724	0.504	0.730
Res. W.	0.088	1.146	0.091	1.107	0.108	1.013
			Secondary			
u	0.052	1.166	0.054	1.130	0.050	1.169
e_F	0.450	0.913	0.468	0.884	0.431	0.914
e_I	0.216	0.816	0.210	0.877	0.253	0.828
e_S	0.282	1.428	0.268	1.459	0.267	1.471
np	0.557	7.904	0.532	10.229	0.554	8.510
h_u	0.529	1.033	0.511	1.075	0.558	1.040
Y_W	2.258	0.944	2.318	0.933	2.191	0.940
Y_C	0.949	0.447	1.027	0.458	0.928	0.445
$E[w\|e_F]$	1.486	0.937	1.776	0.887	1.405	0.877
$E[w\|e_I]$	1.073	0.930	1.299	0.880	0.954	0.847
$E[w\|e_S]$	1.065	0.854	1.304	0.804	1.018	0.798
Res. W.	0.700	0.704	0.755	0.681	0.705	0.689
			Tertiary			
u	0.061	0.870	0.063	0.865	0.059	0.861
e_F	0.681	1.073	0.703	1.067	0.658	1.062
e_I	0.125	1.005	0.115	1.050	0.151	1.025
e_S	0.132	0.776	0.119	0.748	0.132	0.801
np	0.357	3.571	0.319	4.242	0.367	3.755
h_u	0.378	1.031	0.366	1.047	0.392	1.047
Y_W	5.246	1.010	5.375	1.007	5.102	1.000
Y_C	3.166	0.728	3.428	0.749	3.041	0.709
$E[w\|e_F]$	3.083	1.020	3.721	0.961	2.883	0.952
$E[w\|e_I]$	1.912	1.045	2.383	0.992	1.656	0.952
$E[w\|e_S]$	1.738	0.914	2.175	0.878	1.619	0.846
Res. W.	1.213	0.863	1.345	0.852	1.182	0.832

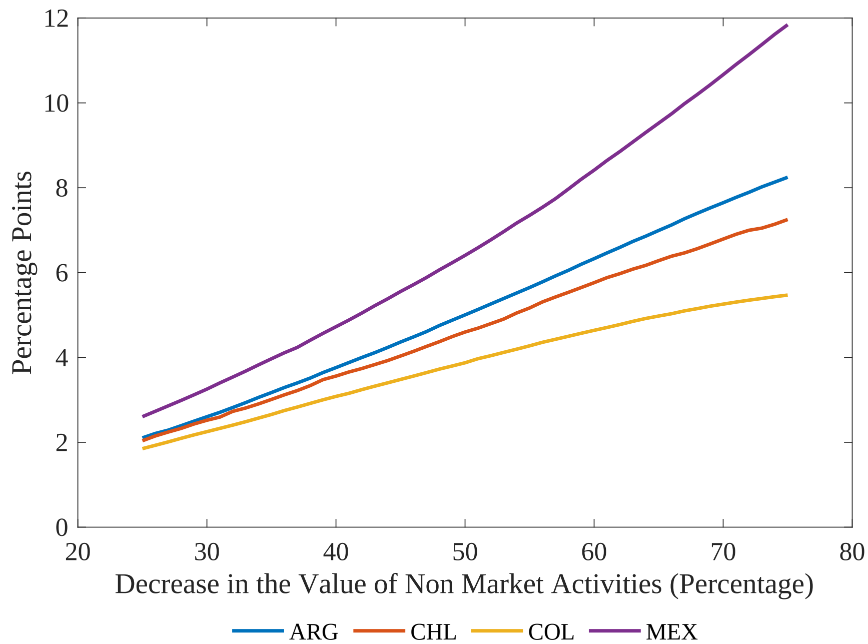
F Additional Material on Policy Experiments

Figure F.1: Child-care Provision Policy: Impact on Female Participation Rates



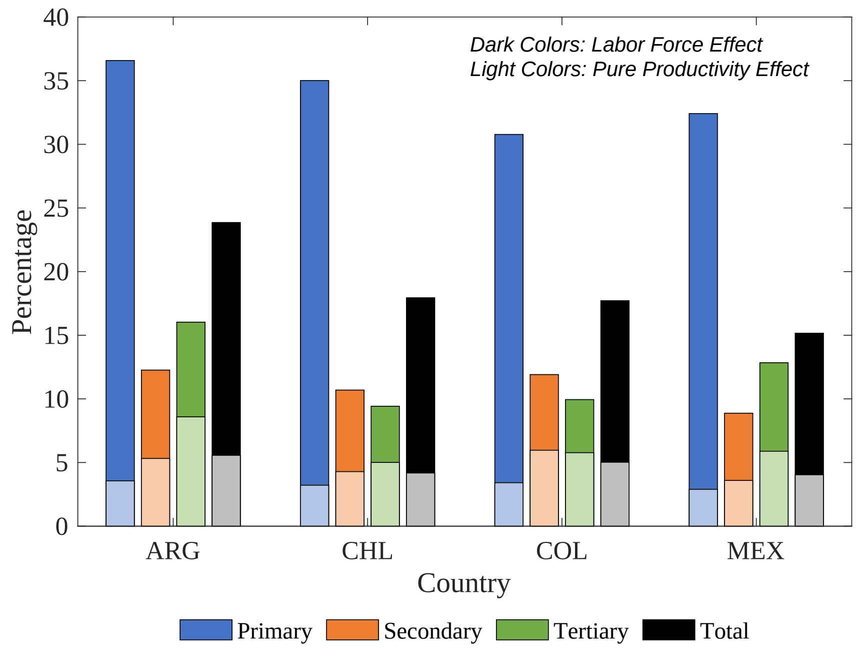
Note: Figure reports percentage points changes in female participation rates as a result of *policy experiment 1*: A range between 25% and 75% of reductions in the average value of non-participation for mother with children aged 5 or younger. See Section 6 for more details.

Figure F.2: Child-care Provision Policy: Impact on Output per Capita



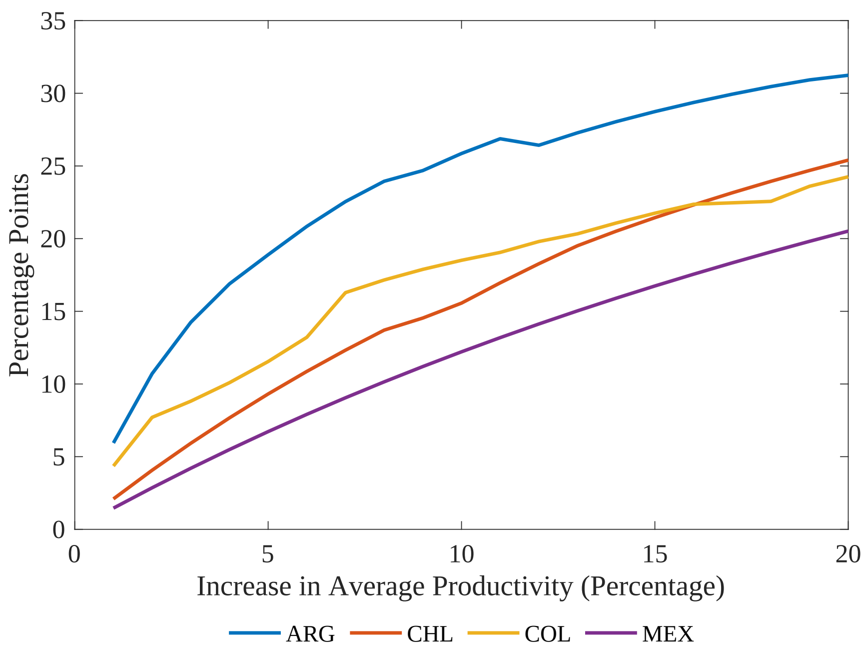
Note: Figure reports percentage points changes in output as a result of *policy experiment 1*: A range between 25% and 75% of reductions in the average value of non-participation for mother with children aged 5 or younger is considered. See Section 6 for more details.

Figure F.3: Increase Female Productivity Policy: Impact on Output per Capita by Channel



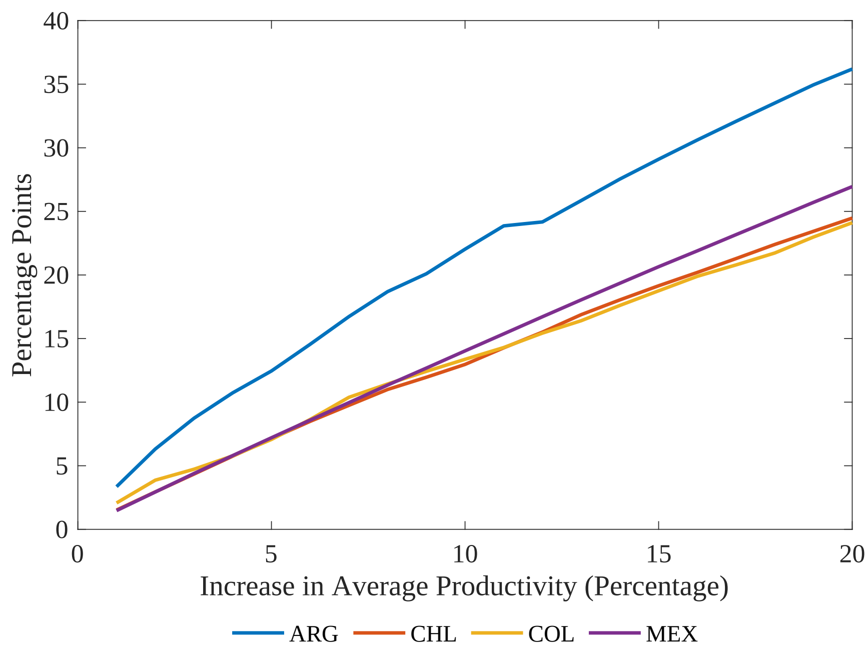
Note: Figure reports percentage points changes in output as a result of *policy experiment 2*: increasing the average productivity of women by 10%. See Section 6 for more details. The overall increase is decomposed in the portion due to the 10% productivity increase (Pure Productivity Effects) and the portion due to the increase in participation resulting from the productivity increase (Labor Force Effect).

Figure F.4: Increase Female Productivity Policy: Impact on Female Participation Rates



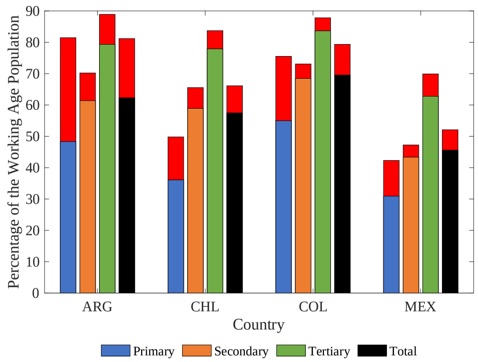
NOTE: Figure reports percentage points changes in participation rates as a result of *policy experiment 2*: A range between 1% and 20% increasing the average productivity of women is considered. See Section 6 for more details.

Figure F.5: Increase Female Productivity Policy: Impact on Output per Capita

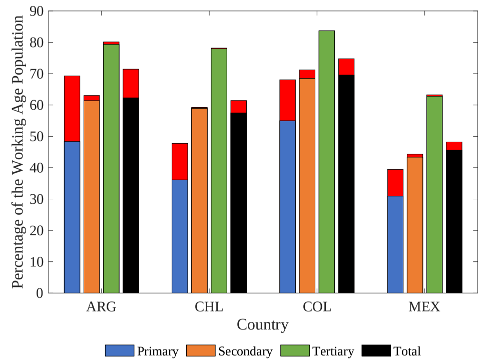


Note: Figure reports percentage points changes in output as a result of *policy experiment 2*: A range between 1% and 20% increasing the average productivity of women is considered. See Section 6 for more details.

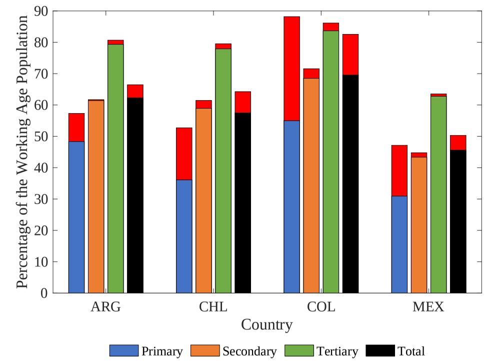
Figure F.6: Policy 5: Increase Female Productivity by Sector Policy



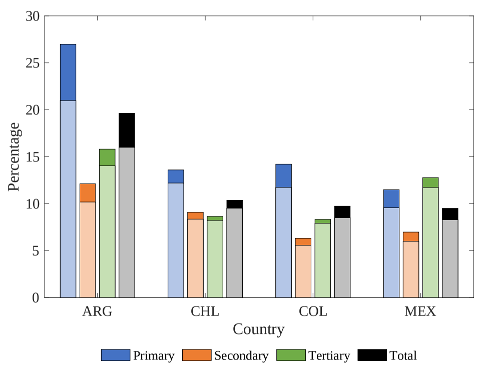
(a) Impact on Female Part. Rates (Formal)



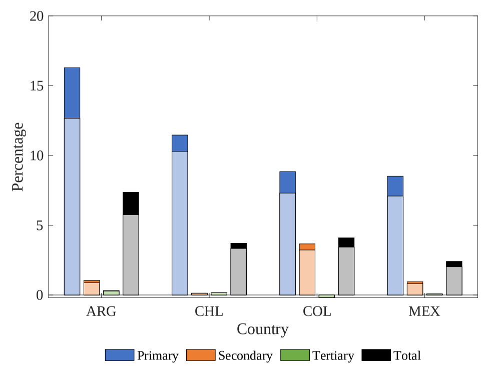
(b) Impact on Female Part. Rates (Informal)



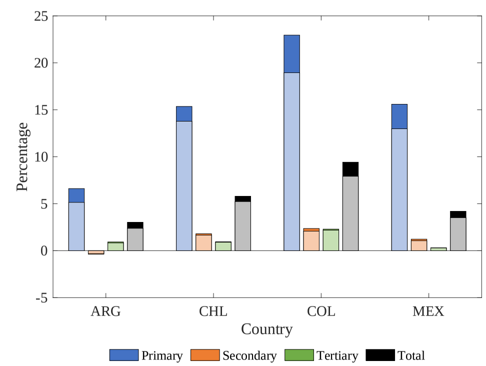
(c) Impact on Female Part. Rates (Self-Emp.)



(d) Impact on Output per Capita (Formal)



(e) Impact on Output per Capita (Informal)



(f) Impact on Output per Capita (Self-Emp.)

NOTE: The policy reported in panels (a) and (d) increases women average productivity by 10% in the formal sector; the one in panels (b) and (e) increases women average productivity by 10% in the informal sector; the one in panels (c) and (f) increases women average productivity by 10% for self-employed. In panels (a), (b) and (c) the overall length of the column is the post-policy participation rate. The red darker segment is the impact of the policy. In panels (d), (e) and (f) we report the percentage points changes in output per capita as a result of the policy. Light colored bars represent the effect on output taking into account differences in average weekly hours worked by men and women. See Section 6 for more details.

G Robustness Analysis

This section of the appendix provides robustness checks. The first concerns the distributional assumption on the value of non participation distribution $Q_i(z)$; the second the Nash-bargaining weight β; and the third the mobility rates λ and δ.

The first robustness check is reported in Figure G.1. Since the empirical identification of the value of non participation distribution $Q_i(z)$ is quite limited – we can only use one moment: the proportion of agents non-participating – we assess the importance of the specific distributional assumption we make. We evaluate importance by re-estimating the model under different distributional assumptions and then re-running the relevant policy experiments. In this case, the most relevant experiment is *policy experiment 1* where we reduce in half the average value of non-participation for mother with children aged 5 or younger. It is the most relevant because the policy directly affect non-participation values. We are constrained in the alternative distributional assumptions we can make. First, we can identify and estimate only one parameter. Second, the distribution should be on a positive support. We have chosen to use a lognormal distribution since it satisfies the support condition. In order to make it a one-parameter distribution, we fix the shape parameter σ and estimate only the location parameter. We fix σ at two values: 1 and 0.5.

The original result under the exponential distribution assumption (ED) is in Panel (a). The results under the alternative lognormal distribution assumption (LND) are in Panel (b) and (c). As in Figure **??**, the overall length of the column is the post-policy participation rate and the red darker segment is the impact of the policy. See Section 6 for more details. The results under the alternative distributional assumptions are qualitatively similar to benchmark: same direction of the impact, same ranking of magnitudes between schooling levels, same ranking across countries.

The second robustness check we perform refers to the Nash-bargaining weight β. In the paper, we impose symmetric bargaining for both men and women, fixing the parameter at 0.5. We are forced to do this because we do not have enough data information to identify it, a common problem in the literature (Flinn, 2006; Flabbi, 2010). Still, the assumption may be more restrictive in our context because it does also imply that men and women share the same parameter. There are a number of reasons why that may not be the case. Some contributions have used this parameter as a proxy for possible discrimination, even if the empirical evidence is mixed (Eckstein and Wolpin, 1999; Bartolucci, 2013). Others have suggested that women and men are systematically different in their bargaining process (Castillo et al., 2013), something that could be captured by the parameter. In general, it could be an additional structural parameter over which men and women could differ, just as we currently allow for differences in productivity and mobility rates.

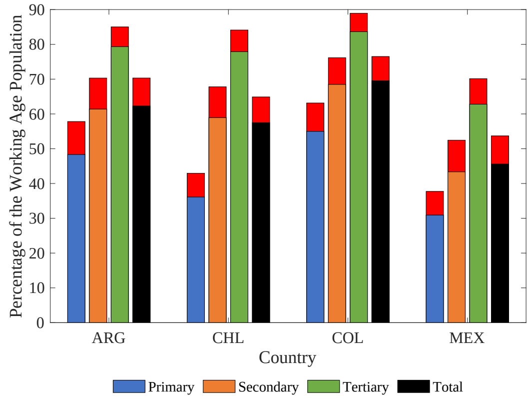
We check robustness with respect to the restriction by focusing on the policy where its impact should be largest: *Policy Experiment 2* where we increase the average productivity of women in the three sectors by 10%. Results of the exercise are reported in Figure G.2. Once again, changes in the parameters deliver result qualitatively similar to benchmark. Primary sees the strongest impact, impact that becomes slightly larger when women have more bargaining power. Across countries, Argentina experience the largest overall impact, the extent of which is almost unaffected by the different parameter combinations. The only country and schooling level where we see important differences is Colombia in the Primary school level: in this case, the impact on primary is significantly reduced when women have a high bargaining power ($\beta_W = 0.6$).

The third robustness exercise concerns the restriction that the arrival and termination rates for formal and informal employees are the same. As we discuss in Section 4.2, we have to impose $\lambda_F = \lambda_I$ and $\delta_F = \delta_I$ not because the theoretical identification cannot be attained but because the empirical identification is very weak for a number of country-education-gender groups. For a significant number of estimation samples we do not have enough data variation to obtain convergence of the likelihood function in the feasible parameters space.

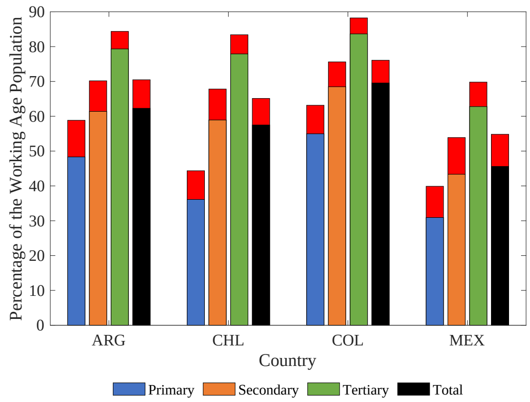
In this robustness section we report results for the one country on which it is possible to attain empirical identification on all estimation samples: Argentina. For Argentina, we estimate the model with and without the restriction. The model with the restriction is the benchmark we estimate in the paper and the model without the restriction allows both the arrival rate λ and the termination rate δ to be different for formals and informals. We use the estimation results to perform a specification test. Since the specification of the model with the restriction is nested in the one of the model without the restriction, it is straightforward to perform Likelihood Ratio tests where the null is the restricted model and the alternative is the unrestricted model. Table G.1 reports statistics and P-values of the test. The restriction is rejected only on one sample out of six: men with Secondary education. Even in this case, the differences in point estimates are not very large.[8]

[8]The arrival rate for formal is 0.1741 and for informal is 0.1106; the termination rate for formal is 0.0154 and for informal 0.0298.

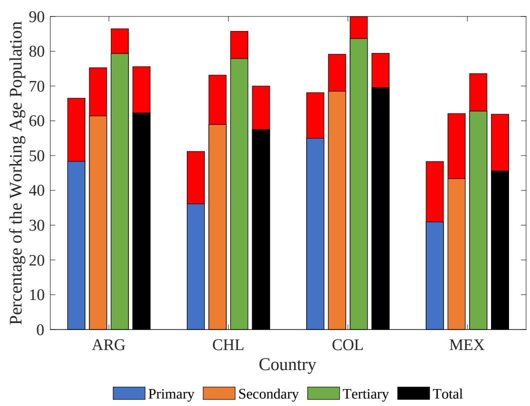
Figure G.1: Robustness Check 1: Child-care Provision Policy using Different Distributional Assumptions for the Value of Non Participation Distribution $Q_i(z)$



(a) Female Participation Rates (ED)



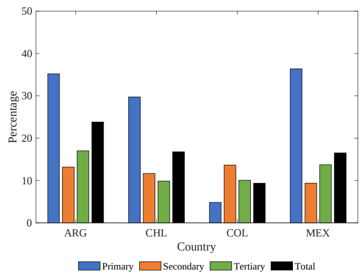
(b) Female Participation Rates (LND $\sigma = 1$)



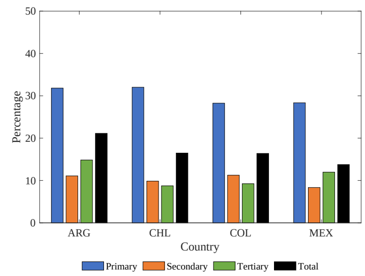
(c) Female Participation Rates (LND $\sigma = 0.5$)

NOTE: The figures report *policy experiment 1* under different parametric assumptions for the $Q_i(z)$ distribution. For each assumption, we re-estimate the model and re-run the experiments. The original result under the exponential distribution assumption (ED) is in Panel (a). The results under the alternative lognormal distribution assumption (LND) are in Panel (b) and (c). As in Figure F.1, the overall length of the column is the post-policy participation rate and the red darker segment is the impact of the policy. See Section 6 for more details.

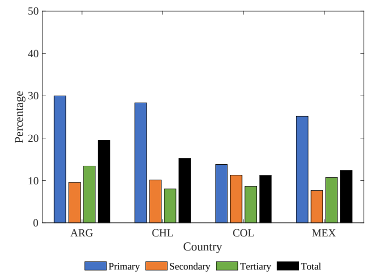
Figure G.2: Robustness Check 2: Increase in Female Productivity Policy using Different Nash Bargaining Coefficients β_W, β_M



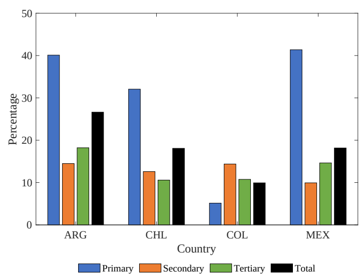
(a) $\beta_M = 0.4, \beta_W = 0.4$



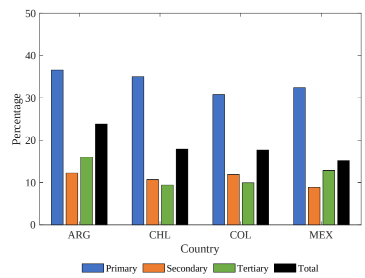
(b) $\beta_M = 0.4, \beta_W = 0.5$



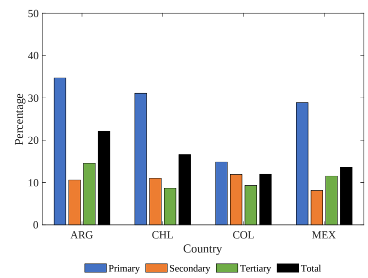
(c) $\beta_M = 0.4, \beta_W = 0.6$



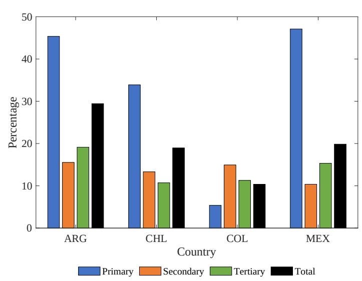
(d) $\beta_M = 0.5, \beta_W = 0.4$



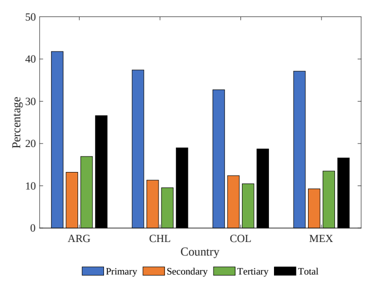
(e) $\beta_M = 0.5, \beta_W = 0.5$



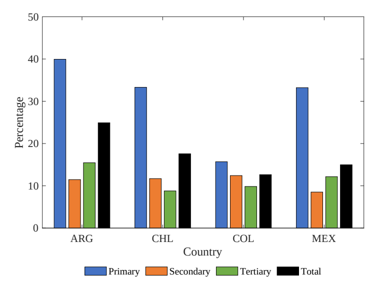
(f) $\beta_M = 0.5, \beta_W = 0.6$



(g) $\beta_M = 0.6, \beta_W = 0.4$



(h) $\beta_M = 0.6, \beta_W = 0.5$



(i) $\beta_M = 0.6, \beta_W = 0.6$

NOTE: The figures report *policy experiment 2* under different values combinations of nash-bargaining coefficients β_W, β_M. For each combination, we re-run the experiments. The original result under symmetric bargaining is reported in Panel (e). All panels report the percentage points changes in output as a result of the policy. As in Figure F.5, we report the effect on output taking into account differences in average weekly hours worked by men and women. See Section 6 for more details.

Table G.1: Likelihood Ratio Test for the restriction $\lambda_F = \lambda_I$ and $\delta_F = \delta_I$

	Argentina			
	Men		Women	
	Test Statistic	P-Value	Test Statistic	P-Value
Primary	0.0015	0.9993	0.0000	1.0000
Secondary	15.8573	0.0004	0.0000	1.0000
Tertiary	0.0384	0.9810	0.8226	0.6628

Note: The Table reports test statistics and P-values of the joint test with: $H_0 : \{\lambda_F = \lambda_I,\, \delta_F = \delta_I\}$

References

Bagger, Jesper, Fran Fontaine, Fabien Postel-Vinay, and Jean-Marc Robin, "Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics," *American Economic Review*, 2014, *104* (6), 1551–96.

Bartolucci, Cristian, "Gender Wage Gaps Reconsidered A Structural Approach Using Matched Employer-Employee Data," *Journal of Human Resources*, 2013, *48* (4), 998–1034.

Bobba, Matteo, Luca Flabbi, and Santiago Levy, "Labor Market Search, Informality and Schooling Investments," Discussion Paper 11170, Institute for the Study of Labor (IZA) 2017.

Borowczyk-Martins, Daniel, Jake Bradley, and Linas Tarasonis, "Racial discrimination in the us labor market: Employment and wage differentials by skill," *Labour Economics*, 2017, *49*, 106–127.

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin, "Wage Bargaining with On-the-Job Search: Theory and Evidence," *Econometrica*, 2006, *74* (2), 323–364.

Castillo, Marco, Ragan Petrie, Maximo Torero, and Lise Vesterlund, "Gender differences in bargaining outcomes: A field experiment on discrimination," *Journal of Public Economics*, 2013, *99*, 35–48.

Daymont, Thomas N. and Paul J. Andrisani, "Job Preferences, College Major, and the Gender Gap in Earnings," *The Journal of Human Resources*, 1984, *19* (3), 408.

Dey, Matthew S. and Christopher J. Flinn, "An Equilibrium Model of Health Insurance Provision and Wage Determination," *Econometrica*, 2005, *73* (2), 571–627.

Eckstein, Zvi and Gerard J. van den Berg, "Empirical labor search: A survey," *Journal of Econometrics*, 2007, *136* (2), 531–564.

______ **and Kenneth I Wolpin**, "Estimating the effect of racial discrimination on first job wage offers," *The Review of Economics and Statistics*, 1999, *81* (3), 384–392.

Flabbi, Luca, "Gender Discrimination Estimation In A Search Model With Matching And Bargaining," *International Economic Review*, 2010, *51* (3), 745–783.

Flinn, Christopher, "Minimum Wage Effects on Labor Market Outcomes under Search, Bargaining and Endogenous Contact Rates," *Econometrica*, 2006, *73*, 1013–1062.

______ **and James Heckman**, "New methods for analyzing structural models of labor force dynamics," *Journal of Econometrics*, 1982, *18* (1), 115 – 168.

Kanbur, Ravi, "Conceptualizing Informality: Regulation and Enforcement.," *Indian Journal of Labour Economics*, 2009, *52* (1), 33–42.

Levy, Santiago, *Good Intentions, Bad Outcomes: Social Policy, Informality, and Economic Growth in Mexico*, Washington DC, The Brooking Institution Press, 2008.

Meghir, Costas, Renata Narita, and Jean-Marc Robin, "Wages and Informality in Developing Countries," *American Economic Review*, 2015, *105* (4), 1509–46.

Pavan, Ronni, "Career choice and wage growth," *Journal of Labor Economics*, 2011, *29* (3), 549–587.