



DEPARTAMENTO DE ECONOMÍA

SDT 265

**ABILITY, SCHOOLING CHOICES AND
GENDER LABOR MARKET
DISCRIMINATION: EVIDENCE FOR
CHILE**

Autor: David
Bravo, Claudia Sanhueza y
Sergio Urzúa.

Santiago, Jul. 2007

**Serie Documentos de Trabajo
N 265**

**Ability, Schooling Choices And Gender Labor Market
Discrimination: Evidence For Chile**

David Bravo

Claudia Sanhueza

Sergio Urzúa

Departamento de Economía
Universidad de Chile

Abstract

This paper presents a comprehensive analysis of the gender differences in the Chilean labor market. We formally deal with the selection of the individuals into schooling levels and its consequences on the gender gaps. Our approach allows for the presence of not only heterogeneity in observable variables but also unobserved heterogeneity. We link this unobserved heterogeneity to unobserved scholastic ability. In the analysis, we utilize a new and rich data set for Chile. This data set contains information on labor market outcomes (including labor history), on schooling attainment and schooling performance, and on a complete set of variables characterizing the family background of the individuals in the sample

Our results show that there exist statistically significant gender differences in several dimensions of the Chilean labor market. Nevertheless, we show that these gaps critically depend on the schooling level of the individuals considered in the analysis. For example, the results indicate that there are no gender differences in labor market variables among college graduates (except in the case of hourly wages).

We interpret our results with prudence. Specifically, instead of interpreting our findings as decisive evidence of the existence of discrimination in the Chilean labor market, we argue that future research based on better information might indeed explain some of the unexplained labor market gaps presented in this paper. In this context, our results represent a new and important attempt to provide a full understanding of the structural causes of gender gaps in the Chilean labor market but they are not conclusive.

Keywords:

Labor Market Discrimination; unobserved heterogeneity; Gender gaps.



ABILITY, SCHOOLING CHOICES AND GENDER LABOR MARKET DISCRIMINATION: EVIDENCE FOR CHILE¹

David Bravo (dbravo@econ.uchile.cl)
Centro de Microdatos, Departamento de Economía, U.de Chile
Claudia Sanhueza (csanhueza@econ.uchile.cl)
Centro de Microdatos, Departamento de Economía, U.de Chile
Sergio Urzúa (s-urzua@northwestern.edu)
Department of Economics, Northwestern University

First Draft, December 2006
This Draft, July 2007

Abstract

This paper presents a comprehensive analysis of the gender differences in the Chilean labor market. We formally deal with the selection of the individuals into schooling levels and its consequences on the gender gaps. Our approach allows for the presence of not only heterogeneity in observable variables but also unobserved heterogeneity. We link this unobserved heterogeneity to unobserved scholastic ability. In the analysis, we utilize a new and rich data set for Chile. This data set contains information on labor market outcomes (including labor history), on schooling attainment and schooling performance, and on a complete set of variables characterizing the family background of the individuals in the sample

Our results show that there exist statistically significant gender differences in several dimensions of the Chilean labor market. Nevertheless, we show that these gaps critically depend on the schooling level of the individuals considered in the analysis. For example, the results indicate that there are no gender differences in labor market variables among college graduates (except in the case of hourly wages).

We interpret our results with prudence. Specifically, instead of interpreting our findings as decisive evidence of the existence of discrimination in the Chilean labor market, we argue that future research based on better information might indeed explain some of the unexplained labor market gaps presented in this paper. In this context, our results represent a new and important attempt to provide a full understanding of the structural causes of gender gaps in the Chilean labor market but they are not conclusive.

Keywords: Labor Market Discrimination; unobserved heterogeneity; Gender gaps.

¹ This paper was prepared as part of the project "Gender Discrimination and Economic Outcomes in Chile" of the Centro de Microdatos (Departamento de Economía, Universidad de Chile) under the direction of Bravo (corresponding author). The authors acknowledge research support from the Research Network of the Inter-American Development Bank. We would like to thank Andrea Moro and Hugo Ñopo for helpful comments.



ABILITY, SCHOOLING CHOICES AND GENDER LABOR MARKET DISCRIMINATION: EVIDENCE FOR CHILE

David Bravo (dbravo@econ.uchile.cl)
Centro de Microdatos, Departamento de Economía, U.de Chile
Claudia Sanhueza (csanhueza@econ.uchile.cl)
Centro de Microdatos, Departamento de Economía, U.de Chile
Sergio Urzúa (s-urzua@northwestern.edu)
Department of Economics, Northwestern University

First Draft, December 2006
This Draft, July 2007

Abstract

This paper presents a comprehensive analysis of the gender differences in the Chilean labor market. We formally deal with the selection of the individuals into schooling levels and its consequences on the gender gaps. Our approach allows for the presence of not only heterogeneity in observable variables but also unobserved heterogeneity. We link this unobserved heterogeneity to unobserved scholastic ability. In the analysis, we utilize a new and rich data set for Chile. This data set contains information on labor market outcomes (including labor history), on schooling attainment and schooling performance, and on a complete set of variables characterizing the family background of the individuals in the sample

Our results show that there exist statistically significant gender differences in several dimensions of the Chilean labor market. Nevertheless, we show that these gaps critically depend on the schooling level of the individuals considered in the analysis. For example, the results indicate that there are no gender differences in labor market variables among college graduates (except in the case of hourly wages).

We interpret our results with prudence. Specifically, instead of interpreting our findings as decisive evidence of the existence of discrimination in the Chilean labor market, we argue that future research based on better information might indeed explain some of the unexplained labor market gaps presented in this paper. In this context, our results represent a new and important attempt to provide a full understanding of the structural causes of gender gaps in the Chilean labor market but they are not conclusive.

I. Introduction

Gender gaps in a variety of labor and educational outcomes (e.g. wages, earnings, employment, schooling levels) are well documented. The structural reasons behind these gaps, however, are not fully understood.

This paper contributes to the literature by studying gender differences in a framework in which schooling decisions and labor market outcomes are

endogenously determined. Our framework also allows individual heterogeneity not only from the point of view of observable characteristics but also unobserved variables. We assume that individuals know this additional source of heterogeneity, and they base their schooling and labor market decisions on it. Unobserved heterogeneity plays a crucial role in our approach.

Ours is a challenging task for several reasons. First, a comprehensive analysis of gender differences in a variety of outcomes is subject to the usual and irremediable data limitations. Second, the natural complexity associated with econometric models of multiple, endogenous, and correlated outcomes makes these models usually not very empirically appealing. And finally, the fact that we allow individuals decisions to depend on variables unobserved by the researcher but known to the agent represents an additional challenge of our approach. Nevertheless, we deal with each of these difficulties. First, we utilize a new data set from Chile that contains detailed information on labor market and schooling outcomes at the individual level. Second, we postulate a simple factor structure model based on economic theory that simplifies the manner we can deal with multiple endogenous variables. And finally, we interpret this factor as unobserved heterogeneity since the researcher does not need to know the individual factor (although it is assumed to be known by the individual). We argue that the factor represents a combination of different scholastic skills (cognitive and noncognitive skills).

As previously mentioned, we implement our approach using new information from Chile. The Chilean case provides an interesting example of apparently huge gender gaps in different dimensions of the labor market. Table 1 presents basic information for a variety of schooling and labor market outcomes obtained from a sample of males and females with ages between 28 and 40 years.²

The evidence in Table 1 provides an initial flavor of the gender differences that motivate the idea of this paper. A comparison of the schooling outcomes (Panel A in Table 1) leads to conclude that, in average, (i) women are slightly more educated than males, (ii) women are less likely to repeat a grade in both primary and secondary school, and (iii) women show a better performance in school than males (measured by the average grade in secondary school). However, this educational advantage of women over men seems to have no consequences on the labor market. The evidence in Panel B illustrates this point. It shows that males overwhelmingly dominate females in every single dimension of the labor market (monthly earnings, employment, and experience).

This paper studies the factors explaining these gender differences in labor market and schooling outcomes.

The paper is organized as follows. Section II describes the data. Section III presents evidence on the differences in labor market outcomes between males and females

² The information comes from the Social Protection Survey 2002 of Chile (SPS02) which is the source of information used in this paper. This survey is described in detail in Section II.

using a conventional approach. Section IV introduces our model and discusses its empirical implementation. Section V presents a discussion of our results. Section VI concludes.

II. Data

This paper uses information from the Chilean Social Protection Survey 2002 (SPS02). This survey was designed to identify and analyze the most important determinants of the social security decisions (participation in the social security system) among Chileans. In order to do this, a representative sample of 17,246 participants of the Chilean pension system was interviewed between June of 2002 and January of 2003. For each individual in the sample, the survey collected information on household composition (ages, genders and schooling levels of the household members as well as their relations with the interviewee), current employment status, different sources of income, schooling (maximum schooling attained, average grades in primary and secondary school, characteristics of the primary and secondary school attended), family history (mother's and father's education, characteristics of the place of residence where the individual grew up, and number of previous relationships), labor history since age 15 or since 1980 depending on the year the individual became 15 years old (periods of employment, unemployment and inactivity), training programs (information on the three most important training programs since 1980), expectations (job, retirement and life), savings (instruments and amounts), and a set of variables describing the individual's knowledge of the characteristics and performance of the Chilean pension system.

Information regarding the Social Protection Survey can be found in Bravo (2004) and Bravo, Behrman, Mitchell and Todd (2006)³.

We use a sample of individuals with ages in the range of 28 and 40 years. This group represents approximately the 21% of the original sample (3,566 versus 17,246).⁴

We restrict the ages of sample for several reasons. First, since the information on labor history begins only in 1980 (or since age 15), by using individuals 28-40 years old we assure that our sample report complete labor histories from age 18. Second, since schooling is an important ingredient of our analysis, by excluding individuals 27

³ Data and technical documentation can be found in www.proteccionsocial.cl

⁴ Our sample is obtained after considering the following exclusions. We first exclude the military sample (57 individuals) and individuals reporting as occupation "family member without salary" (12 individuals). Then, we exclude individuals 27 years old or younger and 41 years old or older. With this the sample reduces from 17,177 to 5,439. Finally, individuals with missing values in any of the following variables are excluded: "years of education", "mother's education", "father's education", "growing up in poverty" and "growing up in a single parent household". This exclusion reduces the sample to the final 3,566 individuals. It is worth noting that the final exclusion is required since for each individual we need to have valid values for the controls entering in the schooling decision model presented in Section III.1.

years old and younger we focus our attention on individuals that have most likely reached their final schooling level.⁵

Finally, it is worth noting that the current Chilean schooling system was designed only in the early 80s. Therefore, since our analysis includes information on the characteristics of the primary and secondary schools in which the individual was enrolled, by restricting the analysis to the individuals with ages 28-40, we assure that such information is available for most of our sample.

Table A.1 presents the summary statistics of the variables used in this paper.

III. The Conventional Gender Gap Analysis

The gender differences in labor market outcomes are usually analyzed in the context of linear models in which the variable of interest is regressed on the gender dummy variable and set of additional controls. The coefficient associated with the gender dummy is interpreted as the estimated gender gap. Given its popularity, our first attempt to quantify gender gaps follows closely this idea. Table 2 presents the results from the following model of (log) hourly wages ($\ln W$):

$$\ln W = \alpha + \varphi \text{Gender} + \beta X + U \quad (1)$$

where *Gender* represents the gender dummy (*Gender*=1 if individual is Male and 0 if Females), *X* represents individual's observable characteristics, and *U* is the error term in the regression. In this simple model, the (conditional) gender gap is simply φ . Each column in Table 2 represents a different specification of (1). In particular, column (A) presents the results of a model in which we include the characteristics of both the place of residence and occupation in the vector of controls *X*. Column (B) adds a set of variables controlling for the individual's accumulated experience and column (C) adds to the controls in (B) a set of variables controlling for schooling levels. The results indicate that males make approximately 23% more than females in terms of hourly wages. This gender gap is statistically significant regardless of the column analyzed.

The last model in Table 1 (column D) includes a correction for the fact that the labor market outcome is reported only for individuals working (Heckman, 1974). This is particularly important given the gender differences in employment rates reported in Table 1 (panel B). Thus, the model in column D is:

⁵ A more general analysis of the schooling decisions would require a dynamic model for schooling choices. The SPS02 does not allow us to carry out such analysis.

$$\ln W = \alpha + \varphi \text{Gender} + \beta X + U \text{ if wage is observed } (D=1)$$

$$D = 1[\gamma Z + V > 0] \quad (2)$$

where $1[A]$ is an indicator function that takes a value of 1 if A is true and zero otherwise, Z is a vector of observables and V represents the unobservables. $D=1[.]$ is the censoring rule for wages. In Z we include variables such as number of children, whether or not the individual grew up in a poor household, mother's and father's occupational status. The estimated gap after correcting for selection is 29% and it is statistically significant. Thus, after controlling for selections, we not only find a significant but larger gender gap in wages (compare to the ones estimated without using the correction). This fact illustrates the importance of paying particular attention to individual's endogenous decisions (in this case employment decisions) when analyzing gender gaps. We exploit this point in the following section.

The analysis of gender gaps in wages is interesting and important but it represents only one dimension of many among which males and females can differ. We first extend our analysis to the case of monthly hours worked. We model (log) hours worked using a linear-in-parameter models similar to (1) and the same set of controls as the ones utilized for wages. Table 3 presents the estimates of gender gaps in this case. The structure of this table is identical to the one in Table 2. The results from columns (A), (B) and (C) suggest that males work approximately 11% more hours per month than females. This difference is statistically significant and it is stable across the three specifications. However, the last column in Table 3 presents (again) a different story. Unlike the results for wages, the correction for selection significantly reduces the gender gap in hours worked. The estimated gap is only 0.04% and it is not statistically significant.

We also extend our analysis to employment status. In this case, we use a probit model instead of a linear regression model. Table 4 presents the results for three different specifications. For each specification, we present both estimated coefficients and estimated marginal effects.⁶ The results indicate that males are 22% more likely to report an employment (during the month previous to the date of the interview) than females when schooling and experience are excluded as controls. When schooling or schooling and experience are included as controls the estimated gap is 14%. The gaps are statistically significant regardless of the specification.

In summary, the results show that men dominate women in every labor market outcome. Additionally, the results are robust across different specifications and only in the case of hours worked and after controlling for selection we find neither sizeable nor statistically significant gender differences.

Notice that up this point we have utilized the individual's schooling decisions and accumulated experience as exogenous regressors. However, in principle these

⁶ The marginal effects are computed at the mean values of the variables in the model.

variables can also be subject to gender differences. Tables 5 and 6 present evidence on this point. The implications of separate analyses of schooling choices and accumulated experience on our previous results are left for the next section where they are discussed in the context of a more general framework than the one used here.⁷

We model accumulated experience assuming that, whatever experience level is observed in the sample, it is the result of a decision involving three alternatives: less than 10 years of experience, between 10 and 15 years of experience, and more than 15 years of experience. This decision is assumed to depend on the schooling level of the individual as well as on his family background (mother's and father's education, broken home, age, and growing up in poverty). Given this set up, we compute the gender gaps in accumulated experience by estimating a multinomial probit model. Table 5 presents the estimated coefficients and marginal effects. The estimates associated with the gender dummy are all significant and suggest that males are considerably more likely to report more experience than females. Specifically, males are 40% less likely to report less than 10 years of experience and 29% more likely to report more than 15 years of experience than females.

The analysis of gender differences in schooling decisions is also relevant in the context of the previous results. On the one hand, if males are in fact more likely to report higher schooling levels than females (after controlling for observable characteristics), then the gender differences in labor market outcomes (including accumulated experience) could be simply interpreted as the result of gender differences in accumulated human capital. On the contrary, if females are more likely to report higher schooling levels than males, then the estimated gender differences in labor market outcomes could be interpreted as downward biased estimates of the actual gaps.

Table 6 sheds light on existence of gender gaps in schooling decisions. It presents the coefficients and marginal effects obtained from a multinomial schooling choice model. The model is estimated using the maximum schooling levels reported by the individuals in the sample. The schooling levels considered are: primary school, secondary school, some post-secondary education, and complete tertiary education (college graduates). The results show that (if anything) females are more likely than males to reach higher schooling levels.

The advantage of females over males in schooling achievement/attainment is confirmed in Table 7. This table presents the estimated gender gaps for three variables measuring schooling performance: probability of a grade repeated during primary school, probability of a grade repeated during secondary school, and average grades during secondary school. For each variable we consistently observe that females outperform males. Males are 7% and 4% more likely to repeat a grade

⁷ This is particularly important if we consider that schooling decisions and accumulated experience are probably endogenous variables in the context of the models presented in Tables 2, 3 and 4. The model presents in the next section deals with this possibility.

during primary and secondary school, respectively, and males in average have a significantly lower grades during high school than females (0.31 points of test's standard deviation).

Therefore, the evidence presented in Tables 6 and 7 leads us to conclude that females should be better prepared than males to face the labor market. This also implies that by not including the gender differences in schooling variables our previous results might be underestimating the actual unexplained gender gaps (or discrimination). We analyze this possibility by introducing a more general model in which schooling decisions, schooling achievement, employment decisions, accumulated experience, hours worked and hourly wages are modeled jointly.

IV. A Model of Schooling and Labor Market Outcomes under Unobserved Heterogeneity

The model in this section follows the analysis in Heckman, Stixrud and Urzua (2006). Heckman, Stixrud and Urzua (2006) postulate and estimate a model with two underlying sources of unobserved heterogeneity that they interpreted as abilities (cognitive and noncognitive abilities). Conditioning on the observables, these factors account for all of the dependence across choices in the model. They show that both abilities play a crucial role explaining a variety of labor market and behavioral outcomes.

In this paper we postulate the existence of only one underlying factor representing unobserved heterogeneity. This is mainly due to the fact that we do not have a set of cognitive and noncognitive variables in the SPS02 sample. Consequently, we interpret the source of unobserved heterogeneity as a combination of both cognitive and noncognitive abilities.⁸ The identification of its distribution is discussed in Section III.4 below.

Let θ denote the unobserved heterogeneity or latent ability. We assume this latent ability determines the individual's schooling and labor market outcomes, and that there are not intrinsic differences between males and females regarding θ , so that we can work with an overall distribution for θ .⁹

⁸ We expect to extend our model to a multi-factor model in which we can precisely distinguish between cognitive and noncognitive abilities.

⁹ The alternative would have been the estimation of gender specific distributions. We consider this an attractive possibility. However, given the data limitations (sample size) and the large number of parameters in the model, we prefer to follow a simple analysis by considering an overall distribution for θ . Future research should consider the potential differences in unobserved heterogeneity between males and females.

IV.1. The Model for Schooling

Each agent chooses the level of schooling, among \bar{S} possibilities, such that he maximizes his benefit. Let I_s represent the net benefit associated with each schooling level s ($s=\{1, \dots, \bar{S}\}$) and assume the following linear-in-the-parameters model for I_s :

$$I_s = \varphi_s \text{Gender} + \beta_s X_s + \alpha_s \theta + e_s \quad \text{for } s = 1, \dots, \bar{S} \quad (3)$$

here φ_s represents the gender gap associated with the schooling level s , X_s is a vector of observed variables determining schooling, β_s is the associated vector of parameters, α_s is the factor loading associated with the latent ability, and e_s represents an idiosyncratic component assumed to be independent of θ , and X_s . The individual components $\{e_s\}_{s=1}^{\bar{S}}$ are mutually independent. All of the dependence across schooling choices comes through the observable, X_s , and the latent ability θ .

The agent chooses the level of schooling with the highest benefit. Formally,

$$s^* = \underset{s \in \{1, \dots, \bar{S}\}}{\operatorname{argmax}} \{I_s\} \quad (4)$$

where s^* denotes the individual's chosen schooling level. Notice that conditional on X_s (with $s=1, \dots, \bar{S}$) and θ , equations (3) and (4) can be interpreted as a standard discrete choice model.

IV.2. The Model for Accumulated Experience

The model also treats accumulated experience as an endogenous outcome. Specifically, after deciding the schooling levels, agents are assumed to pick their experience levels \bar{A} different alternatives. As in the schooling model, given the schooling level s , we assume a linear-in-the-parameters specification for the benefits associated with the experience level $a(s)$ ($I_{a(s)}$):

$$I_{a(s)} = \varphi_{a(s)} \text{Gender} + \beta_{a(s)} X_a + \alpha_{a(s)} \theta + e_{a(s)} \quad \text{for } a(s) = 1, \dots, \bar{A} \text{ and } s = 1, \dots, \bar{S}$$

where $\varphi_{a(s)}$ is the gender gap, X_a is the vector of observed variables, $\beta_{a(s)}$ is the associated vector of parameters, $\alpha_{a(s)}$ is the factor loading, and $e_{a(s)}$ represents an

idiosyncratic component assumed to be independent of θ , and X_a . The individual components $\{e_{a(s)}\}_{a=1}^{\bar{A}}$ for any s are mutually independent. Finally, the observed experience level $A^*(s^*)$ (where s^* represents the schooling level observed in the data) is interpreted as

$$A^*(s^*) = \arg \max_{a(s^*) \in \{1, \dots, \bar{A}\}} \{I_{a(s)}\}.$$

IV.3. The Model for Hourly Wages and Monthly Hours Worked

For hourly wages and monthly hours worked, we consider schooling/experience specific models. Consider first the model for wages. Denote by s and $a(s)$ the schooling and experience level attained by the individual. Wages ($Y_{a(s)}$) are modeled using a linear specification:

$$\ln Y_{a(s)} = \varphi_{Y,a(s)} \text{Gender} + \beta_{Y,a(s)} X_Y + \alpha_{Y,a(s)} \theta + e_{Y,a(s)} \quad \text{for } s = 1, \dots, \bar{S} \text{ and } a(s) = 1, \dots, \bar{A}$$

where $\varphi_{a(s)}$ is the gender gap, X_Y is a vector of observed controls, $\beta_{Y,a(s)}$ is the vector of coefficients, $\alpha_{a(s)}$ is the coefficient associated with the latent ability, and $e_{Y,a(s)}$ represents an idiosyncratic error term such that $e_{Y,a(s)} \perp (\theta, X_Y)$ for any $a(s) (= 1, \dots, \bar{A})$ and $s (= 1, \dots, \bar{S})$.

A parallel strategy is used to model hours worked. Let $H_{a(s)}$ denote the monthly hours worked given schooling level s and experience level $a(s)$. Thus, we assume

$$\ln H_{a(s)} = \varphi_{H,a(s)} \text{Gender} + \beta_{H,a(s)} X_H + \alpha_{H,a(s)} \theta + e_{H,a(s)} \quad \text{for } s = 1, \dots, \bar{S} \text{ and } a(s) = 1, \dots, \bar{A}$$

where $\varphi_{H,a(s)}$ is the gender gap, X_H is a vector of observed controls, $\beta_{H,a(s)}$ is the vector of coefficients associated with X_H , $\alpha_{H,a(s)}$ is the parameters associated with the latent ability, and $e_{H,a(s)}$ represents an idiosyncratic error term such that $e_{H,a(s)} \perp (\theta, X_H)$ for any $a(s) (= 1, \dots, \bar{A})$ and $s (= 1, \dots, \bar{S})$.

IV.4. The Model for Employment

Let $I_{E,a(s)}$ denote the net benefit associated with the alternative of having an employment (versus the alternatives of unemployment or out of the labor force) given the schooling level s and the accumulated experience $a(s)$. As in the previous cases, we assume a linear-in-the-parameters specification for $I_{E,a(s)}$:

$$I_{E,a(s)} = \varphi_{E,a(s)} \text{Gender} + \beta_{E,a(s)} X_E + \alpha_{E,a(s)} \theta + e_{E,a(s)} \quad \text{for } s = 1, \dots, \bar{S}, \quad a(s) = 1, \dots, \bar{A} \quad (5)$$

where $\varphi_{E,a(s)}$, $\beta_{E,a(s)}$, X_E , $\alpha_{E,a(s)}$, and $e_{E,a(s)}$ are defined as before. Finally, the error term is such that $e_{E,a(s)} \perp (\theta, X_E)$ for any $a(s) (= 1, \dots, \bar{A})$ and $s (= 1, \dots, \bar{S})$.

We use (5) to model the employment decisions observed in the data. Specifically, if we let $D_{E,a(s)}$ denote a binary variable such that is equal to 1 if the individual is employed and 0 otherwise, we estimate a binary model assuming that $D_{E,a(s)} = 1[I_{E,a(s)} > 0]$ where $1[\cdot]$ is (again) the indicator function.

IV.5. Schooling Performance: The Measurement System

The identification of the model can be established using the arguments developed in Carneiro, Hansen, and Heckman (2003) and Hansen, Heckman, and Mullen (2004). The identification strategy assumes the existence of a set of measurements. As explained in the next section, these measurements are associated to the individual's schooling performance.

Let T_i ($i=1, \dots, n_C$) denote the i -th measure. We distinguish the unobserved ability from the observed measure T_i . This is important since T_i is likely to depend on the characteristics of school as well as on the family background of the individuals by the time of the test. Thus, if X_T denote these characteristics, we have

$$T_i = \beta_{T_i} X_T + \alpha_{T_i} \theta + e_{T_i} \quad \text{for } i = 1, \dots, n_C$$

where $e_{T_i} \perp (\theta, X_T)$ and $e_{T_i} \perp e_{T_j}$ for any $i, j \in \{1, \dots, n_C\}$ such that $i \neq j$.

Since there are no intrinsic units for the latent ability, we need to normalize one of the loadings in the system to unity to set the scale of the latent ability. Therefore, for some T_i ($i=1, \dots, n_C$), we set $\alpha_{T_i} = 1$.

Notice that our assumptions imply that conditional on observables (variables contained in X), the dependence across all measurements, choices and outcomes come through the unobserved heterogeneity (θ). Notice that if θ were observed, we could use a matching type of approach to control for this dependence (selection). Instead, we estimate the distribution of the unobserved ability and then control for the dependence. Finally, we assume that θ measures the same thing for males and females.

IV.6. Implementing the Model

The model with unobserved heterogeneity has the following ingredients: the schooling decision problem, the linear models for hourly wages and monthly hours worked (by schooling level s and experience level $a(s)$), the models for employment (by schooling level s and experience level $a(s)$), the model for accumulated experience (by schooling level), and finally, the system of measurements or school achievement. Unobserved heterogeneity appears as determinant of each of these components. In this paper we assume that θ is distributed according to a two-component mixture of normals. Formally,

$$\theta \sim p_1 N(\mu_1, \Sigma_1^2) + (1 - p_1) N(\mu_2, \Sigma_2^2).$$

with this assumption we allow a flexible functional form for the unobserved heterogeneity.

Following the empirical strategy utilized in Section III, we estimate the schooling choice model and the experience models using multinomial probit models. Then, we implicitly assume that the idiosyncratic shocks in the equations describing the net utilities are assumed to be jointly normally distributed. The four schooling levels study used here are: primary school, secondary school (or high school), some tertiary education (or some college graduates), and complete tertiary education (or college graduates). For accumulated experience we use three categories: less than 10 years of experience, between 10 and 15 years of experience, and more than 15 years of experience.

In estimating the model, and since there is no sequential decision process, we use the schooling and experience level reported at the time of the interview.¹⁰

¹⁰ In the case of experience, we use the retrospective information provided by the respondent (labor history). The labor history is reported from age 15 or since 1980 depending on the year the individual became 15 years old. For details see Section II.

For the models of wages and hours worked we use the information for the month previous to the interview. The same applies in the case of employment status. This is consistent with what we use in Section III.

The measurement system uses the following variables: Average Grade during Secondary Education, Repeated Grade during Primary Education and Repeated Grade during Secondary Education.

We normalize the mean of the factor to zero, and we normalize the loading to be equal to one in the equation for the Average Grade during Secondary Education

Tables 8A and 8B display the variables used in the empirical implementation of the model, as well as the normalization assuring the identification of the model. The model is estimated using Markov Chain Monte Carlo Methods (MCMC). See Appendix A for a formal discussion of the method used in this paper.

V. Main Results

Table 9 presents the gender gaps in hourly wages obtained from the model with unobserved heterogeneity. The estimated gaps are in general sizeable and statistically significant. We do not observe clear patterns either by schooling and/or experience levels, although we consistently estimated the largest gender gaps among college graduates (regardless of the experience level considered). In this group we estimate that males make between 36% and 38% more per hour than females. These numbers are larger than those presented in Section III. But Table 9 also presents a range for the gender gap in wages which goes from -6% (but non significant) for high school dropouts reporting less than 10 years of experience to 38% for college graduates with between 10 and 15 years of experience. Importantly, in only two cases we estimate a gender gap below 15%. Therefore, our evidence indicates the existence of wage differentials that cannot be explained by observed or unobserved characteristics.

As in the case of wages, the results obtained for hours worked show a range of values for the gender gaps. These are presented in Table 10. We observe that the point estimates are between -6% (high school dropouts with less than 10 years of experience) and 18% (high school dropouts with between 10 and 15 years of experience). In this case however, less than a half of the estimates are statistically significant. For example, among high school graduates and college graduates we do not find significant gender differences. This is consistent with the evidence presented in Section III, although the numbers in Table 10 show a broader picture of the gender gaps (if any) in hours worked.

Table 11 presents the results for employment. Two are the main results here. First, we observe, in general, a reduction in the estimated gap when we move from low to high experience levels (the only exception is observed among high school graduates). Second, the results suggest that schooling also helps to reduce the estimated gaps (there are only two exceptions in Table 11). In fact, among college graduates the estimated coefficients are -0.12 and -0.23 for experience levels “between 10 and 15 years” and “more than 15 years”, respectively,^{11,12} so the gap favors females in this case. However, as in the case of hours worked, only few estimates are statistically significant, and when significant, they are usually associated with low schooling and experience levels.

Table 12 presents the results obtained for the four multinomial choice models used to study accumulated experience. The evidence in Table 12 shows how the gender gap reduces with schooling. Specifically, the significant gender differences estimated for high school dropouts and high school graduates are 100% larger than the ones obtained among individuals with some college. Interestingly, among college graduates we do not find significant differences between genders.

Our analysis of gender gaps in variables associated with the labor market leads us to conclude that (1) there are differences between males and females that cannot be explained with observable or unobservable characteristics, and that, in general, (2) these differences are larger among individuals reporting low schooling level and they almost vanish among the more educated individuals.¹³

The model also allows us to analyze the gender differences in schooling attainment and schooling achievement. It is worth recalling that the evidence presented in Section III already suggested that females outperform males in these two dimensions (see Tables 6 and 7). Table 13 and 14 repeat that analysis but now incorporating unobserved heterogeneity (latent ability).

Table 13 presents the gender gaps in schooling decisions. The results show (again) that females are more likely than males to reach higher schooling levels. When compared with those in Table 6, we see that the effects are now larger. Something similar occurs in the case of “repeating a grade during primary school”, “repeating a grade during secondary school”, and “average grades during high school”. The results are shown in Table 14. The evidence in this table suggests that females outperform males, that the differences are statistically significant and that they are larger than the ones presented in Table 7. Specifically, when comparing the estimated gender gaps across tables we obtain 18% (0.26 versus 0.22) and 41% (0.17 versus 0.12) increments in the gender coefficient associated with “repeating a grade in primary school” and “repeating a grade in secondary school”, respectively.

¹¹ For the group of individuals reporting more than 15 years of experience and a college degree, the gender dummy perfectly predicts the labor status: the 29 women in this category reported a job during the week previous to the interview.

¹² These coefficients are the point estimates of the parameters associated with the gender dummy variable, so they need to be interpreted cautiously since they do not represent the marginal effects.

¹³ The only exception to this point, and an important one, comes from the analysis of hourly wages.

In the case of “average grade during secondary school” we obtain an increment of 6.4% in the gender gap (0.33 versus 0.31).

VI. Can Unobserved Heterogeneity Explain the Gender Gaps in the Labor Market?

From the evidence presented in this paper we must conclude that this is still an open question. Our results do indicate that, after controlling for unobserved heterogeneity, there are non significant gender differences in a variety of labor market variables among educated individuals (e.g., hours worked, accumulated experience, employment), but we still find gender differences among the other schooling groups. These differences can in principle be interpreted as “pure” discrimination. However, this interpretation requires several qualifications.

First, our empirical strategy assumes that a one dimensional model of unobserved heterogeneity is sufficient to capture and control for selection (endogeneity) across different margins (decisions). Nevertheless, previous studies have shown the existence of at least two underlying sources of unobserved heterogeneity when explaining labor market outcomes and social behavior.¹⁴ In this context, our one-dimensional model might be only partially capturing the unobserved heterogeneity in the data. The consequences of incorporating additional sources of essential heterogeneity on our results are hard to predict. Thus, in principle, we cannot discard the possibility that what we interpret as “unexplained gaps” in the one dimensional case, can be in fact “explained” by, for example, heterogeneity in other unobserved traits (e.g. noncognitive abilities) or preferences (e.g. preferences for leisure).¹⁵

Second, and following up on the previous point, it is interesting to notice that in our results the coefficients associated with what we identify as unobserved heterogeneity are not always significant. The strongest effect of unobserved heterogeneity are obtained for the schooling variables (Tables 13 and 14), and accumulated experience (Table 12). Although the effects are sizeable for the other outcomes, they are usually non-statistically significant. This suggests that our source of unobserved

¹⁴ See Heckman, Stixrud, and Urzua, 2006; Urzua, 2006; Cunha, Heckman and Navarro, 2004.

¹⁵ It is worth noting that the assumption of a single source of unobserved heterogeneity can be relaxed depending on the availability of more comprehensive information at the individual level. These needs for better and more comprehensive information come from the identification argument of the models. Recall that the source of unobserved heterogeneity in this paper is identified using the schooling achievement variable. In order to identify additional sources of heterogeneity we would need additional variables in the measurement system. The availability of information on personality traits, IQ tests, or time preferences could allow the identification and estimation of more general models of unobserved heterogeneity.

heterogeneity is more closely related to scholastic ability¹⁶ which apparently is not significantly *valued* in the Chilean labor market after schooling and experience levels are taken into account. Nevertheless, there might be other sources of unobserved heterogeneity that are in fact *priced* in the labor market. This again illustrates the potential benefits of extending the model to multiple dimensions of unobserved heterogeneity

A different consideration regarding the robustness and interpretation of our results can be made by noticing that we implement the model by assuming the existence of a single distribution of unobserved heterogeneity in the sample. The consequences of allowing gender-specific distributions on our previous results are (again) hard to predict, but we believe that the complications of such extension would most likely dominate any potential new insights. This since the identification of gender-specific distribution has additional complications and it relies on even stronger assumptions than the one already made.¹⁷ Besides, from an intuitive point of view, we do not find a priori deep reasons to believe that there are gender differences in the distributions of unobserved heterogeneity. It is because of this remarks that the estimation of gender-specific distribution is left for future research.

¹⁶ Before we interpret the source unobserved heterogeneity as a combination of cognitive and noncognitive abilities.

¹⁷ Specifically, even though we can assure the identification of gender-specific variance/covariance matrices, the identification of gender-specific mean differences in the distribution of unobserved heterogeneity would require the existence of at least one discrimination free variable. The selection and existence of such variable(s) is non trivial is arguably as well. See Urzua (2006) for details.

VII. Conclusions

In this paper we present a comprehensive analysis of the gender gaps in a variety of labor market outcomes for Chile. The analysis is carried out using two different approaches. The first approach follows the literature by estimating linear and nonlinear models of a variety of variables on different observable controls and the gender dummy. This approach does not pay attention to potential selection problems (endogeneity). The second approach is more general. It allows for the presence of individuals' unobserved heterogeneity that is assumed to be the cause of the endogeneity problems in the conventional approach.

Our main results are robust across the approaches. They suggest the existence of gender gaps in labor market variables that cannot be explained by observable or unobservable characteristics or by underlying selection mechanisms generating endogeneity. Nevertheless, the findings from the model with unobserved heterogeneity indicate that the gender gaps critically depend on the schooling level of the individuals considered in the analysis. This is particular important among college graduates. For this group, the gender differences are in general non-significant.

The evidence also demonstrates that females outperform males in schooling achievement and schooling performance. This is observed regardless of the approach, but we find the stronger effects in the model with unobserved heterogeneity. These gender differences favoring women represent an argument against the conventional idea that labor market differences can be interpreted as the result of differences in the human capital between genders. Obviously, this conclusion assumes that the utilized variables are good proxies of the actual human capital accumulated by the individuals.

Overall, the estimates in this paper could lead us to conclude that women are effectively discriminated in the labor market with the largest gender gaps observed among the less educated people. However, we prefer to interpret our results cautiously. We believe that the availability of better data and the estimation of even more general models than the one considered here could indeed explain some of the unexplained estimated gender gaps.

VIII. References

Bravo, D. (2004), *Análisis y principales resultados. Primera Encuesta de Protección Social*. Subsecretaría de Previsión Social, Ministerio del Trabajo y Previsión Social y Centro de Microdatos, Departamento de Economía, Universidad de Chile. Julio. Disponible en www.proteccionsocial.cl.

Bravo, D., J.Behrman, O.Mitchell and P.Todd (2006), *Encuesta de Protección Social 2004: Presentación General y Principales Resultados*. Subsecretaría de Previsión Social, Ministerio del Trabajo y Previsión Social y Centro de Microdatos, Departamento de Economía, Universidad de Chile. Enero. Disponible en www.proteccionsocial.cl

Carneiro, P., K. Hansen, and J. Heckman (2003). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review* 44 (2), 361–422. 2001 Lawrence R. Klein Lecture.

Cunha, F., J. Heckman, and S. Navarro. Separating Uncertainty from Heterogeneity in Life Cycle Earnings. The 2004 Hicks Lecture, Oxford Economic Papers 57(191-261), 1-72

Hansen, K., J. Heckman, and K. Mullen (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics* 121 (1-2), 39–98.

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica* 42(4), 679.

Heckman, J. (1981). Statistical models for discrete panel data. In C. Manski and D. Mc-Fadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 114–178. Cambridge, MA: MIT Press.

Heckman, J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3).

Subsecretaría de Previsión Social, *Encuesta de Protección Social*, disponible en www.proteccionsocial.cl

Urzua, S. (2006). Racial Labor Market Gaps: The Role of Abilities and Schooling Choices. Working paper, University of Chicago.

6 Appendix A: The Computation of the Model

We use the Markov Chain Monte Carlo (MCMC) method to estimate the model introduced in Section 3. This appendix illustrates the algorithm using a simplified version of the model. The model used here is a two-sector Roy economy. Formally, let I denote the net utility associated with the choice problem of selecting between two sectors. Let Y_0 and Y_1 denote the associated potential outcomes. Thus,

$$\begin{aligned} Y_1 &= X_1\beta_1 + \varepsilon_1 \\ Y_0 &= X_0\beta_0 + \varepsilon_0 \\ D &= 1(I > 0) \end{aligned}$$

with $I = Z\gamma + V$. In addition, assume that

$$\begin{aligned} V &= \theta\alpha_V + U_V \\ \varepsilon_1 &= \theta\alpha_1 + U_1 \\ \varepsilon_0 &= \theta\alpha_0 + U_0 \end{aligned}$$

$U_1 \perp\!\!\!\perp U_0 \perp\!\!\!\perp U_V$, and $f \perp\!\!\!\perp (U_0, U_1, U_V)$, and

$$\begin{aligned} U_1 &\sim N(0, \sigma_{U_1}^2) \\ U_0 &\sim N(0, \sigma_{U_0}^2) \\ U_V &\sim N(0, \sigma_{U_V}^2) \end{aligned}$$

Following the identification strategy discussed in Section 3.4, we assume the existence of a set of measurements

$$T_h = Q_h\eta_h + \theta\delta_h + U_{T_h}$$

where $U_{T_h} \sim N(0, \sigma_{U_{T_h}}^2)$ with $h = 1, \dots, H$.

Finally, the unobserved factor is assumed to be distributed according to a mixture of K -normal distributions, i.e.

$$\theta \sim MN\left(\{\mu_k\}_{k=1}^K, \{\Sigma_{fk}\}_{k=1}^K, \{p_k\}_{k=1}^K\right)$$

such that $E(\theta) = 0$.

The likelihood function of the model is given by:

$$\Gamma(Y, T, D; \Theta) = \prod_{j=1}^N \int \left[\prod_{h=1}^H \frac{1}{\sqrt{2\pi}\sigma_{U_{T_h}}} \exp\left(-\frac{1}{2\sigma_{U_{T_h}}^2}(T_{h,j} - Q_{h,j}\eta_h - \theta_j\delta_h)^2\right) \right] \left[\frac{1}{\sqrt{2\pi}\sigma_{U_1}} \exp\left(-\frac{1}{2\sigma_{U_1}^2}(Y_{1,j} - X_j\beta_1 - \theta_j\alpha_1)^2\right) \Phi(-Z_j\gamma - \theta_j\alpha_V) \right]^{D_i} \left[\frac{1}{\sqrt{2\pi}\sigma_{U_0}} \exp\left(-\frac{1}{2\sigma_{U_0}^2}(Y_{0,j} - X_j\beta_0 - \theta_j\alpha_0)^2\right) (1 - \Phi(-Z_j\gamma - \theta_j\alpha_V)) \right]^{1-D_i} dF(\theta_j).$$

6.1 The Gibbs Sampler

We start by specifying the priors. For the parameters in the models for wages $(\beta_j, \alpha_j, 1/\sigma_{U_j}^2)$:

$$\begin{aligned} \beta_j &\sim N(\beta_j^0, \sigma_{\beta_j}^0 I) \\ \alpha_j &\sim N(\alpha_j^0, \sigma_{\alpha_j}^0 I) \\ 1/\sigma_{U_j}^2 &\sim \text{Gamma}(a_{U_j}, b_{U_j}). \end{aligned}$$

For the parameters in the models for test scores $(\eta_h, \delta_h, 1/\sigma_{U_{T_h}}^2)$:

$$\begin{aligned} \eta_h &\sim N(\eta_h^0, \sigma_{\eta_h}^0 I) \\ \delta_h &\sim N(\delta_h^0, \sigma_{\delta_h}^0 I) \\ 1/\sigma_{U_{T_h}}^2 &\sim \text{Gamma}(a_{U_{T_h}}, b_{U_{T_h}}). \end{aligned}$$

For the parameters in the model for the latent variable (γ, α_V) :

$$\begin{aligned} \gamma &\sim N(\gamma^0, \sigma_{\gamma}^0 I) \\ \alpha_V &\sim N(\alpha_V^0, \sigma_{\alpha_V}^0 I). \end{aligned}$$

For the parameters in the factor's distribution $(\{\mu_k\}_{k=1}^K, \{\Sigma_{\theta,k}\}_{k=1}^K, \{p_k\}_{k=1}^K)$:

$$\begin{aligned} p &\sim \text{Dirichlet}(a_1, \dots, a_K) \\ \Sigma_{\theta,k} &\sim \text{Gamma}(v_k^0, V_k^0) \\ \mu_k &\sim N(\mu_k^0, \sigma_{\mu_k}^0). \end{aligned}$$

Notice that in what follows we denote $\Gamma_{\theta,l} = \Sigma_{\theta,l}^{-1}$, i.e $\Gamma_{\theta,l}$ denotes the precision for the l component.

Let $\lambda = (\alpha, \beta, \tau, \gamma, \eta, \theta, \Sigma)$. Using the block structure of the model we can derive the formulae for the conditional posteriors:

For wages:

$$f(\beta_j/\alpha_j, \tau_{U_j}, \theta, Y, D) \propto \exp \left\{ -\frac{\tau_{U_j}}{2} \left(\sum_{i:D=j} (Y_i - X_i\beta_j - \alpha_j\theta_i)^2 \right) - \frac{\sigma_{\beta_j}^0}{2} (\beta_j - \beta_j^0)' (\beta_j - \beta_j^0) \right\} \text{ for } i = 0, 1$$

consequently,

$$\beta_j \sim N \left(\left(\tau_{U_j} \sum_{i:D=j} X_i X_i' + \tau_{\beta_j}^0 I \right)^{-1} \left(\tau_{U_j} \sum_{i:D=j} X_i (Y_i - \alpha_j \theta_i) + \tau_{\beta_j}^0 I \beta_j^0 \right), \left(\tau_{U_j} \sum_{i:D=j} X_i X_i' + \tau_{\beta_j}^0 I \right)^{-1} \right) \quad (4)$$

Similarly for α_j ,

$$\alpha_j \sim N \left(\left(\tau_{U_j} \sum_{i:D=j} \theta_i^2 + \tau_{\alpha_j}^0 \right)^{-1} \left(\tau_{U_j} \sum_{i:D=j} \theta_i (Y_i - X_i \beta_j) + \tau_{\alpha_j}^0 \alpha_j^0 \right), \left(\tau_{U_j} \sum_{i:D=j} \theta_i^2 + \tau_{\alpha_j}^0 \right)^{-1} \right). \quad (5)$$

For test scores:

Following the same logic as in the previous case, it is straightforward to obtain that:

$$\eta_h \sim N \left(\left(\tau_{U_{T_h}} \sum_{i=1}^N Q_i Q_i' + \tau_{\eta_h}^0 I \right)^{-1} \left(\tau_{U_{T_h}} \sum_{i=1}^N Q_i (T_{h,i} - \delta_h \theta_i) + \tau_{\eta_h}^0 I \eta_h^0 \right), \left(\tau_{U_{T_h}} \sum_{i=1}^N Q_i Q_i' + \tau_{\eta_h}^0 I \right)^{-1} \right) \quad (6)$$

and

$$\delta_h \sim N \left(\left(\tau_{U_{T_h}} \sum_{i=1}^N \theta_i^2 + \tau_{\delta_h}^0 \right)^{-1} \left(\tau_{U_{T_h}} \sum_{i=1}^N \theta_i (T_{h,i} - \eta_h Q_i) + \tau_{\delta_h}^0 \delta_h^0 \right), \left(\tau_{U_{T_h}} \sum_{i=1}^N \theta_i^2 + \tau_{\delta_h}^0 \right)^{-1} \right). \quad (7)$$

For probit:

We need a completion for D . Let D^* be the latent variable. Thus,

$$\begin{aligned}
f(\gamma, \alpha_V, D^*/D) &= f(D/\alpha_V, \gamma, D^*)f(D^*/\alpha_V, \gamma)f(\gamma)f(\alpha_V) \\
&= \left[\left(\frac{1}{\sigma_{\alpha_V}^0} \right)^{Dim(\alpha_V)/2} \exp \left(-\frac{\sigma_{\alpha_V}^0}{2} (\alpha_V - \alpha_V^0) (\alpha_V - \alpha_V^0)' \right) \right] \\
&\quad \times \left[\left(\frac{1}{\sigma_{\gamma}^0} \right)^{Dim(\gamma)/2} \exp \left(-\frac{\sigma_{\gamma}^0}{2} (\gamma - \gamma^0) (\gamma - \gamma^0)' \right) \right] \\
&\quad \times \prod_{i=1}^N \left\{ [1(D_i^* > 0)\phi(D_i^* - Z_i\gamma - \alpha_V\theta_i)]^{D_i} \right. \\
&\quad \quad \left. \times [1(D_i^* < 0)\phi(D_i^* - Z_i\gamma - \alpha_V\theta_i)]^{1-D_i} \right\}
\end{aligned}$$

Now,

$$\begin{aligned}
f(\gamma/\alpha_V, \theta, D^*, D, Z) &\propto \prod_{i=1}^N \phi(D_i^* - z_i\gamma - \alpha_V\theta_i) \\
&\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^N (D_i^* - z_i\gamma - \alpha_V\theta_i)^2 - \frac{\sigma_{\gamma}^0}{2} (\gamma - \gamma^0) (\gamma - \gamma^0)' \right\}
\end{aligned}$$

Consequently,

$$\gamma \sim N \left((z'z + \tau_{\gamma}^0 I)^{-1} (z' (D^* - \alpha_V\theta) + \tau_{\gamma}^0 I \gamma^0), (z'z + \tau_{\gamma}^0 I)^{-1} \right) \quad (8)$$

Using the same logic as in (5) we obtain:

$$\alpha_V \sim N \left(\left(\sum_{i=1}^N \theta_i^2 + \tau_{\alpha_V}^0 \right)^{-1} (Z' (D^* - \alpha_V\theta) + \tau_{\alpha_V}^0 \alpha_V^0), \left(\sum_{i=1}^N \theta_i^2 + \tau_{\alpha_V}^0 \right)^{-1} \right). \quad (9)$$

Finally,

$$f(D_i^*/\alpha_V, \theta, \gamma, D) \propto [1(D_i^* > 0)\phi(D_i^* - Z_i\gamma - \alpha_V\theta_i)]^{D_i} [1(D_i^* < 0)\phi(D_i^* - Z_i\gamma - \alpha_V\theta_i)]^{1-D_i}$$

Therefore, we sample D_i^* from:

$$D_i^* = \begin{cases} TN_{[0, \infty)}(Z_i\gamma + \alpha_V\theta_i, 1) & \text{if } D_i = 1 \\ TN_{(-\infty, 0]}(Z_i\gamma + \alpha_V\theta_i, 1) & \text{if } D_i = 0 \end{cases} \quad (10)$$

Precisions:

The posterior distributions for the precisions are:

$$\tau_{U_j} \sim G \left(\frac{\sum_{i:D=j} 1}{2} + a_{U_j}, \left(\frac{1}{2} \sum_{i:D=j} (Y_i - X_i \beta_j - \alpha_j \theta_i)^2 \right) + b_{U_j} \right) \quad \text{for } i = 0, 1 \quad (11)$$

$$\tau_{U_{T_h}} \sim G \left(\frac{N}{2} + a_{U_{T_h}}, \left(\frac{1}{2} \sum_{j=1}^N (T_{h,i} - Q_i \eta_h - \delta_h \theta_i)^2 \right) + b_{U_{T_h}} \right) \quad \text{for } h = 1, \dots, H. \quad (12)$$

Factor's Distribution:

Since we restrict the mean to be zero: $\mu_L = -\sum_{l=1}^{L-1} \frac{p_l}{p_L} \mu_l$, we have that

$$\begin{aligned} & \exp \left(-\frac{\tau_\mu^0}{2} (\mu_1 - \mu^0) (\mu_1 - \mu^0)' - \frac{1}{2} \sum_{i,1} (\theta_i - \mu_1) \Gamma_{\theta,1} (\theta_i - \mu_1)' \right) \\ & \quad \vdots \\ & \times \exp \left(-\frac{\tau_\mu^0}{2} \left(-\sum_{l=1}^{L-1} \frac{p_l}{p_L} \mu_l - \mu^0 \right) \left(-\sum_{l=1}^{L-1} \frac{p_l}{p_L} \mu_l - \mu^0 \right)' - \frac{1}{2} \sum_{i,L} \left(\theta_i + \sum_{l=1}^{L-1} \frac{p_l}{p_L} \mu_l \right) \Gamma_{\theta,L} \left(\theta_i + \sum_{l=1}^{L-1} \frac{p_l}{p_L} \mu_l \right)' \right) \\ & \quad = \exp \left(-\frac{\tau_\mu^0}{2} (\mu_1 - \mu^0) (\mu_1 - \mu^0)' - \frac{1}{2} \sum_{i,1} (\theta_i - \mu_1) \Gamma_{f,1} (\theta_i - \mu_1)' \right) \\ & \quad \quad \quad \vdots \\ & \times \exp \left(\begin{array}{l} -\frac{\tau_\mu^0}{2} \left(\frac{p_1}{p_L} \right)^2 \left(\mu_1 - \left(-\sum_{l=2}^{L-1} \frac{p_l}{p_1} \mu_l - \frac{p_L}{p_1} \mu^0 \right) \right) \left(\mu_1 - \left(-\sum_{l=2}^{L-1} \frac{p_l}{p_1} \mu_l - \frac{p_L}{p_1} \mu^0 \right) \right)' \\ -\frac{1}{2} \left(\frac{p_1}{p_L} \right)^2 \sum_{i,L} \left(\mu_1 - \left(-\frac{p_L}{p_1} \theta_{L,i} - \sum_{l=2}^{L-1} \frac{p_l}{p_1} \mu_l \right) \right) \Gamma_{\theta,L} \left(\mu_1 - \left(-\frac{p_L}{p_1} \theta_{L,i} - \sum_{l=2}^{L-1} \frac{p_l}{p_1} \mu_l \right) \right)' \end{array} \right) \end{aligned}$$

thus,

$$\mu_1 \sim N \left(\left(\tau_\mu^0 + N_1 \Gamma_{\theta,1} + \left(\frac{p_1}{p_L} \right)^2 (\tau_\mu^0 + N_L \Gamma_{\theta,L}) \right)^{-1} \begin{pmatrix} \tau_\mu^0 \mu^0 + \Gamma_{\theta,1} \sum_{i=1,1} \theta_i \\ + \left(\frac{p_1}{p_L} \right)^2 \tau_\mu^0 \left(- \sum_{l=2}^{L-1} \frac{p_l}{p_1} \mu_l - \frac{p_L}{p_1} \mu^0 \right) \\ + \left(\frac{p_1}{p_L} \right)^2 \Gamma_{\theta,L} \sum_{i=1,L} \left(- \frac{p_L}{p_1} \theta_i - \sum_{l=2}^{L-1} \frac{p_l}{p_1} \mu_l \right) \end{pmatrix}, \left(\tau_\mu^0 + \Gamma_{\theta,1} + \tau_\mu^0 \left(\frac{p_1}{p_L} \right)^2 + \left(\frac{p_1}{p_L} \right)^2 \Gamma_{\theta,L} \right)^{-1} \right)$$

For the probabilities,

$$\Pr(p) \propto \frac{1}{B(\alpha)} \prod_{l=1}^L p_l^{\alpha_l + N_l - 1}$$

so,

$$p \sim D(\alpha_1 + N_1, \dots, \alpha_L + N_L).$$

For the group indicator

$$z_i \sim MN \left(\frac{p_1 \phi(\theta_i, \mu_1, \Gamma_{\theta,1})}{\sum_l p_l \phi(\theta_i, \mu_l, \Gamma_{\theta,l})}, \dots, \frac{p_L \phi(\theta_i, \mu_L, \Gamma_{\theta,L})}{\sum_l p_l \phi(\theta_i, \mu_l, \Gamma_{\theta,l})} \right)$$

Following the same argument as in the case of the other precisions, we have that the posterior distribution for the precisions $\Gamma_{\theta,l}$ follows a Gamma distribution.

Finally the *posterior for the factor*:

$$\begin{aligned} \Pr(\theta_i / \lambda, Y_i, T_i, D_i = 1) &\propto \exp \left(- \frac{\tau_{U_1}}{2} (Y_i^1 - X_i \beta_1 - \theta_i \alpha_1)^2 \right) \\ &\times \exp \left(- \frac{1}{2} (D_i^* - Z_i \gamma - \alpha_V \theta_i)^2 \right) \\ &\times \exp \left(- \sum_{h=1}^H \frac{\tau_{U_{T_h}}}{2} (T_{i,h} - Q_{i,h} \eta_h - \theta_i \delta_h)^2 \right) \\ &\times \exp \left(- \frac{1}{2} \theta_i^2 \Gamma_\theta \right) \\ &= \exp \left(- \frac{\tau_{U_1}}{2} (Y_i^1 - X_i \beta_1 - \theta_i \alpha_1)^2 - \frac{1}{2} (D_i^* - Z_i \gamma - \alpha_V \theta_i)^2 \right. \\ &\quad \left. - \sum_{h=1}^H \left(\frac{\tau_{U_{T_h}}}{2} (T_{i,h} - Q_{i,h} \eta_h - \theta_i \delta_h)^2 \right) - \frac{1}{2} \theta_i^2 \Gamma_\theta \right) \end{aligned}$$

To simplify notation let $\theta_i^* = Y_i - X_i\beta_j$, $\theta_{i,k}^* = T_{i,k} - Q_{i,k}\theta_k$, $\theta_i^{***} = D_i^* - Z_i\gamma$. Then

$$\Pr(\theta_i/\lambda, Y_i, T_i, D_i = 1) \propto \exp \left(\begin{array}{c} -\frac{\tau_{U_1}}{2}(\theta_i^* - \theta_i\alpha_1)^2 - \frac{1}{2}(\theta_i^{***} - \alpha_V\theta_i)^2 \\ -\sum_{k=1}^K \left(\frac{\tau_{U_{T_k}}}{2}(\theta_{i,k}^* - \theta_i\delta_k)^2 \right) - \frac{1}{2}\theta_i^2\Gamma_\theta \end{array} \right).$$

Notice that

$$\exp \left(-\frac{\tau_{U_1}}{2}(\theta_i^* - \theta_i\alpha_1)^2 - \frac{1}{2}(\theta_i^{***} - \alpha_V\theta_i)^2 \right) \propto \exp \left(\begin{array}{c} -\frac{\tau_{U_1}}{2}(\theta_i - \bar{\theta}_i^*)' \alpha_1' \alpha_1 (\theta_i - \bar{\theta}_i^*) \\ -\frac{1}{2}(\theta_i - \bar{\theta}_i^{**})' \alpha_V' \alpha_V (\theta_i - \bar{\theta}_i^{**}) \end{array} \right)$$

where $\bar{\theta}_i^* = (\alpha_1'\alpha_1)^{-1}(\alpha_1\theta_i^*)$ and $\bar{\theta}_i^{**} = (\alpha_V'\alpha_V)^{-1}(\alpha_V\theta_i^{**})$. Then,

$$\exp \left(-\frac{\tau_{U_1}}{2}(\theta_i^* - \theta_i\alpha_1)^2 - \frac{1}{2}(\theta_i^{***} - \alpha_V\theta_i)^2 \right) \propto \exp \left(-\frac{1}{2}(\theta_i - \bar{\theta}_i)'\tilde{\Sigma}(\theta_i - \bar{\theta}_i) \right)$$

where $\bar{\theta}_i = (\tau_{U_1}\alpha_1'\alpha_1 + \alpha_V'\alpha_V)^{-1}(\tau_{U_1}\alpha_1'\alpha_1\bar{\theta}_i^* + \alpha_V'\alpha_V\bar{\theta}_i^{**})$ and $\tilde{\Sigma} = (\tau_{U_1}\alpha_1'\alpha_1 + \alpha_V'\alpha_V)$.

Repeating the same argument multiple times we get,

$$\Pr(\theta_i/\lambda, Y_i, T_i, D_i = 1) \sim N(\bar{\mu}_{\theta,i}, \bar{\Sigma}_\theta) \quad (13)$$

with

$$\bar{\mu}_{\theta,i} = \left(\tau_{U_1}\alpha_1'\alpha_1 + \alpha_V'\alpha_V + \sum_{h=1}^H \tau_{U_{T_h}}\delta_h'\delta_h + \Gamma_\theta \right)^{-1} \left(\tau_{U_1}\alpha_1'\alpha_1\bar{\theta}_i^* + \alpha_V'\alpha_V\bar{\theta}_i^{**} + \sum_{h=1}^H \tau_{U_{T_h}}\delta_h'\delta_h\bar{\theta}_{h,i}^* \right)$$

and

$$\bar{\Sigma}_\theta = \left(\tau_{U_1}\alpha_1'\alpha_1 + \alpha_V'\alpha_V + \sum_{h=1}^H \tau_{U_{T_h}}\delta_h'\delta_h + \Gamma_\theta \right)^{-1}$$

where $\bar{\theta}_{h,i}^* = (\delta_h'\delta_h)^{-1}(\delta_h^*\theta_{i,h})$.

Finally, the gibbs sampling procedure is:

1. Choose initial values for the parameters, and an arbitrary first draw for the factor. (I used $\theta^{(m)} \sim N(0,1)$)

For $m = 1, M$

1. Sample $D^*(j)^{(m)}$ for $j = 1, \dots, N$ according to (10)

2. Sample $\theta(j)^{(m)}$ for $j = 1, \dots, N$ according to (13)
3. Sample $\beta_i^{(m)}$ ($i = 1, 2$) according to (4)
4. Sample $\alpha_i^{(m)}$ ($i = 1, 2$) according to (5)
5. Sample $\gamma^{(m)}$ according to (8)
6. Sample $\alpha_V^{(m)}$ according to (9)
7. Sample $\tau_i^{(m)}$ ($i = 1, 2$) according to (11)
8. Sample $\tau_T^{(m)}$ according to (12)
9. Sample $\tau_\theta^{(m)}$ according to the respective Gama distribution.

We iterate over m until converge.

The model introduced in Section 3 is simply a more general version of this two-sectors Roy model.

Table 1. Means of Schooling and Labor Market Outcomes by Gender
SPS02

Variable (Dummy=1 if Apply)	Females		Males	
	Mean	Std. Dev	Mean	Std. Dev
A. School Information				
Maximum Schooling Level = Primary Education	0.11	0.32	0.17	0.38
Maximum Schooling Level = Secondary Education	0.51	0.50	0.49	0.50
Maximum Schooling Level = Some Tertiary Education	0.26	0.44	0.24	0.43
Maximum Schooling Level = Complete Tertiary Education	0.11	0.31	0.10	0.30
Repeat a Grade in Primary School	0.22	0.41	0.30	0.46
Repeat a Grade in Secondary School	0.20	0.40	0.24	0.43
Average Grade in Secondary School ^(a)	0.16	0.98	-0.17	1.00
B. Labor Market Variables				
Monthly Earnings	215,266	214,323	285,140	360,046
Hours Worked per Week	43.41	11.74	48.17	9.81
Hourly Wage	1,292	1,257	1,636	4,649
Working During Last Month	0.59	0.49	0.82	0.39
Less than 10 years of Experience	0.56	0.50	0.25	0.43
Between 10 and 15 years of Experience	0.26	0.44	0.34	0.47
More than 15 years of Experience	0.18	0.39	0.41	0.49
Number of Observations	1,765		1,801	

Note: The numbers presented in this table corresponds to the sample of individuals with ages between 28 and 40 years old at the time of the interview.

Table 2. The Gender Gap in Hourly Wages
SPS02

Variables	(A)	(B)	(C)	(D)
Male	0.24	0.23	0.23	0.29
	(0.03)	(0.03)	(0.03)	(0.03)
Schooling ^(a)				
Secondary Education	-	-	0.29	0.30
	-	-	(0.04)	(0.04)
Some Tertiary Education	-	-	0.49	0.50
	-	-	(0.04)	(0.05)
Complete Tertiary Education	-	-	0.90	0.92
	-	-	(0.06)	(0.06)
Experience ^(b)				
Between 10 and 15 years of Experience	-	0.04	0.05	0.14
	-	(0.03)	(0.03)	(0.03)
More than 10 years of Experience	-	0.04	0.10	0.19
	-	(0.03)	(0.03)	(0.04)
Residence ^(c)				
Central	-0.15	-0.15	-0.15	-0.15
	(0.04)	(0.04)	(0.04)	(0.04)
South	-0.04	-0.04	-0.05	-0.004
	(0.04)	(0.04)	(0.04)	(0.04)
Santiago	0.22	0.22	0.21	0.24
	(0.03)	(0.03)	(0.03)	(0.03)
Type of Job ^(d)				
Employer or Self-Worker	-0.13	-0.13	-0.10	-0.11
	(0.03)	(0.03)	(0.03)	(0.03)
Domestic Service	-0.08	-0.08	-0.04	-0.06
	(0.07)	(0.07)	(0.07)	(0.07)
Occupations ^(e)				
Professionals	0.09	0.10	-0.18	-0.17
	(0.07)	(0.07)	(0.07)	(0.07)
Technicians and associate professionals	-0.33	-0.33	-0.27	-0.25
	(0.07)	(0.07)	(0.07)	(0.07)
Clerks	-0.71	-0.72	-0.56	-0.53
	(0.07)	(0.07)	(0.06)	(0.06)
Service workers and shop and market sales workers	-1.08	-1.08	-0.84	-0.83
	(0.07)	(0.07)	(0.07)	(0.07)
Skilled agricultural and fishery workers	-1.35	-1.36	-0.96	-0.93
	(0.08)	(0.08)	(0.08)	(0.09)
Craft and related trades workers	-1.05	-1.05	-0.77	-0.74
	(0.06)	(0.06)	(0.06)	(0.06)
Plant and machine operators and assemblers	-1.11	-1.11	-0.85	-0.82
	(0.07)	(0.07)	(0.07)	(0.07)
Elementary occupations	-1.28	-1.28	-0.94	-0.91
	(0.07)	(0.07)	(0.07)	(0.07)
Constant	7.63	7.61	7.04	6.75
	(0.07)	(0.07)	(0.08)	(0.10)
Correction for Selection	No	No	No	Yes

Notes: (a) The baseline category is Primary Education; (b) The baseline category is Less than 10 years of experience; (c) The baseline category is North (I to III regions). Central represents IV-VII regions (including the XIII region), South represents VIII-XII regions; (d) The baseline category is Public and Private Employees; (e) The baseline category is Legislators, senior officials and managers. For each model Schooling corresponds to the declared schooling level for each individual in the sample. Specification (D) includes the same controls as (C) but is estimated including a correction for selection. The variables used in the first stage are number of children, mother's occupational situation, father's occupational situation, and whether or not the individual grew up in a poor household. Standard Errors are presented in parentheses.

Table 3. The Gender Gap in Monthly Hours Worked
SPS02

Variables	(A)	(B)	(C)	(D)
Male	0.12	0.11	0.11	0.004
	(0.02)	(0.02)	(0.02)	(0.02)
Schooling ^(a)				
Secondary Education	-	-	-0.01	-0.04
	-	-	(0.02)	(0.02)
Some Tertiary Education	-	-	0.02	-0.03
	-	-	(0.03)	(0.02)
Complete Tertiary Education	-	-	-0.03	-0.04
	-	-	(0.04)	(0.03)
Experience ^(b)				
Between 10 and 15 years of Experience	-	0.08	0.08	-0.07
	-	(0.02)	(0.02)	(0.02)
More than 10 years of Experience	-	0.08	0.08	-0.08
	-	(0.02)	(0.02)	(0.02)
Residence ^(c)				
Central	-0.002	-0.005	-0.01	0.02
	(0.03)	(0.03)	(0.03)	(0.03)
South	-0.05	-0.05	-0.05	-0.10
	(0.02)	(0.02)	(0.02)	(0.03)
Santiago	0.02	0.02	0.02	-0.03
	(0.02)	(0.02)	(0.02)	(0.02)
Type of Job ^(d)				
Employer or Self-Worker	-0.20	-0.20	-0.20	-0.05
	(0.02)	(0.02)	(0.02)	(0.02)
Domestic Service	-0.11	-0.12	-0.12	0.01
	(0.04)	(0.04)	(0.04)	(0.03)
Occupations ^(e)				
Professionals	-0.30	-0.28	-0.27	-0.22
	(0.04)	(0.04)	(0.04)	(0.03)
Technicians and associate professionals	-0.24	-0.24	-0.25	-0.17
	(0.04)	(0.04)	(0.04)	(0.03)
Clerks	-0.18	-0.18	-0.19	-0.16
	(0.04)	(0.04)	(0.04)	(0.03)
Service workers and shop and market sales workers	-0.19	-0.19	-0.19	-0.11
	(0.04)	(0.04)	(0.04)	(0.03)
Skilled agricultural and fishery workers	-0.18	-0.20	-0.20	-0.16
	(0.05)	(0.05)	(0.05)	(0.04)
Craft and related trades workers	-0.16	-0.17	-0.18	-0.13
	(0.04)	(0.04)	(0.04)	(0.03)
Plant and machine operators and assemblers	-0.12	-0.13	-0.13	-0.06
	(0.04)	(0.04)	(0.04)	(0.03)
Elementary occupations	-0.24	-0.25	-0.25	-0.17
	(0.04)	(0.04)	(0.04)	(0.03)
Constant	3.95	3.91	3.92	4.21
	(0.04)	(0.04)	(0.05)	(0.04)
Correction for Selection	No	No	No	Yes

Notes: (a) The baseline category is Primary Education; (b) The baseline category is Less than 10 years of experience; (c) The baseline category is North (I to III regions). Central represents IV-VII regions (including the XIII region), South represents VIII-XII regions; (d) The baseline category is Public and Private Employees; (e) The baseline category is Legislators, senior officials and managers. For each model Schooling corresponds to the declared schooling level for each individual in the sample. Specification (D) includes the same controls as (C) but is estimated including a correction for selection. The variables used in the first stage are number of children, mother's occupational situation, father's occupational situation, and whether or not the individual grew up in a poor household. Standard Errors are presented in parentheses.

Table 4. The Gender Gap in Employment
SPS02

Variables	(A)		(B)		(C)	
	Coefficient	Marg. Effect	Coefficient	Marg. Effect	Coefficient	Marg. Effect
Male	0.67	0.22	0.42	0.14	0.41	0.14
	(0.05)	(0.02)	(0.05)	(0.02)	(0.05)	(0.02)
Background ^(a)						
Number of Children	-0.09	-0.03	-0.08	-0.03	-0.04	-0.01
	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)
Age	0.02	0.01	-0.03	-0.01	-0.04	-0.01
	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)
Mother's Occupation	-0.05	-0.02	-0.03	-0.01	-0.06	-0.02
	(0.05)	(0.02)	(0.05)	(0.02)	(0.05)	(0.02)
Father's Occupation	-0.27	-0.08	-0.21	-0.07	-0.13	-0.04
	(0.29)	(0.08)	(0.30)	(0.09)	(0.31)	(0.09)
Growing Up in Poverty	-0.24	-0.08	-0.27	-0.09	-0.14	-0.05
	(0.05)	(0.02)	(0.05)	(0.02)	(0.05)	(0.02)
Schooling ^(b)						
Secondary Education	-	-	-	-	0.26	0.09
	-	-	-	-	(0.07)	(0.02)
Some Tertiary Education	-	-	-	-	0.59	0.17
	-	-	-	-	(0.08)	(0.02)
Complete Tertiary Education	-	-	-	-	1.22	0.27
	-	-	-	-	(0.12)	(0.01)
Experience ^(c)						
Between 10 and 15 years of Experience	-	-	0.66	0.20	0.73	0.21
	-	-	(0.06)	(0.02)	(0.06)	(0.02)
More than 10 years of Experience	-	-	0.88	0.26	1.04	0.29
	-	-	(0.07)	(0.02)	(0.08)	(0.02)
Residence ^(d)						
Central	-0.08	-0.03	-0.09	-0.03	-0.06	-0.02
	(0.08)	(0.03)	(0.08)	(0.03)	(0.08)	(0.03)
South	0.22	0.07	0.24	0.08	0.24	0.08
	(0.08)	(0.03)	(0.08)	(0.03)	(0.08)	(0.03)
Santiago	0.24	0.08	0.24	0.08	0.18	0.06
	(0.06)	(0.02)	(0.06)	(0.02)	(0.06)	(0.02)
Constant	0.03	-	1.25	-	0.91	-
	(0.37)	-	(0.40)	-	(0.42)	-

Notes: (a) Mother's and Father's Education are dummy variables that take a value of one if the respective parent worked as asalaried and zero otherwise; (b) The baseline category is Primary Education; (c) The baseline category is Less than 10 years of experience; (d) The baseline category is North (I to III regions). Central represents IV-VII regions (including the XIII region), South represents VIII-XII regions. Standard Errors are presented in parentheses.

Table 5. The Gender Gap in Accumulated Experience
SPS02

Variables ^(b)	Less Than 10 Years ^(a)		Between 10 and 15 Years ^(a)		More than 15 Years ^(a)
	Coefficient	Marg. Effect	Coefficient	Marg. Effect	Marg. Effect
Male	1.11 (0.07)	-0.40 (0.02)	1.92 (0.09)	0.11 (0.02)	0.29 (0.02)
Secondary Education	0.26 (0.11)	-0.04 (0.03)	-0.08 (0.12)	0.09 (0.03)	-0.04 (0.02)
Some College	0.08 (0.13)	0.04 (0.04)	-0.61 (0.14)	0.08 (0.03)	-0.13 (0.02)
College Graduates	-0.07 (0.16)	0.11 (0.04)	-1.16 (0.19)	0.07 (0.04)	-0.18 (0.02)
Mother's Years of Schooling	-0.01 (0.01)	0.002 (0.003)	0.00 (0.01)	-0.002 (0.003)	-0.0003 (0.003)
Father's Years of Schooling	-0.02 (0.01)	0.01 (0.003)	-0.04 (0.01)	-0.003 (0.003)	-0.01 (0.003)
Growing Up in Poverty	-0.05 (0.08)	0.003 (0.02)	0.06 (0.09)	-0.02 (0.02)	0.02 (0.02)
Growing Up in Broken Home	-0.15 (0.17)	0.01 (0.05)	0.16 (0.21)	-0.06 (0.04)	0.05 (0.03)
Age	0.11 (0.01)	-0.07 (0.00)	0.42 (0.01)	-0.01 (0.00)	0.08 (0.00)
Constant	-4.10 (0.40)		-15.07 (0.55)		

Notes: (a) The experience levels correspond to the accumulated experience declared during the interview. Post-secondary education includes includes technical education (complete and incomplete). (b) The schooling level corresponds to the schooling level declared in the sample. Post-secondary education includes includes technical education (complete and incomplete).

Table 6. The Gender Gap in Schooling Decisions
SPS02

Variables	Primary School		Secondary School		Some Post-Secondary Education		College Graduates
	Coefficient	Marg. Effect	Coefficient	Marg. Effect	Coefficient	Marg. Effect	Marg. Effect
Male	-0.30 (0.08)	0.04 (0.01)	-0.33 (0.09)	-0.02 (0.02)	-0.30 (0.10)	-0.02 (0.02)	-0.004 (0.01)
Mother's Years of Schooling	0.08 (0.02)	-0.02 (0.00)	0.15 (0.02)	-0.01 (0.00)	0.17 (0.02)	0.02 (0.00)	0.01 (0.002)
Father's Years of Schooling	0.05 (0.01)	-0.01 (0.00)	0.13 (0.01)	-0.02 (0.00)	0.16 (0.02)	0.02 (0.00)	0.01 (0.002)
Growing Up in Poverty	-0.59 (0.08)	0.11 (0.01)	-0.84 (0.09)	0.00 (0.02)	-0.84 (0.11)	-0.08 (0.02)	-0.03 (0.01)
Growing Up in Broken Home	0.43 (0.17)	-0.09 (0.03)	0.83 (0.21)	0.00 (0.04)	0.34 (0.23)	0.11 (0.03)	-0.01 (0.03)
Age	-0.03 (0.01)	0.004 (0.001)	-0.04 (0.01)	-0.002 (0.002)	-0.01 (0.01)	-0.004 (0.002)	0.003 (0.001)
Constant	1.07 (0.40)		-0.68 (0.45)		-2.35 (0.52)		

Notes: The schooling level corresponds to the schooling level declared in the sample. Post-secondary education includes includes technical education (complete and incomplete).

Table 7. The Gender Gap in Schooling Achievement
SPS02

Variables	Repeating a Grade in Primary School		Repeating a Grade in Secondary School		Average Score during Secondary School ^(a)
	Coefficient	Marg. Effect	Coefficient	Marg. Effect	Coefficient
Male	0.22	0.07	0.12	0.04	-0.31
	(0.05)	(0.01)	(0.05)	(0.02)	(0.04)
Mother's Education ^(b)					
Secondary Education	-0.06	-0.02	0.02	0.01	0.05
	(0.06)	(0.02)	(0.06)	(0.02)	(0.04)
Some Tertiary Education	0.14	0.04	-0.06	-0.02	0.11
	(0.19)	(0.06)	(0.20)	(0.06)	(0.13)
Complete Tertiary Education	-0.29	-0.08	-0.13	-0.04	0.35
	(0.22)	(0.06)	(0.20)	(0.05)	(0.13)
Father's Education ^(b)					
Secondary Education	-0.17	-0.05	-0.05	-0.01	0.14
	(0.06)	(0.02)	(0.06)	(0.02)	(0.04)
Some Tertiary Education	-0.51	-0.13	-0.29	-0.08	0.23
	(0.16)	(0.03)	(0.16)	(0.04)	(0.10)
Complete Tertiary Education	-0.41	-0.11	-0.11	-0.03	0.21
	(0.16)	(0.04)	(0.15)	(0.04)	(0.10)
Background					
Growing Up in Poverty	0.25	0.08	-0.04	-0.01	-0.16
	(0.05)	(0.02)	(0.06)	(0.02)	(0.04)
Growing Up in Broken Home	-0.38	-0.13	0.04	0.01	0.10
	(0.11)	(0.04)	(0.13)	(0.04)	(0.09)
School Characteristics ^(c)					
Urban Primary School	-0.20	-0.07	-	-	0.02
	(0.08)	(0.03)	-	-	(0.08)
Urban Secondary School	-	-	0.40	0.10	0.22
	-	-	(0.24)	(0.05)	(0.16)
Private-Subsized Primary School	-0.10	-0.03	-	-	0.07
	(0.07)	(0.02)	-	-	(0.06)
Cooperation - Primary School	-0.45	-0.12	-	-	0.22
	(0.59)	(0.12)	-	-	(0.35)
Private Primary School	-0.27	-0.08	-	-	0.09
	(0.12)	(0.03)	-	-	(0.09)
Private-Subsized Secondary School	-	-	-0.21	-0.06	0.13
	-	-	(0.06)	(0.02)	(0.05)
Cooperation - Secondary School	-	-	-0.42	-0.10	0.15
	-	-	(0.26)	(0.05)	(0.17)
Private Secondary School	-	-	-0.41	-0.10	0.24
	-	-	(0.12)	(0.03)	(0.10)
Constant	-0.18	-	-1.15	-	-0.33
	(0.13)	-	(0.27)	-	(0.17)

Notes: (a) The average score is standardized to have mean 0 and variance 1 in the population; (b) The baseline category is Primary Education; (c) In the case of the dummies controlling for the type of management the baseline category is Public School. Standard Errors are presented in parentheses.

Table 8A. Variables in the empirical implementation of the model
Outcome Equations

Variables	Hourly Wage ^(a)	Monthly Hours Worked ^(a)	Employment ^(a)	Accumulated Experience ^(b)	Educational Choice Model ^(c)
Gender Dummy	Yes	Yes	Yes	Yes	Yes
Region of Residence	Yes	Yes	Yes	-	-
Growing Up in Broken Home	-	-	-	-	Yes
Mother's Education	-	-	-	Yes	Yes
Father's Education	-	-	-	Yes	Yes
Growing Up in Poverty	-	-	-	Yes	Yes
Age	-	-	Yes	Yes	Yes
Type of Occupation	Yes	Yes	-	-	-
Type of Job	Yes	Yes	-	-	-
Unobserved Ability	Yes	Yes	Yes	Yes	Yes

Notes: (a) Hourly wages, monthly hours worked and employment models are estimated for four different schooling categories (primary, secondary, some tertiary and complete tertiary) and three different levels of accumulated experience (less than 10 years, between 10 and 15 years, and more than 15 years). In each case, the labor market outcome refers to the previous month individual's outcome; (b) Accumulated experience is modeled with a multinomial choice model. The categories considered are: less than 10 years, between 10 and 15 years, and more than 15 years. The level of accumulated experience is the total work experience reported at the time of the interview;

Table 8B. Variables in the empirical implementation of the model
Auxiliary Measures

Variables	Average Grade in Secondary Education	Repeat Any Grade in Primary School	Repeat Any Grade in Secondary School
Primary School in a Urban Area (Dummy)	Yes	Yes	-
Secondary School in a Urban Area (Dummy)	Yes	-	Yes
Growing Up in Broken Home	Yes	Yes	Yes
Mother's Education	Yes	Yes	Yes
Father's Education	Yes	Yes	Yes
Growing Up in Poverty	Yes	Yes	Yes
Primary School System (Public, Private, etc.)	Yes	Yes	-
Secondary School System (Public, Private, etc.)	Yes	-	Yes
Unobserved Ability	Yes	Yes	1.0

Table 9. Model with Essential Heterogeneity
Gender Gap in Hourly Wages, by Schooling Level and Accumulated Experience ^(a)
SPS02

Variable	High School Dropouts			High School Graduates			Some Post-Secondary Education			College Graduates		
	Less than 10 Years	Between 10 and 15 Years	More than 15 Years	Less than 10 Years	Between 10 and 15 Years	More than 15 Years	Less than 10 Years	Between 10 and 15 Years	More than 15 Years	Less than 10 Years	Between 10 and 15 Years	More than 15 Years
Male	-0.06	0.30	0.07	0.35	0.15	0.19	0.23	0.35	0.15	0.38	0.38	0.36
	(0.29)	(0.18)	(0.14)	(0.09)	(0.06)	(0.05)	(0.08)	(0.08)	(0.10)	(0.08)	(0.15)	(0.20)
Employer or Self-Worker ^(b)	-0.41	-0.34	-0.30	0.19	-0.19	-0.23	0.22	0.06	-0.12	0.00	-0.10	0.41
	(0.37)	(0.15)	(0.10)	(0.11)	(0.08)	(0.06)	(0.12)	(0.14)	(0.14)	(0.15)	(0.20)	(0.39)
Domestic Service	-0.52	0.18	-0.27	-0.11	-0.13	0.16	-	0.08	-1.44	-0.11	-	-
	(0.37)	(0.24)	(0.20)	(0.16)	(0.17)	(0.17)	-	(0.28)	(0.61)	(0.50)	-	-
Professionals ^(c)	-	-	-	-1.04	-	-0.52	0.26	0.42	-0.11	-0.21	-0.18	-0.18
	-	-	-	(0.63)	-	(0.48)	(0.22)	(0.36)	(0.29)	(0.14)	(0.24)	(0.30)
Technicians and associate professionals	-	-	-	-0.96	-0.36	-0.40	0.03	0.26	-0.41	-0.11	-0.46	-0.03
	-	-	-	(0.28)	(0.20)	(0.15)	(0.18)	(0.25)	(0.23)	(0.18)	(0.30)	(0.35)
Clerks	-	-	-0.83	-1.22	-0.48	-0.43	-0.38	-0.11	-0.53	-0.55	-0.79	-0.56
	-	-	(0.60)	(0.25)	(0.18)	(0.13)	(0.19)	(0.26)	(0.24)	(0.19)	(0.39)	(0.47)
Service workers and shop and market sales workers	-	-0.57	-0.48	-1.66	-0.61	-0.84	-0.64	-0.46	-0.59	-	-0.84	-
	-	(0.48)	(0.32)	(0.25)	(0.18)	(0.13)	(0.20)	(0.26)	(0.24)	-	(0.52)	-
Skilled agricultural and fishery workers	0.39	-0.55	-0.78	-1.75	-0.83	-0.92	-0.44	0.47	-0.81	-	-	-
	(0.57)	(0.40)	(0.29)	(0.37)	(0.24)	(0.16)	(0.63)	(0.46)	(0.38)	-	-	-
Craft and related trades workers	0.25	-0.38	-0.57	-1.44	-0.68	-0.67	-0.57	-0.27	-0.63	-	-	-1.22
	(0.34)	(0.38)	(0.28)	(0.26)	(0.17)	(0.12)	(0.22)	(0.26)	(0.25)	-	-	(0.85)
Plant and machine operators and assemblers	0.69	-0.36	-0.57	-1.52	-0.72	-0.81	-0.80	-0.52	-1.11	-	0.59	-1.37
	(0.56)	(0.40)	(0.30)	(0.26)	(0.17)	(0.12)	(0.27)	(0.27)	(0.32)	-	(0.69)	(0.59)
Elementary occupations	0.35	-0.67	-0.63	-1.66	-0.80	-0.88	-0.91	-0.55	-1.18	-0.97	-1.75	-
	(0.32)	(0.38)	(0.28)	(0.26)	(0.20)	(0.13)	(0.28)	(0.30)	(0.28)	(0.38)	(0.50)	-
Central	0.48	-0.11	-0.13	-0.29	-0.21	-0.15	-0.05	-0.19	0.11	-0.13	0.38	-0.58
	(0.51)	(0.22)	(0.13)	(0.12)	(0.09)	(0.07)	(0.15)	(0.15)	(0.23)	(0.17)	(0.28)	(0.38)
South	0.41	-0.13	-0.03	-0.24	-0.07	0.03	-0.10	-0.13	0.11	0.03	0.42	-0.12
	(0.43)	(0.21)	(0.14)	(0.11)	(0.09)	(0.08)	(0.13)	(0.14)	(0.23)	(0.15)	(0.27)	(0.34)
Santiago	-0.25	0.36	0.19	0.22	0.25	0.27	0.08	0.20	0.12	0.35	-0.05	0.39
	(0.35)	(0.16)	(0.11)	(0.10)	(0.07)	(0.06)	(0.11)	(0.10)	(0.13)	(0.11)	(0.19)	(0.26)
Intercept	5.56	7.04	7.04	8.10	7.36	7.37	7.28	7.12	7.52	8.07	7.98	8.80
	(0.47)	(0.51)	(0.34)	(0.27)	(0.19)	(0.13)	(0.21)	(0.30)	(0.30)	(0.25)	(0.43)	(0.55)
Unobserved Heterogeneity	-0.20	0.71	-0.03	0.17	0.13	-0.19	-0.32	-0.30	-0.004	-0.39	-0.50	-0.89
	(0.13)	(0.43)	(0.07)	(0.15)	(0.11)	(0.09)	(0.15)	(0.17)	(0.20)	(0.22)	(0.39)	(0.51)

Notes: The accumulated experience corresponds to the retrospective information reported by the individual at the time of the interview. The schooling level corresponds to the schooling level declared in the sample. Post-secondary education includes includes technical education (complete and incomplete). (b) For the characteristics of the type of job (employer or self-worker and domestic service), the baseline category is Public and Private Employees; (c) For the set of variables controlling for occupation characteristics (from Professionals to Elementary Occupations in this table) the baseline category is Legislators, senior officials and managers.

Table 10. Model with Essential Heterogeneity
Gender Gap in Hours Worked, by Schooling Level and Accumulated Experience ^(a)
SPS02

Variable	High School Dropouts			High School Graduates			Some Post-Secondary Education			College Graduates		
	Less than 10 Years	Between 10 and 15	More than 15 Years	Less than 10 Years	Between 10 and 15 Years	More than 15 Years	Less than 10 Years	Between 10 and 15 Years	More than 15 Years	Less than 10 Years	Between 10 and 15	More than 15 Years
Male	-0.06	0.18	0.14	0.08	0.10	0.07	0.17	0.12	0.10	0.02	0.00	0.08
	(0.15)	(0.13)	(0.08)	(0.07)	(0.03)	(0.03)	(0.05)	(0.04)	(0.05)	(0.06)	(0.08)	(0.10)
Employer or Self-Worker ^(b)	-0.01	-0.12	-0.17	-0.53	-0.06	-0.16	-0.24	-0.31	-0.22	-0.09	-0.02	-0.07
	(0.18)	(0.11)	(0.06)	(0.09)	(0.04)	(0.03)	(0.08)	(0.07)	(0.07)	(0.11)	(0.10)	(0.20)
Domestic Service	0.18	-0.09	0.08	-0.25	0.00	-0.18	-	0.07	-0.75	-0.09	-	-
	(0.19)	(0.18)	(0.11)	(0.14)	(0.08)	(0.09)	-	(0.14)	(0.30)	(0.36)	-	-
Professionals ^(c)	-	-	-	-0.96	-	-0.08	-0.34	-0.45	-0.33	-0.20	-0.29	-0.15
	-	-	-	(0.55)	-	(0.26)	(0.15)	(0.18)	(0.15)	(0.10)	(0.13)	(0.15)
Technicians and associate professionals	-	-	-	-0.34	-0.05	-0.21	-0.42	-0.38	-0.22	-0.19	-0.19	-0.31
	-	-	-	(0.24)	(0.10)	(0.08)	(0.12)	(0.12)	(0.11)	(0.13)	(0.15)	(0.17)
Clerks	-	-	0.31	-0.21	-0.11	-0.25	-0.27	-0.34	-0.22	-0.02	-0.15	-0.02
	-	-	(0.34)	(0.22)	(0.09)	(0.07)	(0.13)	(0.13)	(0.12)	(0.14)	(0.19)	(0.24)
Service workers and shop and market sales workers	-	0.36	-0.27	-0.11	-0.11	-0.23	-0.38	-0.31	-0.28	-0.02	0.00	-
	-	(0.36)	(0.20)	(0.21)	(0.09)	(0.07)	(0.13)	(0.13)	(0.12)	(0.28)	(0.27)	-
Skilled agricultural and fishery workers	0.01	-0.10	-0.09	-0.31	-0.06	-0.22	0.11	-0.29	-0.09	-	-	-
	(0.29)	(0.29)	(0.19)	(0.30)	(0.12)	(0.09)	(0.41)	(0.23)	(0.18)	-	-	-
Craft and related trades workers	-0.11	0.07	-0.06	-0.23	-0.05	-0.23	-0.27	-0.33	-0.09	-	-	0.38
	(0.17)	(0.28)	(0.18)	(0.22)	(0.08)	(0.06)	(0.15)	(0.13)	(0.12)	-	-	(0.43)
Plant and machine operators and assemblers	0.06	0.03	-0.08	-0.09	-0.02	-0.16	-0.06	-0.25	-0.43	-	-0.06	0.02
	(0.28)	(0.30)	(0.20)	(0.23)	(0.09)	(0.06)	(0.18)	(0.13)	(0.16)	-	(0.36)	(0.29)
Elementary occupations	-0.16	0.01	-0.09	-0.28	-0.09	-0.27	-0.68	-0.64	-0.05	-	-0.52	-
	(0.16)	(0.28)	(0.18)	(0.22)	(0.10)	(0.07)	(0.18)	(0.15)	(0.14)	-	(0.26)	-
Central	0.05	-0.07	0.05	-0.04	-0.02	-0.02	0.13	-0.05	-0.12	-0.07	-0.26	-0.20
	(0.25)	(0.16)	(0.08)	(0.10)	(0.05)	(0.04)	(0.09)	(0.07)	(0.12)	(0.12)	(0.14)	(0.20)
South	-0.19	-0.01	-0.15	0.00	0.00	-0.04	0.07	-0.04	-0.07	-0.19	-0.22	-0.31
	(0.22)	(0.16)	(0.08)	(0.09)	(0.04)	(0.04)	(0.09)	(0.07)	(0.12)	(0.11)	(0.14)	(0.17)
Santiago	-0.09	-0.02	-0.02	0.07	0.05	0.01	-0.03	0.03	0.09	0.02	0.02	0.03
	(0.17)	(0.12)	(0.06)	(0.09)	(0.04)	(0.03)	(0.07)	(0.05)	(0.07)	(0.08)	(0.10)	(0.14)
Intercept	4.35	3.54	4.01	3.91	3.85	4.05	3.92	4.17	4.06	3.90	4.04	4.05
	(0.27)	(0.37)	(0.22)	(0.22)	(0.09)	(0.07)	(0.14)	(0.14)	(0.15)	(0.17)	(0.21)	(0.27)
Unobserved Heterogeneity	0.40	-0.39	0.23	-0.15	0.01	0.03	0.18	0.06	0.08	0.08	0.17	-0.04
	(0.13)	(0.26)	(0.06)	(0.12)	(0.05)	(0.05)	(0.10)	(0.08)	(0.10)	(0.13)	(0.17)	(0.22)

Notes: The accumulated experience corresponds to the retrospective information reported by the individual at the time of the interview. The schooling level corresponds to the schooling level declared in the sample. Post-secondary education includes includes technical education (complete and incomplete). (b) For the characteristics of the type of job (employer or self-worker and domestic service), the baseline category is Public and Private Employees; (c) For the set of variables controlling for occupation characteristics (from Professionals to Elementary Occupations in this table) the baseline category is Legislators, senior officials and managers.

Table 11. Model with Essential Heterogeneity
Gender Gap in Employment Status, by Schooling Level and Accumulated Experience
SPS02

Variable	High School Dropouts ^(a)			High School Graduates			Some Post-Secondary Education			College Graduates ^(b)	
	Less than 10 Years	Between 10 and 15	More than 15 Years	Less than 10 Years	Between 10 and 15	More than 15 Years	Less than 10 Years	Between 10 and 15	More than 15 Years	Less than 10 Years	Between 10 and 15 Years
Male	1.40	0.35	-0.10	0.98	0.30	0.34	0.80	0.36	0.18	-0.12	-0.23
	(0.54)	(0.35)	(0.35)	(0.14)	(0.12)	(0.13)	(0.14)	(0.20)	(0.24)	(0.25)	(0.47)
Central	-	-	-	-0.17	-0.09	0.00	0.03	0.42	0.42	-0.39	-
	-	-	-	(0.17)	(0.21)	(0.22)	(0.23)	(0.32)	(0.48)	(0.61)	-
South	-	-	-	0.16	0.05	-0.06	0.29	1.17	1.45	-0.36	-
	-	-	-	(0.17)	(0.20)	(0.22)	(0.23)	(0.37)	(0.62)	(0.59)	-
Santiago	-	-	-	0.17	0.27	0.10	-0.06	0.20	0.10	-0.19	-
	-	-	-	(0.14)	(0.16)	(0.16)	(0.18)	(0.26)	(0.33)	(0.33)	-
Number of Children	0.09	0.04	0.11	-0.11	-0.07	-0.07	-0.04	-0.03	0.01	0.14	0.17
	(0.10)	(0.13)	(0.10)	(0.05)	(0.05)	(0.05)	(0.07)	(0.09)	(0.10)	(0.13)	(0.21)
Intercept	-2.20	-0.26	0.13	-0.08	0.62	0.79	0.06	0.40	0.38	1.13	3.16
	(1.17)	(0.54)	(0.61)	(0.17)	(0.20)	(0.23)	(0.21)	(0.29)	(0.44)	(0.72)	(2.09)
Unobserved Heterogeneity	-1.65	-1.63	-1.64	0.37	0.21	0.10	0.07	0.07	0.22	0.59	-1.69
	(1.35)	(1.27)	(1.62)	(0.22)	(0.27)	(0.25)	(0.29)	(0.44)	(0.54)	(0.70)	(1.81)

Notes: The accumulated experience corresponds to the retrospective information reported by the individual at the time of the interview. The schooling level corresponds to the schooling level declared in the sample. Post-secondary education includes technical education (complete and incomplete). (a) Among high school dropouts, the characteristics of the place of residence perfectly predict the labor status, so those variables are excluded in these cases. (b) For the group of individuals reporting more than 15 years of experience and a college degree, the gender dummy perfectly predicts the labor status: the 29 women in these category reported to be working (34 out of 37 males report to be working). Since the gender coefficient is the main interest of this table we do not include this model here.

Table 12. Model with Essential Heterogeneity
Gender Gap in Accumulated Experience, by Schooling Level
SPS02

Variables	High School Dropouts		High School Graduates		Some College		College Graduates	
	Between 10 and 15 Years	More than 15 Years	Between 10 and 15 Years	More than 15 Years	Between 10 and 15 Years	More than 15 Years	Between 10 and 15 Years	More than 15 Years
Male	1.47	2.95	1.48	2.50	0.75	1.02	0.14	0.41
	(0.23)	(0.24)	(0.12)	(0.15)	(0.14)	(0.21)	(0.28)	(0.42)
Mother's Years of Schooling	0.04	-0.02	0.03	0.03	-0.01	0.05	-0.04	-0.11
	(0.05)	(0.05)	(0.02)	(0.02)	(0.02)	(0.04)	(0.06)	(0.09)
Father's Years of Schooling	-0.04	0.00	0.00	-0.02	-0.02	0.00	0.09	-0.06
	(0.05)	(0.04)	(0.02)	(0.02)	(0.02)	(0.04)	(0.06)	(0.10)
Growing Up in Poverty	-0.21	-0.03	-0.16	-0.14	0.02	0.25	-0.30	0.79
	(0.24)	(0.22)	(0.12)	(0.14)	(0.19)	(0.26)	(0.42)	(0.69)
Age	-0.05	0.26	0.11	0.46	0.20	0.61	0.39	0.78
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.06)	(0.06)	(0.11)
Intercept	1.37	-10.45	-4.22	-17.06	-6.98	-22.97	-16.69	-25.84
	(1.29)	(1.21)	(0.64)	(0.97)	(1.00)	(2.73)	(3.54)	(4.42)
Unobserved Heterogeneity	1.19	-0.20	1.23	1.57	0.53	2.09	2.95	-1.71
	(0.74)	(0.17)	(0.28)	(0.36)	(0.45)	(0.96)	(1.55)	(3.10)

Notes: The accumulated experience corresponds to the retrospective information reported by the individual at the time of the interview. The table presents the results for three multinomial choice models (each for each schooling level). The baseline category is less than 10 years of accumulated experience.

Table 13. Model with Essential Heterogeneity
Gender Gap in Schooling Decisions
SPS02

Variable	Secondary School	Some Post- Secondary	College Graduates
Male	-0.47 (0.11)	-0.55 (0.13)	-0.61 (0.30)
Mother's Years of Schooling	0.13 (0.02)	0.23 (0.03)	0.41 (0.06)
Father's Years of Schooling	0.09 (0.02)	0.21 (0.02)	0.44 (0.07)
Growing Up in Poverty	-0.03 (0.01)	-0.03 (0.02)	0.08 (0.04)
Growing Up in Broken Home	0.53 (0.22)	1.02 (0.30)	0.46 (0.71)
Age	-0.81 (0.11)	-1.25 (0.15)	-2.20 (0.46)
Intercept	1.10 (0.51)	-1.66 (0.64)	-12.93 (2.99)
Unobserved Heterogeneity	1.90 (0.38)	3.52 (0.48)	10.90 (1.96)

Notes: The schooling level corresponds to the schooling level declared in the sample. Post-secondary education includes technical education (complete and incomplete). The baseline category is Primary School.

Table 14. Model with Essential Heterogeneity
Gender Gap in Hours Schooling Achievement
SPS02

Variables ^(a)	Repeating a Grade in Primary School	Repeating a Grade in Secondary School	Average Score during Secondary School ^(b)
Male	0.26 (0.05)	0.17 (0.06)	-0.33 (0.04)
Mother: Secondary Education	-0.08 (0.07)	-0.04 (0.07)	0.10 (0.04)
Mother: Some Tertiary Education	0.06 (0.21)	-0.20 (0.22)	0.23 (0.12)
Mother: Complete Tertiary Education	-0.30 (0.25)	-0.18 (0.23)	0.38 (0.12)
Father: Secondary Education	-0.20 (0.06)	-0.12 (0.07)	0.20 (0.04)
Father: Some Tertiary Education	-0.61 (0.18)	-0.44 (0.18)	0.30 (0.10)
Father: Complete Tertiary Education	-0.40 (0.17)	-0.17 (0.17)	0.24 (0.09)
Growing Up in Poverty	0.28 (0.06)	0.04 (0.07)	-0.23 (0.04)
Growing Up in Broken Home	-0.09 (0.08)	0.45 (0.26)	0.01 (0.08)
Urban Primary School	-0.40 (0.12)	-	0.13 (0.09)
Urban Secondary School	-	0.01 (0.16)	0.19 (0.14)
Private-Subsized Primary School	0.00 (0.08)	-	-0.01 (0.05)
Cooperation - Primary School	-0.57 (0.68)	-	0.26 (0.32)
Private Primary School	-0.11 (0.13)	-	-0.06 (0.08)
Private-Subsized Secondary School	-	-0.20 (0.07)	0.12 (0.05)
Cooperation - Secondary School	-	-0.47 (0.28)	0.16 (0.15)
Private Secondary School	-	-0.28 (0.14)	0.17 (0.09)
Intercept	-0.36 (0.14)	-1.17 (0.30)	-0.40 (0.16)
Unobserved Heterogeneity	-0.98 (0.09)	-1.22 (0.12)	1.00 -

Notes: (a) In the case of mother's and father's education the baseline category is Primary Education. In the case of the dummies controlling for the type of management the baseline category is Public School. (b) The average score is standardized to have mean 0 and variance 1 in the population. Standard Errors are presented in parentheses.

Table A1. Descriptive Statistics SPS02 by Gender

Variable (Dummy=1 if Apply)	Females		Males	
	Mean	Std. Dev.	Mean	Std. Dev.
Age	33.76	3.76	33.71	3.79
A. School Information				
Maximum Schooling Level = Primary Education	0.11	0.32	0.17	0.38
Maximum Schooling Level = Secondary Education	0.51	0.50	0.49	0.50
Maximum Schooling Level = Some Tertiary Education	0.26	0.44	0.24	0.43
Maximum Schooling Level = Complete Tertiary Education	0.11	0.31	0.10	0.30
<i>A.1. Primary School</i>				
Primary School in Urban Area	0.91	0.29	0.89	0.31
Repeating a Grade in Primary School	0.22	0.41	0.30	0.46
Was Primary School Public?	0.77	0.42	0.81	0.39
Was Primary School Private-Subsidized?	0.16	0.37	0.13	0.33
Was Primary School Managed by a Corporation?	0.00	0.05	0.00	0.04
Was Primary School Private?	0.07	0.25	0.06	0.23
<i>A.2. Secondary School</i>				
Secondary School in Urban Area	0.98	0.14	0.99	0.12
Repeating a Grade in Secondary School	0.20	0.40	0.24	0.43
Was Secondary School Public?	0.70	0.46	0.70	0.46
Was Secondary School Private-Subsidized?	0.23	0.42	0.22	0.42
Was Secondary School Managed by a Corporation?	0.01	0.08	0.02	0.13
Was Secondary School Private?	0.07	0.25	0.06	0.24
Average Grade in Secondary School	0.16	0.98	-0.17	1.00
B. Family Background				
Mother's Employment - Asalaried	0.56	0.50	0.55	0.50
Father's Employment - Asalaried	0.99	0.07	0.99	0.09
Total Number of Children	1.64	1.19	1.47	1.22
Mother's Education (years of schooling)	7.51	3.77	7.42	3.69
Father's Education (years of schooling)	8.14	4.11	7.91	4.00
Growing up under Poverty	0.28	0.45	0.35	0.48
Growing up in a Broken Home	0.96	0.20	0.96	0.20
C. Labor Market Variables				
Monthly Earnings	215,266	214,323	285,140	360,046
Hours Worked per Week	43.41	11.74	48.17	9.81
Hourly Wage	1,292	1,257	1,636	4,649
Working During Last Month	0.59	0.49	0.82	0.39
Total Work Experience since Jan. 1980	113.43	66.00	165.02	63.52
Less than 10 years of Experience	0.56	0.50	0.25	0.43
Between 10 and 15 years of Experience	0.26	0.44	0.34	0.47
More than 15 years of Experience	0.18	0.39	0.41	0.49
<i>C.1 Type of Job</i>				
Asalaried	0.81	0.39	0.80	0.40
Employer or Self-Worker	0.11	0.32	0.20	0.40
Domestic Service	0.08	0.27	0.00	0.02
<i>C.2 Type of Occupation</i>				
Administrative and Managerial Workers	0.03	0.17	0.06	0.24
Professionals	0.13	0.34	0.08	0.27
Technicians and associate professionals	0.14	0.35	0.11	0.32
Clerks	0.26	0.44	0.10	0.30
Service workers and shop and market sales workers	0.22	0.42	0.09	0.29
Skilled agricultural and fishery workers	0.01	0.09	0.06	0.23
Craft and related trades workers	0.04	0.19	0.23	0.42
Plant and machine operators and assemblers	0.04	0.19	0.17	0.37
Elementary occupations	0.13	0.34	0.10	0.31
D. Place of Residence				
North (I to III Regions)	0.13	0.33	0.11	0.32
Central (IV to VII Regions)	0.65	0.48	0.62	0.49
South (VIII to XII Regions)	0.23	0.42	0.26	0.44
Santiago (Region XIII)	0.43	0.49	0.42	0.49
Number of Observations	1,765		1,801	

Note: The numbers presented in this table corresponds to the sample of individuals with ages between 28 and 40 years old at the time of the interview.

