Session Number: Plenary Session 5
Time: Thursday, August 26, AM

Paper Prepared for the 31st General Conference of
The International Association for Research in Income and Wealth

St. Gallen, Switzerland, August 22-28, 2010

Diploma earning differences by gender in Colombia

Jhon James Mora and Juan Muro

For additional information please contact:

Name: Jhon James Mora
Affiliation: Universidad Icesi and Universidad de Alcalá

Email address: jjmora@icesi.edu.co

This paper is posted on the following website: http://www.iariw.org

Diploma earning differences by gender in Colombia

Jhon James Mora[1] and Juan Muro[2]

Paper Prepared to IARIW 31st General Conference, St-Gallen, Switzerland, August 22-28, 2010

Abstract: This paper discusses the existence of diploma earnings differences by gender in Colombia with a model of sheepskin effects based on pseudo panel data for the period 1996-2000. Our results show a significant and distinctive effect of high school and university degrees among men and women. Thus, additional earnings associated with a high school degree are higher for women than for men, while additional earnings associated with a university degree are higher for men compared to women in Colombia in the period under consideration.

JEL classification: J7;J31; C31
Keywords: Sheepskin effects; Pseudo Panel Data; Gender; Selection bias; Colombia

Introduction

The labor market in Colombia showed from 1981 to 2000 a relevant growth of the rate of female participation. The female rate of participation went up from a low 37 percent in 1981 to a 51 percent in 1998. Meanwhile labor participation of men showed stagnation in the two last decades, i.e. 74 percent in 1981 and 74 percent in 1998 (Luisa Fernanda Bernat, Rocio Ribero and Jaime Tenjo 2004:150). So, the traditional gender gap in labor force participation has narrowed in the last decades of the century. In the same period, however, the returns to education by gender exhibited a steady path with the returns for women always above those for men, around an overall 2 percent in 1989 and 1998 (Bernat et al 2004:161)[3]. This evolution has taken place in a labor market characterized by a high component of screening in the hiring process. In it salaries not only depend on productivity but on credentials. In a labor market with different collectives, majorities and minorities defined by gender and education achievement, the effect of a signaling process through a schooling diploma on wages varies across the diverse collectives. This fact has deep consequences on the design of education and anti-discrimination policies that in order to be effective must be targeted toward heterogeneous groups.

To discuss the effect of schooling degrees on the workers´ labor earnings by gender in Colombia we use Thomas Hungerford and Gary Solon's equation (1987), also known as the "sheepskin effect" equation. Under signaling in the labor market [See Spence (1973)], the Belman and Heywood (1991:721) extension of the sheepskin model allows us to analyze differences in size of the sheepskin effects. Given identical signals of productivity, departures from the mean productivity are driven by the cost of obtaining an inaccurate signal that in turn differs for distinct groups. Collectives with a higher cost of acquiring a diploma will have a lower expected productivity with a low signal and in consequence will receive a lower sheepskin effect. For a high signal the opposite applies.

---

[1] Universidad Icesi and Universidad de Alcalá
[2] Universidad de Alcalá
[3] We lack of comparative figures by gender and education for returns to education.

Empirically, our paper differs from other applications of the sheepskin equation in incorporating the use of pseudo panel data. On the one hand it is a requirement of the kind of available information on the Colombian labor market due we don't have a panel data information – only cross section data in each year; on the other, pseudo panel data conveys a well known set of advantages: the specified repeated cross-section model controls for individual heterogeneity and selection biases, eliminates attrition problems and, last but not least, allows analyzing the effect of diplomas on salaries as a result of a permanent effect of credentials in the labor market and not as a result of a transitory effect of credentials.

Our results based on inference from a discontinuous spline specification of the sheepskin equation show that there is a significant and distinctive effect of high school and university degrees among men and women. The findings reveal that women holding a high school degree had higher additional earnings than men, but men holding a university degree obtained higher additional earnings than women in Colombia in the period 1996-2000. As far as the contrast of selection biases is concerned, the findings show that there is a selection bias in both men and women equations. These results have important political implications in Colombia. First of all, we must discuss the effect of credentials on the labor market. Secondly and most importantly, we must discuss the anti-discriminatory policy to compensate for wage differences between men and women.

Sheepskin effect equation

Michael Spence (1973, 2002) and Kenneth Arrow (1973) made contributions that gave rise to a considerable amount of work relative to the debate of human capital and signaling. The theory of human capital postulated by Gary Becker (1964) contends that education (and on-the-job training) directly increases an individual's productivity and, thus, his/her salary. Therefore, each additional year of schooling brings about a proportional salary increase. On the other hand, Spence's (1973) and Arrow's (1973) theories of both signaling and screening suggest that in the years in which a degree is earned this salary increase is more than proportional because schooling degrees provide either indications of a worker's productivity or the grounds for signaling or screening.

In the mid nineteen eighties, Thomas Hungerford and Gary Solon (1987:175) found evidence to confirm that "wages will rise faster with each extra year of education when an extra year also conveys a certificate". Therefore, a diploma has its own value aside from the number of years of schooling. Similarly, using cohorts from 1979 and 1991 in a cross-section model, Dale Belman and John Heywood (1997) found empirical evidence confirming that degrees do have an effect on salaries in US. However, Belman and Heywood (1997:634) do not control by number of individuals in the cohort and their statement that: "The sheepskin effects of a college degree display a clear pattern of declining effects within cohorts" is weak when we observed 1,592 individuals in the last cohort and 18,438 individuals in the first cohort.

On the other hand, the results could change if instead of a sample of homogeneous individuals there were different groups under analysis. Richard Freeman (1977) showed that black academics who have published books earn more than their fellow white academics. Steven Culler, Edmund Becker and Robert Ohsteldt (1987) also found that female doctors who achieve medical certification see a higher increase of their earnings than their male counterparts. Devra Golbe (1985) showed that statistical discrimination in the signaling model could translate into higher earnings for the minorities than for the majorities in response to a signal of high productivity. Dale Belman and John Heywood (1991) followed Golbe´s postulate (1985) and incorporated it to Hungerford and Solon´s (1987) sheepskin equation. Dale Belman and John Heywood (1991) specifically stated that the main reason that accounts for the differences between both groups is that the cost of achieving a signal is higher for minorities than for majorities. Their findings about men and women in general as well as about African American men and women clearly show that when a signal is low, the associated earnings are higher for the majority group compared with the minority group. When the signal is high, however, the minority group sees higher earnings than the majority group.

In the Colombian case, Bernat et al. (2004:161) analyze gender differences in the returns to education and find that these are always greater for women than for men in the equations corrected for selectivity. The authors found an 11 percent return from each additional year of education for men and a 13 percent return from each additional year of education for women in 1998. Bernat et al. (2004) follow the traditional human capital equation and don't discuss the non-linearity effects over the returns to education.

The first applications of the sheepskin equation in Colombia made by Jhon James Mora (2003) don't either discuss the gender differences in the returns of diploma or the use of pseudo panel data. Estimating the effect of diplomas using pseudo panel data entails the use of the following model:

$$\begin{aligned} LnWh_{i(t),t} &= \alpha_1 S_{i(t),t} + \alpha_2 exp_{i(t),t} + \alpha_3 exp^2{}_{i(t),t} + \beta_0 S11_{i(t),t} \\ &+ \beta_1 S11_{i(t),t} * (S\text{-}11)_{i(t),t} + \beta_2 S16_{i(t),t} + \beta_3 S16 * (S\text{-}16)_{i(t),t} \\ &+ \beta_4 S17_{i(t),t} + C_{i(t)} + \nu_{i(t)} + \mu_{i(t),t} \end{aligned} \qquad (1)$$

In model (1) $LnWh_{i(t),t}$ $S_{i(t),t}$ and $exp_{i(t),t}$ are the logarithm of wages per hour, years of schooling, and potential experience (age-s-6), respectively, for a worker i in a given period of time t. Sub-index i(t) denotes the fact that individuals are different in each period of time, that is we have a pseudo panel. $S11_{i(t),t}$ is a dummy variable with a value of one if an individual has completed 11 years of schooling or more, and variable $S16_{i(t),t}$ is a also a dummy with a value of one if an individual has completed 16 years of schooling or more. $\nu_{i(t)}$ represents the deviation of the effect of the cohort after breaking down fixed individual effects. Therefore, if there are any fixed individual effects, these will be consistent with fixed effects in the cohort. In the sheepskin effect model, these effects could imply that a diploma signals differently and it is also related with individual heterogeneity, which is associated with the institution that awarded the diploma. Lastly, $\mu_{i(t),t}$ represents idiosyncratic error. We use a cohort variable dummy defined based on the year of birth of a sample of individuals from the seven largest cities in Colombia to instrument a numeric variables such as education, experience, experience square, s11(s-11), and s16(s-16) consistent with Robert Moffitt's (1993) pseudo panel technique. Given that variables $s11_{i(t),t}$, $s16_{i(t),t}$ and $s17_{i(t),t}$ are dummy variables, these variables are not subject to error correction (Angus Deaton 1985).

The sheepskin effect is reflected by parameters $\hat{\beta}_0$ and $\hat{\beta}_2$, which show that salary levels change from a high school diploma (s ≥ 11) to a university diploma (s ≥ 16). This means that salaries will have a change that is more than proportional when there are sheepskin effects. If parameters $\hat{\beta}_0$ and $\hat{\beta}_2$ are both positive and statistically significant, then a salary increase will be much greater for employees who hold a diploma than the increase associated with an additional year of schooling for a particular cohort. S17 and interaction terms, s11(s-11) and s16(s-16), allow for slope changes in the returns to education (Thomas Hungerford and Gary Solon 1987:176).

We expect that if women are more educated on average, that is, women are majority in the educated Colombian labor market, then the returns to a high school diploma will be greater for women than men – the low signal – and the returns to a university diploma will be greater for men than women – the high signal – in accord to Dale Belman and John Heywood (1997). Empar Pons (2006) also discusses diploma effects by gender and finds that there exist only diploma effects for men in the Spanish labor market where women are more educated than men. However, her results must be taken cautiously due to the unbalanced size of the groups in her sample; actually men are a 33 percent more than women.

Data set and variable description

In Colombia there is no panel survey statistics on household labor supply data. Our sample comes from the National Housing Survey (*NHS*) which consists of a time series of independent and representative cross-sections collected from 1984 to 2000 by the National Agency of Statistics (DANE). Since 2000, the DANE has collected information about the labor market through another mechanism called Continuous Housing Survey. Because of this, information before and after 2000 is not comparable. In each year, the modules of working individuals, personal characteristics, work force, and education were linked. The data for variables as schooling years, age, labor earnings, number of working hours, and kind of occupation were obtained through this link.

We start by defining our sample based on the year of birth and developed pseudo panel data. Table 1 shows an exact definition of the year of birth:

Insert table 1

Table 1 above shows the number of individuals who belong to a given cohort in a particular year. Individuals selected in the sample are aged between 16 and 44. Table 1 also shows that there are not substantial differences between cohorts. The higher difference is in the third cohort with 10 percent in favor of men.

With these cohorts we selected the data in each year and the mean variables for men and women per year were as follows:

Insert table 2

As shown in Table 2 above, the logarithms of real salaries in Colombian pesos are higher for men than for women in all the years of the sample and, although women's salaries have been increasing since 1996, there is still a salary difference between men and women.[4] The average

[4] We use a Colombian Consumer Price Index to adjust to real pesos.

number of years of school is higher for women than for men. This shows a rather discouraging result for women: even if they are more educated than men, women earn less than men.

With regard to the percentage of men and women who have at least completed more than eleven years of education, N_s, the percentage of women is higher than that of men in any year. On the other hand, the percentage of women is not very different from the percentage of men in the overall sample.

Estimation results

Table 3 shows the sheepskin results differentiated by gender. We separated the sample by gender because we considered different separating equilibrium for men and women in the labor market (Spence 1973; Pons 2006). On the other hand, separating women from men allows creating a different labor participation model for each group. The findings of this regression are shown below:

Insert table [3]

Because of segmentation in the Colombian labor market (TH. Macnac 1991; Armando Galvis 2002; Jhon James Mora 2006), dummy variables for each cohort and city were used in all regressions (including the results listed in Table 4 below) and identification in pseudo panel data is done when we use cohort dummies (Moffitt 1993). [5] J-Hansen's over identification test shows that there are no over-identification problems, that is, we have good instruments.

Table 3 above shows that the returns on the successful completion of the tenth grade for women are between 8 percent and 9 percent with a confidence interval of 95 percent. For men, the additional earnings associated with a high school degree are between 6 percent and 7 percent with a confidence interval of 95 percent.

Men's additional years of experience have a return from 4 percent to 6 percent with a confidence interval of 95 percent on completion of the tenth grade of high school. For women, experience results in a return of 3 percent to 4 percent with a confidence interval of 95 percent Meanwhile, the profit from an additional year of experience is higher for women than for men. In this respect, Bernat et al. (2004) arrived at similar results.

With regard to the sheepskin effect on salaries, the findings reveal that women holding a high school degree have higher additional earnings than men, but men holding a university degree obtain higher additional earnings than women. For women, the additional earnings associated with a high school degree are between 12 percent and 18 percent with a 95 percent confidence interval. For men, the additional earnings associated with a high school degree are between 5 percent and 10 percent with a 95 percent of confidence interval. The additional

[5] Jairo Núñez and Fabio Sánchez (2003) discuss the analysis of the effect of schooling degrees on labor earnings in Colombia by using cross-section data and following the theory of human capital. The authors mixed all period in a pooled Mincer equation with cohorts, disregarding measurement errors when individuals are different in each period.

earnings for men and women who hold a university degree are from 31 percent to 68 percent and from 16 percent to 66 percent, respectively, with a 95 percent confidence interval. Following Wald's contrasting approach, these results about the joint statistical significance of high school and university degrees are significant on any level of statistical significance.

While the total number of men is only slightly different from that of women, a calculation of the number of workers holding a degree shows that women represent a majority in the number of both high school and university graduates (Table 2), which accounts for the results shown in Table 3 and is consistent with the minority and majority groups in Belman's and Heywood's sheepskin model.

The last column in table 3 shows sheepskin estimation weighted with square root of the cohort size. If the size of the cohort is very high then we weighted it (Angus Deaton 1985:114). There are no particular reasons for weighting, but we present these results in the spirit of the discussion. The results show that the difference between the sheepskin effects in men and women does not experience substantial change when we ponder by the square root of the cohort size.

Selection bias is a problem that affects all groups in the sheepskin equation because it is possible to find individuals that obtain a diploma with the objective of signaling, but they don't work in the moment of the application of the interview. For this reason we made a contrast of the existence of selection biases based on the methodology proposed by James Heckman (1979) and extended by Jhon James Mora and Juan Muro (2006, 2007) in the pseudo panel case.

Jhon James Mora and Juan Muro (2006, 2007) proposed a contrast of the existence of selection bias by modeling the selection process using an IV-probit consisting of several instrumental variables. The dependent variable in the IV-probit is equal to 1 when an individual was participating in the labor market, or otherwise zero. The years of education, wealth, marital status, and household size were used as independent variables, and dummy variables for cohorts and cities were used as instruments. [6] For each group we estimated a different participation model, and the results of the probit participation model using instrumental variables for men and women are listed below:

Insert table 4

In table 4 above $S_{i(t),t}$ stands for number of years of education, $W_{i(t),t}$ is a dummy for wealth, $M_{i(t),t}$ is a dummy for married and $H_{i(t),t}$ is a dummy for household size. We have 85,540 individuals in the sample consisting of 39,015 women and 46,525 men. For these individuals we estimated a participation model using an IV-probit. With regard to the participation model, the findings show that the participation in the labor market increases as the number of schooling years increases. In the case of females, the wealth and marital status result in a decrease of the participation in the labor market. In the case of males, the wealth results in a decrease of their participation in the labor market. Also we control of industries and incorporate dummys of the

---

[6] Following the work of Luis Eduardo Arango and Carlos Esteban Posada (2001), wealth was incorporated as a dummy variable that takes the value of 1 if the family either owned a house (without any outstanding mortgage debt), or the neighborhood is middle class or higher (according to the official urban classification system). The size of a household was built as a dummy variable that equals either one, if there are more than two people in a household, or zero otherwise.

economic sectors such as agricultural, minery, electricity, manufacturing, building, trade, transports, and financial services.

After incorporating Mills' inverse ratio, obtained from the IV-probit, it was found to be significant for women and men, that is, there is a selection bias problem. The sheepskin effect with selection bias shows that for women, the additional earnings associated with a high school degree are a 12 percent with a 95 percent confidence interval. For men, the additional earnings associated with a high school degree are a 6 percent with a 95 percent confidence interval. The additional earnings for men and women who hold a university degree are a 49 percent and 41 percent, respectively, with a 95 percent confidence interval.

Policy implications and Conclusions

The findings presented in this paper show that there are in deed differences in the additional earnings on schooling degrees for men and women in Colombia with pseudo panel data for the time period from 1996 to 2000. High school degrees translate into greater additional earnings for women than for men, but holding a university degree, on the contrary, represents greater additional earnings for men compared to those of women.

Empar Pons (2006)'s work on gender differences in diplomas did not arrive at these results, but our results show that the discussion of diploma gender differences is more complex. Is it enough that there are differences in the size of groups of individuals to reveal diploma differences between groups? Or on the contrary in the sheepskin model the most important difference between groups is the years of education by group? What variable must be chosen to consider a majority group? Well, this is an open question. We show that if the difference in years of education is a critical aspect, then a diploma difference between men and women arises.

These results have important political implications in Colombia. First, we must discuss whether a diploma could become a barrier of entry to the Colombian labor market to occupations and professions, for which holding a degree is not necessary. Second, implementing a policy to correct the difference between diploma earnings is not easy. The literature reviews show the following types of antidiscriminatory policies: affirmative action, equal pay to equal work and direct subsidies to female work.[7] But it is not clear what you do in the diploma differences with more than one signal. Bernat, et al. (2004) show that most policies are oriented to direct or indirect subsidies to female work in Latin America. If the most important differences between men and women in Colombian labor market are university diploma additional earnings, and additionally a university level is more expensive to employers, antidiscriminatory policy oriented to women's subsidy, i.e. maternity compensation, nursery schools, etc. could compensate these wage differences between men and women.

[7] Affirmative action policy in Colombia exists since 2000, with the Colombian government's law of quotas. According to this law, 30 executive positions in the government should be held for women.

References

Arango, Luis Eduardo and Carlos Esteban Posada. 2001. “Participación laboral en Colombia.” *Borradores de Economía* 217, Banco de la Republica.

Arrow, Kenneth. 1973. “Higher Education as a Filter.” *Journal of Public Economics* 2:193-216.

Becker, G. S. (1964). Human Capital, New York, Columbia.

Belman, Dale and John Heywood. 1991. "Sheepskin Effects in the Returns to Education: An Examination of Women and Minorities." *The Review of Economics and Statistics* 73(4): 720-725.

Belman, Dale and Jhon Heywood. 1997. “Sheepskin Effects by Cohort: Implications of Job Matching in a Signaling Model.” *Oxford Economic Papers* 49(4): 623-637.

Bernat, Luisa Fernanda, Rocio Ribero, R and Jaime Tenjo. 2004. “Evolution of Salary Differences between Men and Women in Six Latin American Countries,” In Cludia Piras (ed) Women at Work: Challenges for Latin America, pp.139-185. Inter American Development Bank.

Culler, Steven, Edmund Becker, and Robert Ohsteldt. 1987. “Sex Differences in the Economic Advantages of Physician Board Certification.” *Southern Economic Journal* 54(2): 343-350.

Deaton, Angus. 1985. “Panel Data from Time Series of Cross-Sections.” *Journal of Econometrics* 30(1-2):109-126.

Freeman, Richard. 1977. “The New Job Market for Black Academicians.” *Industrial and Labor Relations Review* 30(2): 161-174.

Galvis, Armando. 2002. “Integración regional de los mercados laborales en Colombia, 1984-2000*.” Documentos de trabajo sobre economía regional* 27.

Golbe, Devra. 1985. “Imperfect Signalling, Affirmative Action, and Black-White Differentials.” *Southern Economic Journal* 51(3): 842-848.

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

Hungerford, Thomas and Gary Solon. 1987. “Sheepskin Effects in the Returns to Education”, *Review of Economics and Statistics* 69(1): 175-177.

Macnac, Th. 1991. “Segmented or Competitive Labor Markets.” *Econometrica* 59(1): 165-187.

Moffitt, Robert. 1993. “Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross-Sections.” *Journal of Econometrics* 59(1-2): 99-123.

Maldonado, D. (2000).”Diferenciales salariales entre los trabajadores del sector publico y del sector privado en Colombia en 1997”, *Desarrollo y Sociedad*, No 45, marzo.

Mora, Jhon James. 2006. “Screening and Sheepskin effects in Colombia” *Colombian Economic Journal* 1: 95-108.

Mora, Jhon James. 2006. “Dinámica de salarios urbanos en Colombia: segmentación o integración en secundaria y universidad” in Jairo Guillermo Isaza and Luís Fernando Ramírez (eds.) *Empleo, pobreza y desigualdad: Una mirada desde la investigación universitaria*, pp.13-35.Universidad de la Salle.

Mora, Jhon James and Juan Muro. 2006. “Bias Selection in Pseudo Panel Data.” International Conference on Mathematical and Statistical Modeling, UCL (Spain).

Mora, Jhon James and Juan Muro. 2007. “Testing for sample selection bias in pseudo panels: theory and Monte Carlo” *Borradores de Economía y Finanzas* 11.

Núñez, Jairo and Fabio Sánchez, F. 2003. “A dynamic analysis of human capital, female work force participation, returns to education and changes in household structure in urban Colombia, 1976-1998.” *Colombian Economic Journal* 1: 110-149.

Pons Empar. 2006. “Diploma Effects by Gender in the Spanish Labour Market*.” Labour* 20(1):139 157.

Spence, Michael. 2002. “Signaling in Retrospect and the Informational Structure of Markets.” *American Economic Review* 92(3): 434-59.

Spence, Michael. 1973. “Job Market Signaling.” *Quarterly Journal of Economics* 87:355-74.

TABLES

Table 1. Definitions of Cohorts and number of individuals in each cohort.

Cohort / Year	1996	1997	1998	1999	2000
Cohort 1 (Born between 1976 and 1980)	1,561	1,761	1,921	1,747	1,885
Cohort 2 (Born between 1971 and 1975)	2,562	2,682	2,492	2,190	1,945
Cohort 3 (Born between 1966 and 1970)	2,700	2,631	2,385	2,010	1,831
Cohort 4 (Born between 1961 and 1965)	2,276	2,316	1,983	1,819	1,730
Cohort 5 (Born between 1956 and 1960)	1,974	1,734	1,674	1,354	1,266

Source: authors' calculations

Table 2. Mean variables for females and males

Variable/ Sex	1996		1997		1998		1999		2000		Total Group	
	F	M	F	M	F	M	F	M	F	M	F	M
Lwhr	6.60	6.75	6.89	7.03	7.08	7.19	7.24	7.38	7.28	7.38	7.00	7.12
S	9.72	9.22	10.04	9.62	10.16	9.66	10.26	9.96	10.42	10.06	10.11	9.67
Exp	12.64	13.12	12.75	13.10	13.29	13.84	14.11	14.35	14.56	15.07	13.42	13.80
N_S (%)	0.55	0.45	0.59	0.41	0.60	0.40	0.62	0.38	0.63	0.37	0.60	0.40
N (%)	0.46	0.54	0.48	0.52	0.48	0.52	0.50	0.50	0.50	0.50	0.48	0.52

Source: authors' calculations

Table 3. Sheepskin effects by gender

Variable	Female	Male	Female Weighted	Male Weighted
Constant	5.327275	5.03494	5.172629	4.991561
	(0.0699982)	(0.0885015)	(0.0505183)	(0.0866764)
$S_{i(t),t}$	0.0892025	0.0685215	0.0999448	0.0789916
	(0.003769)	(0.0033388)	(0.003808)	(0.0033839)
$Exp_{(t),t}$	0.0323303	0.051129	0.041292	0.0609115
	(0.0033029)	(0.003203)	(0.0033715)	(0.0032245)
$Exp^2_{i(t),t}$	-0.0002941	-0.0007972	-0.0002725	-0.0007856
	(0.000069)	(0.0000697)	(0.0000704)	(0.000071)
Sheepskin effects				
$S11_{i(t),t}$	0.1505872	0.084726	0.1520493	0.0896793
	(0.0152101)	(0.0128296)	(0.0157681)	(0.013106)
$S16_{i(t),t}$	0.415288	0.5026064	0.4352614	0.5133008
	(0.126091)	(0.0934782)	(0.1299262)	(0.0912715)
$S11(s-11)_{i(t),t}$	0.112275	0.1320614	0.1114798	0.1315293
	(0.0038624)	(0.0036909)	(0.0039495)	(0.003773)
$S16(s-16)_{i(t),t}$	-0.1721396	-0.1836564	-0.1807363	-0.185669
	(0.0519283)	(0.0324575)	(0.0536532)	(0.0318361
$S17_{i(t),t}$	-0.1644979	-0.207286	-0.1790801	-0.2286654
	(0.0854016)	(0.076489)	(0.087209)	(0.0753112)
Wald Test	$\chi^2(2)=109.22$	$\chi^2(2)=72.44$	$\chi^2(2)=104.65$	$\chi^2(2)=78.29$
F	513.42	377.49	511.28	367.12
R^2	0.5594	0.4775	0.5541	0.4765
Sample size	24,410	26,019	24,410	26,019
Overidentification, J-Hansen	0.000	0.000	0.000	0.000

Robust standard errors in parentheses. Source: authors' calculations

Table 4. Labor force participation and selection bias

Variable	Male-Sel	Male-bias	Female-Sel	Female-bias
$S_{i(t),t}$	0.074 (0.011)	0.109 (0.005)	0.158 (0.011)	0.130 (0.006)
$W_{i(t),t}$	-0.608 (0.061)		-0.256 (0.036)	
$H_{i(t),t}$	0.041 (0.023)	-0.074 (0.013)		-0.054 (0.012)
$M_{i(t),t}$		0.092 (0.008)	-0.344 (0.014)	
$Exp_{i(t),t}$		0.046 (0.003)		0.031 (0.003)
$Exp^2_{i(t),t}$		-0.001 (0.000)		-0.000 (0.000)
Sheepskin effects				
$S11_{i(t),t}$		0.064 (0.013)		0.124 (0.016)
$S16_{i(t),t}$		0.489 (0.092)		0.412 (0.126)
$S11(s-11)_{i(t),t}$		0.102 (0.005)		0.073 (0.006)
$S16(s-16)_{i(t),t}$		-0.191 (0.032)		-0.180 (0.052)
$S17_{i(t),t}$		-0.209 (0.075)		-0.164 (0.085)
Secagri		-0.194 (0.031)		0.172 (0.050)
Secmin		0.013 (0.004)		0.020 (0.006)
Secman		-0.050 (0.010)		0.099 (0.010)
Secelec		0.197 (0.033)		0.325 (0.047)
Secbild		-0.054 (0.014)		0.219 (0.033)
Sectrad		-0.120 (0.011)		0.074 (0.010)
Sectrans		-0.064 (0.015)		0.201 (0.026)
SecFinS		-0.036 (0.013)		0.203 (0.014)
Selection bias $_{i(t),t}$		-0.860 (0.075)		-0.440 (0.050)
Constant	-0.555 (0.108)	6.055 (0.090)	-1.048 (0.110)	5.288 (0.089)
Year effects	No	Yes	No	Yes
F		372.875		499.096
R^2		0.484		0.560
Sample size	46,525	26,019	39,015	24,410
Overidentificatión, J-Hansen		0.000		0.000

Robust standard errors in parentheses. Source: authors' calculations