Gender differences in sheepskin effects: Evidence from synthetic cohort data

Jhon James Mora and Juan Muro

Jhon James Mora
Departamento de Economía.
Universidad ICESI.
Calle 18 No. 122-135. Cali (Colombia).
Phone number: +57+25552334; Fax number :+57+25551441;
e-mail: jjmora@icesi.edu.co

Juan Muro (corresponding autor)
Departamento de Economía.
Facultad de Ciencias Económicas y Empresariales. Universidad de Alcalá.
Plaza de la Victoria 2. Alcalá de Henares 28802. Madrid (Spain).
Phone number: +34+918854270; Fax number :+34+918854201;
e-mail: juan.muro@uah.es

Abstract

We find evidence of gender differences in diploma effects with a model of sheepskin effects estimated with Colombian synthetic cohort data for the period 1996-2000. Our results show a significant and distinctive gender effect of high school and university degrees. While additional earnings associated with a high school degree are higher for women than for men, university degree effects are higher for men in accordance with the heterogeneous-group signalling model.

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

.

1. Introduction

The labour market in Colombia showed from 1981 to 2000 a reduction of the traditional gender gap in labour force participation. While the female labour force participation rate went up from a low 37% in 1981 to a 51% in 1998 the male labour force participation rate remained stable around a 74%. In the same period, however, the returns to education by gender exhibited a constant gap with the returns for women around a 2% above those for men. The evolution has taken place in a labour market characterized by a high component of screening in the hiring process. In a labour market with different collectives, majorities and minorities defined by gender and education achievement, the effect of a signalling process 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 wages by gender we use Hungerford-Solon (1987) sheepskin effect equation. Under signalling Belman-Heywood (1991) extension of the sheepskin model allows us to analyse the gender differences. 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.

Empirically, our paper differs from other sheepskin equation applied analysis in incorporating the use of pseudo panel data by groups. 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 selection bias test is concerned we find a problem of selection bias in both men and women equations.

2. Model and Database

Hungerford-Solon (1987) found evidence to confirm that “wages will rise faster with each extra year of education when an extra year also conveys a certificate”. The results could change if instead of a sample of homogeneous individuals there were different groups under analysis. Golbe (1985) and Belman and Heywood (1991) show that the statistical discrimination in the signalling model could translate into higher earnings for the minorities than for the majorities in response to a signal of high productivity.

Mora (2003) and Mora and Muro (2008) discuss the sheepskin equation in Colombia but they don’t consider differences by groups. Estimating the effect of diplomas using pseudo panel data entails the use of the following model:

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

where $\text{LnWh}_{i(t),t}$ $S_{i(t),t}$ and $exp_{i(t),t}$ are the logarithm of wages per hour, years of schooling, and potential experience (age-s-6), respectively, for individual i in time t; Subindex i(t) denotes we observe different individuals in each time period of time. $S11_{i(t),t}$, $S16_{i(t),t}$ are dummies for completed 11 years of schooling or more and completed 16 years of schooling or more, respectively; sheepskin effects are represented by parameters β_0, β_2. $\nu_{i(t)}$ represents the deviation of the effect of the cohort after breaking down fixed individual effects. 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.[i]

If women are the majority in the educated Colombian labour 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.

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).[ii]

Insert table 1

Table 1 above displays the number of individuals aged between 16 and 44 who belong to a given cohort in a particular year. It shows no substantial differences between cohorts.

Insert table 2

Descriptive statistics in Table 2 above show 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, the salary gap between men and women. The average number of years of school is higher for women than for men and the percentage of women who have at least completed more than eleven years of education is higher than that of men in all years.

3. Empirical Results and Conclusions

Table 3 shows the sheepskin results by gender. Using subsamples by gender reflects not only different equilibrium for men and women in the labour market but allows creating a different labour participation model for each group.

Insert table [3]

Because of segmentation in the Colombian labour market (Macnac 1991; Galvis 2002; Mora and Muro 2015), 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). J-Hansen's over identification test shows that there are no over-identification problems.

Table 3 above shows that the returns on the successful completion of the tenth grade for women are between 8% and 9% with a confidence interval of 95%. For men, the additional earnings associated with a high school degree are between 6% and 7%. Men's additional years of experience have a return from 4% to 6% on completion of the tenth grade of high school.

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% and 18%. For men, the additional earnings associated with a high school degree are between 5 % and 10 %. The additional earnings for men and women who hold a university degree are from 31 % to 68 % and from 16 % to 66 %, respectively. Following Wald's testing 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 by the 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 weight 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 signalling, but they don't work in the moment of the application of the interview. For this reason we made a test of the existence of selection biases based on the methodology proposed by Heckman (1979) and extended by Mora and Muro (2014) in the pseudo panel case. Following Mora and Muro (2014) we model the labour force participation for men and women,

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. With regard to the participation model, the findings show that the participation in the labour 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 labour market. In the case of males, the wealth results in a decrease of their participation in the labour market. Also we control of industries and incorporate dummies of the economic sectors such as agricultural, mining, electricity, manufacturing, building, trade, transports, and financial services.

After incorporating Mills' inverse ratio 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%. For men, the additional earnings associated with a high school degree are a 6%.

The additional earnings for men and women who hold a university degree are a 49% and 41%, respectively.

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 labour 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 anti-discriminatory policies: affirmative action, equal pay to equal work and direct subsidies to female work.[1] 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 labour market are university diploma additional earnings, and additionally a university level is more expensive to employers, anti-discriminatory policy oriented to women's subsidy, i.e. maternity compensation, nursery schools, etc. could compensate these wage differences between men and women.

References

Arango, L. E. and Posada, C.E. 2001. “Participación laboral en Colombia”. *Borradores de Economía* 217. Banco de la República.

---

[1] 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.

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

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

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

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

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

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

Moffitt, R. 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 público y del sector privado en Colombia en 1997”. *Desarrollo y Sociedad*, No 45, marzo.

Macnac, T,h .(1991) "Segmented or Competitive labor Markets". Econometrica,59(1):1 65-187.

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

Mora, J. J, and Muro, J (2008) "Sheepskin effects by cohorts in Colombia", International Journal of Manpower, Vol. 29 Iss: 2, pp.111 - 121

Mora, J. J, and Muro, J. 2014 “Consistent Estimation in Pseudo Panels in the Presence of Selection Bias” *Economics: The Open-Access, Open-Assessment E-Journal*, 8 (2014-43): 1-25.

Mora, J. J, and Muro, J. 2015. “Labor market segmentation in a developing country”. Working paper.

Núñez, J. and 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.

TABLES

Table 1. Cohorts' definition 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

Table 2. Descriptive statistics by gender.

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

Table 3. Sheepskin effects by gender

Variable	Female	Male	Female Weighted	Male Weighted
Constant	5.327	5.035	5.173	4.991
	(0.0699)	(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	0.109	0.158	0.130
	(0.011)	(0.005)	(0.011)	(0.006)
$W_{i(t),t}$	-0.608		-0.256	
	(0.061)		(0.036)	
$H_{i(t),t}$	0.041	-0.074		-0.054
	(0.023)	(0.013)		(0.012)
$M_{i(t),t}$		0.092	-0.344	
		(0.008)	(0.014)	
$Exp_{i(t),t}$		0.046		0.031
		(0.003)		(0.003)
$Exp^2_{i(t),t}$		-0.001		-0.000
		(0.000)		(0.000)
Sheepskin effects				
$S11_{i(t),t}$		0.064		0.124
		(0.013)		(0.016)
$S16_{i(t),t}$		0.489		0.412
		(0.092)		(0.126)
$S11(s-11)_{i(t),t}$		0.102		0.073
		(0.005)		(0.006)
$S16(s-16)_{i(t),t}$		-0.191		-0.180
		(0.032)		(0.052)
$S17_{i(t),t}$		-0.209		-0.164
		(0.075)		(0.085)
Selection bias $_{i(t),t}$		-0.860		-0.440
		(0.075)		(0.050)
Constant	-0.555	6.055	-1.048	5.288
	(0.108)	(0.090)	(0.110)	(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.

---

[i] We use cohort-dummy variables, defined on the year of birth and residence in the seven largest cities in Colombia, as matching instruments.

[ii] Since 2000 the DANE has collected information about the labour 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, labour earnings, number of working hours, and kind of occupation were obtained through this link.