Assignment 5 Undirected: Part 2

Daniel Sánchez Pazmiño

This document presents my code and analysis for my replication and commenting of Linthon-Delgado and Méndez-Heras (2022).

Preliminaries

I load my libraries and then my data and regressions from the `undirected-part2.R` script that is included within the `.rar` file that I uploaded along with this document.

The paper

Linthon-Delgado and Méndez-Heras (2022) is a paper published in the Mexican Journal of Economics and Finance which estimates the gender wage gap in Ecuador using the employment survey from September 2020. It uses a Oaxaca-Blinder decomposition after estimating Mincer earnings functions.

In this asssignment, I replicate and expand Linthon-Delgado and Méndez-Heras (2022) by estimating the same earnings functions but with 2011 and 2012 data, which is not subject to the pandemic's effects. Further, I want to know if the gender wage differential changed from 2011 to 2012, where there was a significant policy change affecting women only. The policy affected women only in that it increased the number of months that a woman who recently gave birth can enjoy a paid break of two hours a day for breastfeeding. Further, the reform required companies to install "breastfeeding stations" in their offices if the company exceeded an employee number threshold. I would like to investigate if employers reduced base earnings for women to account for these new expenses.

Table 1: Descriptive statistics for the continuous variables in the sample

	Mean				Median			
	Male		Female		Male		Female	
	2011	2012	2011	2012	2011	2012	2011	2012
Income	380.98	410.75	322.28	349.41	280.00	300.00	230.00	250.00
Ln Income	5.56	5.65	5.29	5.37	5.63	5.70	5.44	5.52
Age	43.22	43.34	42.10	42.73	42.00	42.00	41.00	42.00
Experience	14.69	14.87	11.10	11.61	10.00	10.00	5.00	6.00

The data

I use the data from the Ecuadorian analogous to Canada's employment survey (LFS), which is the National Surveyt of Employment, Unemployment & Underemployment (ENEMDU, for its initials in Spanish). I use the microdata available to the public at the Ecuadorian National Institute of Statistics and Censuses' (INEC, for its initials in Spanish) website. In my code, I unzip the data for 2011 and 2012 (the December survey) from its `.rar` presentation as downloaded from the website and then proceed to use it for my analysis.

I present some descriptive statistics below, as well as counts and percentages for each of the categorical variable I use in the study.

Methods

Mincer earnings function

The Mincer earnings function develops the natural logarithm of earnings as a function of observable characteristics at the invidual level. I implement such earnings functions based on the common controls and add race, urban/rural and public/private groups to further control for differences. The basic equation is:

$$ln(Y_i) = \beta_0 + \beta_1 Educ_i + \beta_2 Exp_i + \beta_3 Exp_i^2 + \beta_4 X_i + \epsilon_i, \quad (1)$$

where Y_i represents the earnings of individual i, $Educ_i$ is the years of education, Exp_i is the years of work experience, Exp_i^2 is the squared work experience, and X_i is a vector of other individual characteristics, such as age, gender, and occupation. The β coefficients represent the marginal effects of each variable on earnings.

Importantly, I assume experience has diminishing marginal returns to earnings. This is reflected in the quadratic term, which captures the idea that the effect of experience on earnings

Table 2: Percentages (%) for the categorical variables in the sample

		Male		Female	
		2011	2012	2011	2012
Schooling	None	5.37	4.61	6.16	5.55
	Alphabetization Center	0.40	0.57	0.51	0.85
	Basic	54.14	52.81	43.18	43.39
	Higher	17.25	17.75	27.53	27.40
	Secondary	22.84	24.26	22.63	22.81
Marital Status	Single	23.64	25.09	25.08	26.76
	Common Law	20.57	21.55	12.41	12.72
	Divorced	1.67	1.69	4.79	4.67
	Married	45.96	42.82	37.88	35.91
	Separated	5.25	6.15	11.86	11.64
	Widowed	2.92	2.70	7.97	8.30
Race	Mestizo	79.62	82.97	84.05	85.48
	Afroecuadorian/Black	4.19	2.85	3.67	3.33
	Indigenous	6.45	5.27	6.43	5.83
	Montubio	6.02	6.53	2.16	2.18
	Mulatto	1.19	0.90	1.21	1.38
	Other	0.23	0.07	0.15	0.02
	White	2.29	1.42	2.33	1.78
Urban/Rural	Rural	43.97	42.50	31.34	31.67
	Urban	56.03	57.50	68.66	68.33
Private/Public Employment	Not Public	90.90	90.58	85.58	85.47
	Public	9.10	9.42	14.42	14.53

diminishes as experience increases. I estimate the earnings function using OLS, but later test for sticky floor and glass ceiling effects by using quantile regression at the bottom 10% and at the top 95%.

Oaxaca-Blinder decomposition

The Oaxaca-Blinder method is used to decompose differences in economic outcomes between two groups into portions that can be explained by differences in observable characteristics (e.g., education, experience, industry, etc.) and portions that cannot be explained by such differences (e.g., discrimination, unobserved factors).

The method decomposes the difference in means of the outcome variable between two groups (e.g., men and women) into two parts: the part that can be explained by differences in observable characteristics and the part that cannot. The decomposition can be expressed as:

$$\bar{Y}_1 - \bar{Y}_0 = \sum_{i=1}^{K} \beta_i \left(X_{1i} - X_{0i} \right) + \sum_{i=1}^{K} \delta_i \bar{X}_1 + u_1 - \sum_{i=1}^{K} \delta_i \bar{X}_0 - u_0 \tag{2}$$

where $\overline{Y}1$ and $\overline{Y}0$ are the mean values of the outcome variable for the two groups, $X1i$ and $X0i$ are the mean values of the ith characteristic for the two groups, β_i is the coefficient for the ith characteristic in a regression of the outcome variable on the characteristics for the pooled sample, $\overline{X}_1$ and $\overline{X}_0$ are the mean values of the ith characteristic for the two groups, and u_1 and u_0 are the unexplained portions of the difference in means for the two groups.

The first part of the equation represents the explained portion of the difference in means, which is due to differences in observable characteristics between the two groups. The second part of the equation represents the unexplained portion of the difference in means, which may be due to discrimination, unobserved characteristics, or measurement error.

The Oaxaca-Blinder method estimates the explained and unexplained portions of the difference in means by estimating two separate regression equations for the two groups and comparing the predicted values from each equation. In this paper, I implement the two-way Oaxaca-Blinder filter using the *oaxaca* R package.

Results

Estimated earnings functions

Below, I include the results for key earnings functions that I estimated with the above methodology.

Table 3: Estimated earnings functions

	(1)	(2)	(3)	(4)
(Intercept)	4.800***	4.814***	1.174***	1.183***
	(0.036)	(0.036)	(0.051)	(0.052)
year2012	0.108***	0.079***	0.171***	0.152***
	(0.011)	(0.013)	(0.016)	(0.021)
sexFemale	-0.393***	-0.425***	-1.376***	-1.397***
	(0.012)	(0.017)	(0.017)	(0.024)
age	-0.003***	-0.003***	-0.002**	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
schoolingAlphabetization Center	0.001	0.002	0.079	0.079
	(0.081)	(0.081)	(0.106)	(0.106)
schoolingBasic	0.289***	0.289***	0.281***	0.282***
	(0.028)	(0.028)	(0.038)	(0.038)
schoolingHigher	0.952***	0.952***	1.048***	1.048***
	(0.032)	(0.032)	(0.044)	(0.044)
schoolingSecondary	0.546***	0.547***	0.571***	0.571***
	(0.031)	(0.031)	(0.043)	(0.043)
mar_statCommon Law	0.125***	0.125***	-0.185***	-0.185***
	(0.018)	(0.018)	(0.027)	(0.027)
mar_statDivorced	0.178***	0.178***	-0.184***	-0.184***
	(0.036)	(0.036)	(0.047)	(0.047)
mar_statMarried	0.192***	0.192***	-0.110***	-0.110***
	(0.015)	(0.015)	(0.023)	(0.023)
mar_statSeparated	0.104***	0.105***	-0.293***	-0.292***
	(0.021)	(0.021)	(0.032)	(0.032)
mar_statWidowed	0.008	0.008	-0.356***	-0.356***
	(0.031)	(0.031)	(0.044)	(0.044)
raceAfroecuadorian/Black	-0.043	-0.044	-0.131***	-0.131***
	(0.030)	(0.030)	(0.046)	(0.046)
raceIndigenous	-0.261***	-0.261***	-0.290***	-0.290***
	(0.026)	(0.026)	(0.033)	(0.033)
raceMontubio	-0.010	-0.010	-0.115***	-0.115***
	(0.024)	(0.024)	(0.036)	(0.036)
raceMulatto	-0.062	-0.064	-0.121	-0.122
	(0.049)	(0.049)	(0.075)	(0.075)
raceOther	0.003	0.004	-0.192	-0.191
	(0.164)	(0.165)	(0.187)	(0.187)
raceWhite	0.050	0.048	0.026	0.025
	(0.049)	(0.049)	(0.060)	(0.060)

	(1)	(2)	(3)	(4)
urbanUrban	0.347***	0.348***	0.662***	0.662***
	(0.012)	(0.012)	(0.018)	(0.018)
publicPublic	0.706***	0.706***	0.636***	0.637***
	(0.016)	(0.016)	(0.025)	(0.025)
exp	0.021***	0.021***	0.025***	0.025***
	(0.001)	(0.001)	(0.002)	(0.002)
I(exp^2)	0.000***	0.000***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
year2012 × sexFemale		0.063***		0.042
		(0.022)		(0.032)
Num.Obs.	39504	39504	39504	39504
R2	0.357	0.357	0.395	0.395
R2 Adj.	0.356	0.357	0.395	0.395
AIC	543486.1	543471.8	260687.8	260686.5
BIC	543692.1	543686.4	260893.8	260901.1
Log.Lik.	-51852.198	-51844.046	-67488.304	-67486.649
RMSE	0.79	0.79	1.16	1.16
Std.Errors	HC3	HC3	HC3	HC3
Dependent variable	Monthly income	Monthly income	Hourly income	Hourly income

Note: ^^ * p < 0.1, ** p < 0.05, *** p < 0.01

Note: ^^ HC3 heteroskedasticity-robust std. errors are presented in parentheses.

This table presents the estimated earnings functions of four models (1)-(4). Each model is fitted with different covariates that may affect earnings, such as year, sex, age, schooling, marital status, race, urbanization, and work sector. The coefficients of each covariate and interaction terms are reported, along with their standard errors, t-values, and significance levels.

The year2012 variable is also significant with a positive coefficient, implying that average earnings have increased over time. The variable Female is negative and statistically significant, suggesting that females earn less on average than males. The age variable is also significant and negative, indicating that earnings decrease with age. The schooling variables have positive and significant coefficients, indicating that more schooling leads to higher earnings. The last two variables, urbanUrban and publicPublic, are also significant, suggesting that working in urban areas and in the public sector leads to higher earnings.

Model (2) includes interaction terms between year2012 and sexFemale. The coefficient of the interaction term is positive and statistically significant, indicating that the gender earnings gap has decreased over time, contrary to what was hypothesized. All of these models include monthly earnings as the dependent variable. However, Models (3) and (4) use hourly earnings

assuming a 40 hour workweek (the standard for Ecuadorian law calculations), and repeat the analysis.

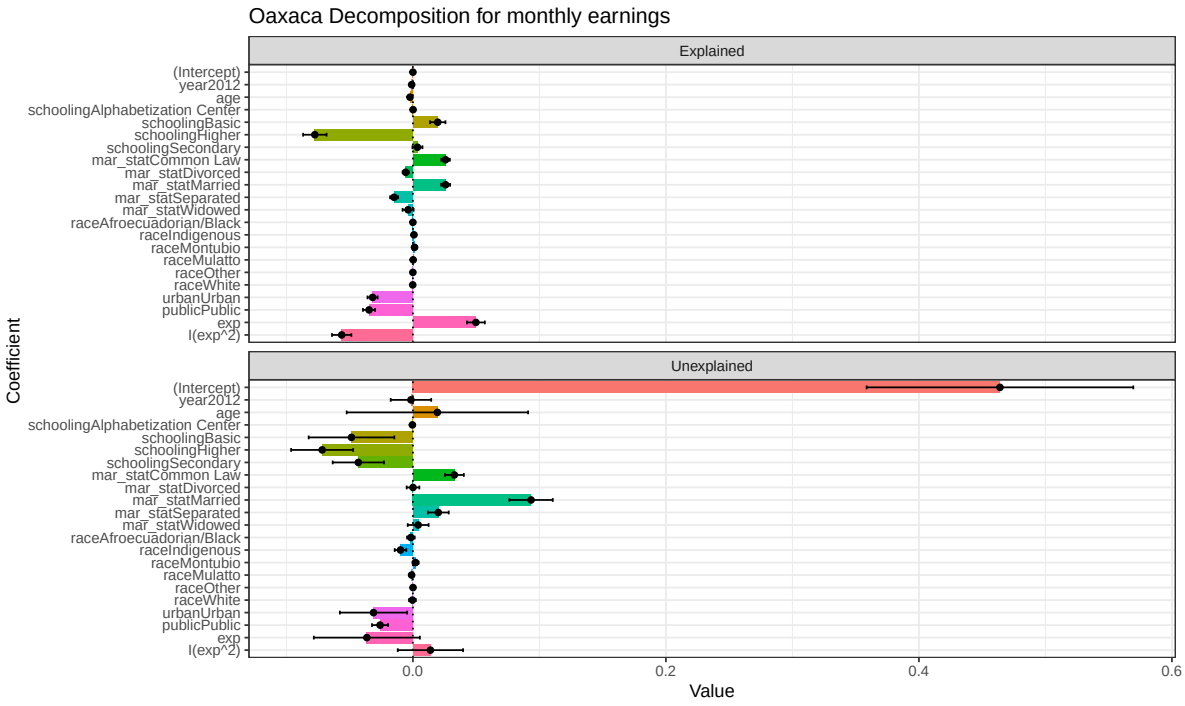
Model (3) includes a quadratic term of experience (exp) in addition to the covariates used in Model (1). The coefficient of exp is positive and significant, implying that earnings increase with experience. The quadratic term I(exp^2) is negative and significant, suggesting that the rate of increase in earnings decreases as experience increases.

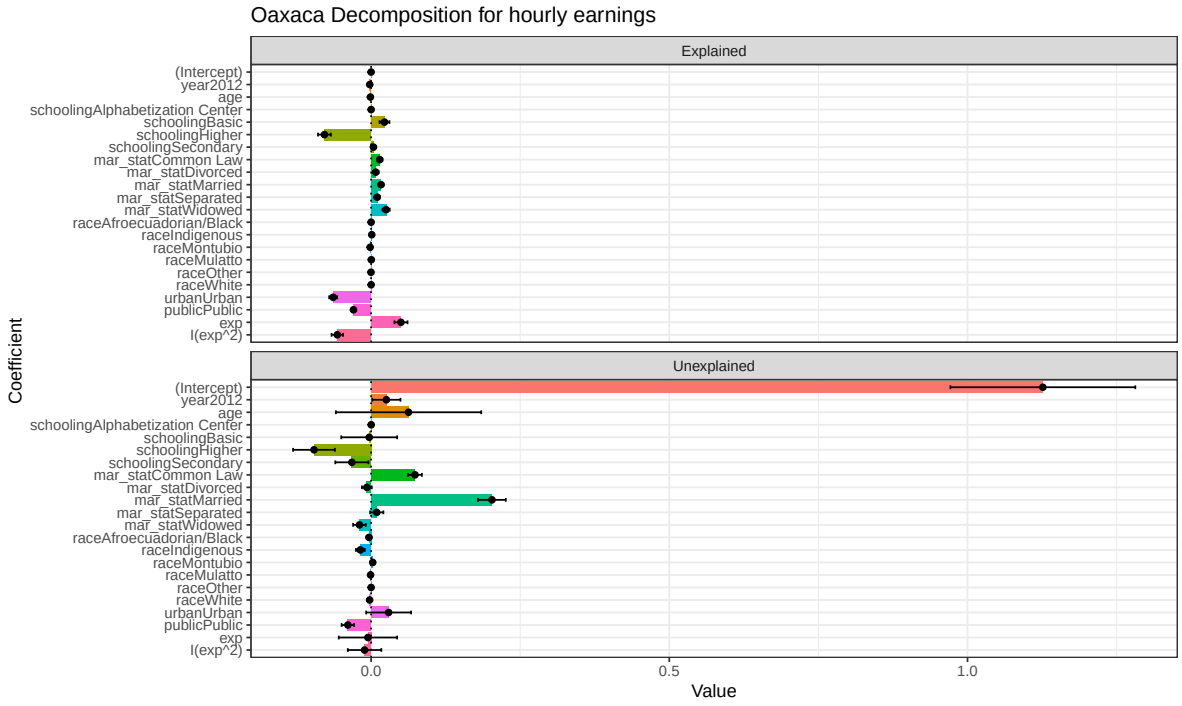
Model (4) includes an interaction term between year2012 and sexFemale in addition to the covariates used in Model (3). The coefficient of the interaction term is not statistically significant, indicating that the gender earnings gap has not changed significantly after controlling for experience and its quadratic term.

Overall, the R^2 values of the models suggest that the covariates included in the models explain around 36-40% of the variation in earnings. The standard errors of the coefficients are calculated using the HC3 method, which is a sample size adjustment to the common heteroskedasticity-adjusted errors HC1 which are default in Stata.

Plots for Oaxaca-Blinder estimations

Below, I present the plots which give the results for the Oaxaca-Blinder decomposition done to a regression model which follows the Mincer earnings function, but does not include gender as two separate regressions are estimated.





What these two decompositions imply is that (1) the discrimination component of women's intercept is highest, which can either mean that discrimination is very high in the country or that we lack enough variables to explain the gap. Further, it is seen that the most unexplained part of these regressions lie around the return of marriage. While married men make more than single men, the same difference appears for women, but it is much greater.

References

Linthon-Delgado, Diego Emilio, and Lizethe Berenice Méndez-Heras. 2022. "Descomposición de La Brecha Salarial de género En El Ecuador." *Revista Mexicana de Economía y Finanzas* 17 (1).