ASSOCIAÇÃO BRASILEIRA DE ESTATÍSTICA
XIV EBEB – Encontro Brasileiro de Estatística Bayesiana
Rio de Janeiro, de 05 a 09 de março de 2018

Is there a wage discrimination in IT careers in Santa Catarina?

Schmidt, A.V.[1]; Pozzobon, F.[2]
[1]Undergraduate Student, Departamento de Ciências Econômicas da Universidade do Estado de Santa Catarina
[2]Professor, Departamento de Ciências Econômicas da Universidade do Estado de Santa Catarina
Contact: aishameriane.schmidt@posgrad.ufsc.br

Poster number 42

Introduction

The lack of female representation in sciences, technology, engineering and mathematics (STEM) careers has been discussed in the literature due to its implications to women and society as a whole. For example, the low participation of women in some clinical trials led to the posterior discovery of different drug effects in the female population when compared to the male clinical-trial participants of the studies. Also, gender imbalance can cause wage gaps between women and men.

STEM careers usually have a higher wage prize in comparison to other professions, due to its higher demand for qualified professionals which still have a low supply, specially in Brazil and in other developing economies. Thus, the low female participation in these areas increases the overall wage gap between men and women - studies point out that women wages are, on average, 30% lower than men's.

The Human Capital Theory establishes a connection between productivity and the workers characteristics (innate or acquired). In this sense, we can think the set of characteristics like years of schooling, age, years at the same occupation, etc, as one factor driving of wages in the labor force. The Becker view of human capital stablishes that human capital is directly connected with production, i.e., human capital increases the worker's productivity in all tasks (ACEMOGLU & AUTOR, 2011).

Introduction (cont.)

The Human Capital Theory enters in an econometric model by the Mincer equations, which allows us to have a quantitative analysis of the effect in salary (which represents the productivity in the neoclassical sense of factor remuneration) from the individual's characteristics. This approach can be expanded by the Oaxaca decomposition to evaluate the wage discrimination between two or more groups.

In the information technology (IT) sector, with professions like computer programmer, there is no expectation that physical attributes would influence over productivity. Thus, there should be no wage gap between men and women in this areas. Nonetheless, Bonini and Pozzobon, using data from the workers in IT industry for the year of 2011, found out that, controlling for age, scholarity and time at the job, women still had salaries that were approximately 15% lower than their male counterparts.

Goals

Using data from the IT sector in 2016 and the multiple linear regression model, our goal is to estimate the gender gap in salaries. As parameters priors the results from Bonini and Pozzobon (2016) are used.

Model and prior specification

The econometric model follows Bonini and Pozzobon (2016) specification where the natural logarithm of the hourly salary is the dependent variable :

$$\ln(wage) = \beta_0 + \beta_1 \cdot S_1 + \beta_2 \cdot S_2 + \beta_3 \cdot S_3 + \beta_4 \cdot Age + \beta_5 \cdot Age^2 + \beta_6 \cdot Time + \beta_7 \cdot Gender + \varepsilon$$

Where:

- ln(wage) is the vector of the natural logarithm of the worker's wages (in Brazilian Reals).
- β_0 is the model intercept, which represents the heterogeneity due to non-observable factors that are common to all workers;
- S_1, S_2 and S_3 are dummy variables representing the maximum scholarity of the worker: incomplete college, complete college and graduate school. When the three variables are equal to zero, the worker had studied until complete high school;
- Age is the worker's age, in years and due to the assumption of decreasing returns of salary w.r.t. age, the square vector of ages, Age^2, was also included in the model;
- $Gender$ is a binary variable which assumes 0 for male workers and 1 if the worker is a woman;
- It is also assumed that the components of the error vector ε are independent, normally distributed, i.e., $\varepsilon \sim N(0_N, h^{-1} I_N)$, and non correlated with the other independent variables.

Prior distributions for the parameters are given by $\beta|h \sim N\left(\underline{\beta}, h^{-1}\underline{V}\right)$, $h \sim Gamma\left(\underline{s}^{-2}, \underline{\nu}\right)$ and their conjoint prior distribution is $\beta, h \sim NG\left(\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{\nu}\right)$ (where G means Gamma and NG is the Normal-Gamma distributions).

We used the multiple linear regression model with conjugate normal-gamma prior (Koop, 2003) and the quantities estimated from Bonini and Pozzobon (2016) were used as means for the coefficients prior distributions. We adopted high values for their variances (10^2).

Since the logarithm of the wages is almost totally comprised between 3.0 and 5.5 and by assumption the errors are normally distributed, we used $s^2 = 1$, implying a prior value of 1 to the precision h. Since our guess about h is rudimentary, we are giving it little weight (in comparison to the data information) through the prior for the degrees of freedom, i.e., we used $\underline{\nu} = 44$.

Data

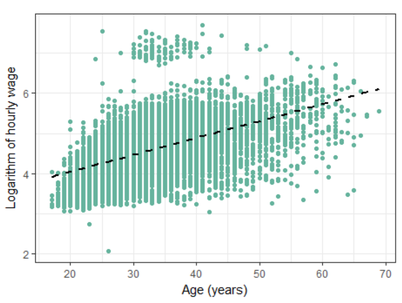
Using data from the Brazilian Labor Office (RAIS data) for the Santa Catarina State, we selected the entire population of 10.919 workers registered with IT occupations in the year of 2016. Descriptives are shown bellow.

	ln(wage) (R$)	Time in the job (mo)	Age (yrs)
Minimum	2.067	0.00	17.00
1st quartile	4.112	14.80	27.00
Median	4.578	36.90	31.00
Mean	4.580	55.41	32.64
3th quartile	5.015	76.50	37.00
Maximum	7.707	507.80	69.00
Std. Deviation	0.656	62.87	8.11

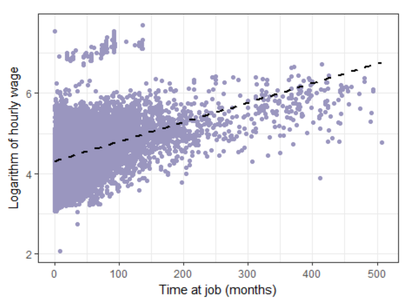
Table 1: Descriptives of the quantitative variables entered in the model.

	n	%
Male	8792	80.59
Incomplete college ed.	1157	10.61
Complete college ed.	8339	76.44
Graduate school ed.	184	1.69

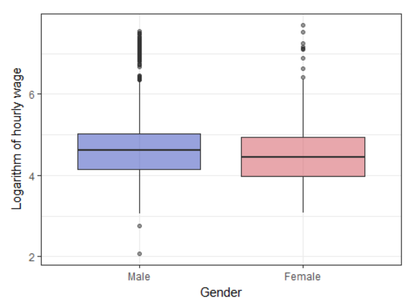
Table 2: Frequencies of categories in the sample. *n* corresponds to the individuals matching the category in the line. Individuals without college or graduate studies corresponds to the ones with maximum scholarity equal to high school.



Graph 1: Scatterplot of hourly wages and age.



Graph 2: Scatterplot of hourly wages and time at the job.



Graph 3: Boxplot of hourly wages by gender.

Posterior Results

Posterior distributions for the parameters are given by $\beta|y \sim t(\overline{\beta}, \overline{s}^2\overline{V}, \overline{\nu})$, $h|y \sim \left(\overline{s}^{-2}, \overline{\nu}\right)$ and their conjoint prior distribution is $\beta, h|y \sim NG(\overline{\beta}, \overline{V}, \overline{s}^2, \overline{\nu})$ with parameters defined as in Koop (2003). Posterior means, std. deviations and probabilities for β_i were calculated using the moments from the corresponding univariate t distribution.

Parameter	Prior mean	Posterior mean	Posterior std. dev.	IC95%	$P(\beta_i > 0)$
β_0	2.07	1.8739	0.0647	[1.747;2.001]	1
β_1	0.12	0.0949	0.0198	[0.0561;0.1338]	1
β_2	0.45	0.4943	0.0153	[0.4642;0.5243]	1
β_3	1.07	0.9525	0.0387	[0.8766;1.0284]	1
β_4	0.13	0.1053	0.0037	[0.0981;0.1126]	1
β_5	-0.00143	-0.0011	≈ 0	[-0.0012;-0.001]	0
β_6	0.00271	0.0030	0.0001	[0.0028;0.0032]	1
β_7(Gender)	**-0.15**	**-0.1334**	**0.0117**	**[-0.1563;-0.1105]**	**0**

Table 3: Posterior results

Conclusions

Posterior estimates reveals that the parameter associated with Gender is negative valued, indicating that the salaries of women, holding all other variables constant, are lower than men's for this sample. In fact, it is likely that a woman working in an IT career in Santa Catarina State have a salary 13% on average lower than her male counterparts.

Although this result is lower than the wage gap due to gender in overall population, the lack of female participation in this careers contributes to overall gap: since women in IT are underrepresented and the professions usually have higher salaries and lower gap, it could indicates that women (in general) are working in jobs with lower salaries.

References

[1] Bonini, P., & Pozzobon, F. (2016). *Discriminação Salarial Feminina e o Prêmio Salarial De TI Na Indústria De Tecnologia Da Região Sul*. Análise Econômica, 34(66).
[2] Koop, G. (2003). *Bayesian econometrics*. Wiley-Interscience, 1st edition.
[3] Schmidt, A. V. (2017). *Estimação bayesiana em modelos lineares com aplicações em economia.* Trabalho de Conclusão de Curso (Bacharelado em Economia). Florianópolis: Universidade do Estado de Santa Catarina, Centro de Ciências da Administração e Socioeconômicas.