Are Female Workers' wages penalized Simply Because They are Women?

By *Jiansong (Jensen) Xu*

2019/11/19

Acknowledgements: I would like to thank *Statistic Canada* for making the Census data available; every part of my study relies on the Census data. I also would like to thank my dear friend and classmate, Jack Madison. He kindly offered his help to solve the technical problems I had; this study would take much longer without his help. Moreover, I would like to appreciate my dear friend, Jun Jiang, for spending countless hours with me designing this study and sharing our ideas.

Introduction

As modern society promotes women's equal right, it is heartening to see that female workers are becoming increasingly ubiquitous. Also, unlike the old society in which women could only work in designated occupations, the labour market of today offers the same opportunities for all genders. Nevertheless, female workers are still being discriminated in many ways. A relevant study in the year 2013 suggested that although the female share of all wage earners increased in 2006 compared to previous years, the wage penalty against women was aggravated on average (Liaw, K.-L., & Xu, L., 2013). This disappointing finding leads to the belief that the wage penalty associated with being females persists in Canada.

The main focus of this paper is to study whether women get paid less simply because they are women. The result can hopefully inform the policymakers about the effectiveness of the previous policies that target gender inequality in Canada, which can further assist the evaluation and improvement of labour market policies. This study estimated a model to explain the wage gap between women and men, and the output shows that there is an apparent wage gap between women and men. However, the magnitude of such difference varies given one’s working experience. Overall, the results suggest that the wage penalty associated with being females does exist, and gender explains a significant portion of this penalty.

Data and Methodology

The regression model I wish to estimate is:

$$\begin{aligned}\ln WAGE_i = \alpha + \beta_f FEMALE_i + \beta_{f,a} FEMALE * POTENEXP_i \\ + \beta_{f,p2} FEMALE * POTENEXP_i^2 + \beta_{f,m} FEMALE * MARRIED_i \\ + \beta_{f,inf} FEMALE * INFANT_i + \beta_{f,psc} FEMALE * PRESCHOOLER_i \\ + \beta_{f,sch} FEMALE * SCHOOLAGE_i + \beta_{f,you} FEMALE * YOUTH_i \\ + \beta_{f,min} FEMALE * MINORITY_i + \beta_p POTENEXP_i + \beta_{p2} POTENEXP_i^2 \\ + \beta_{inf} INFANT_i + \beta_{psc} PRESCHOOLER_i + \beta_{sch} SCHOOLAGE_i \\ + \beta_{you} YOUTH_i + \beta_n i.NAICS_i + \beta_h HDGREE_i + \beta_{pr} PR_i \\ + \beta_w WEEKWRK_i + \beta_m MARRIED_i + \beta_{min} MINORITY_i + \varepsilon_i\end{aligned}$$

where i represents each observation in the sample

$\ln WAGE_i$: the natural log of the wage of i.

$FEMALE_i$: = 1 if i is female, 0 otherwise.

$POTENEXP_i$: potential years of working experience calculated by age – expected years of schooling - 6

$i.NAICS_i$: sector of employment indicator, each sector is treated as a dummy except the base category, such a dummy = 1 if i belongs to the corresponding sector, 0 otherwise.

$HDGREE_i$: the highest level of education i has completed, each degree is treated as a dummy except for the base category, such a dummy = 1 if i has the corresponding degree, 0 otherwise.

$INFANT_i$: = 1 if i has one or more infants, 0 otherwise.

$PRESCHOOLER_i$: = 1 if i has one or more preschool-aged children, 0 otherwise.

$SCHOOLAGE_i$: = 1 if i has one or more school-aged children, 0 otherwise.

$YOUTH_i$: = 1 if i has one or more children who are youth, 0 otherwise.

$MARRIED_i$: = 1 if i is married, 0 otherwise.

PR_i: the province indicator, each province is a dummy except the baseline province,

each dummy = 1 when i is a resident of the corresponding province, 0 otherwise.

$WEEKWRK_i$: weeks worked in 2010

$MINORITY_i$: = 1 if i belongs to a visible minority group, 0 otherwise.

ε_i: random error term.

The parameter of interest is $FEMALE_i$, which is a binary indicator of gender takes value 1 when the individual is female and 0 otherwise. The coefficient of $FEMALE_i$, however, does not have a useful interpretation. To make an inference, we also need to consider years of working experience, marital status and children, because the model contains interactions between gender and these parameters. A quadratic term of experience is also included as I expect the return to experience is not constant. The benefit of working experience starts to decay once the individual has accumulated a significant amount of experience. Therefore, the expected wage gap between women and men is:

$$\frac{\partial \ln WAGE_i}{\partial FEMALE_i} = \beta_f + \beta_{f,a}POTENEXP_i + \beta_{f,p2}POTENEXP_i^2 + \beta_{f,m}MARRIED_i + \beta_{f,inf}INFANT_i + \beta_{f,psc}PRESCHOOLER_i + \beta_{f,sch}SCHOOLAGE_i + \beta_{f,you}YOUTH_i + \beta_{f,min}MINORITY_i.$$

Notice that the coefficients represent the percentage change as the dependent variable is in natural logarithmic form. For a woman and a man whom both have, say, 10 years of experience, no husband/wife, only school-aged children, identities of visible minorities all other variables equal, the expected percentage wage gap is $\left[\left(\beta_f + \beta_{f,a}10 + \beta_{f,p2}100 + \beta_{f,sch} + \beta_{f,min}\right) * 100\right]$ %.

To identify the relationship between wage and gender as causal, the following assumption must satisfy:

$$E[\varepsilon_i \mid FEMALE_i, POTENEXP_i, INFANT_i, PRESCHOOLER_i, SCHOOLAGE_i, YOUTH_i, NAICS_i, HDGREE_i, PR_i, WEEKWRK_i, MARRIED_i, MINORITY_i] = 0$$

That is, there are no omitted variables that correlate with gender and earnings, conditioning on the covariates included in the model. The strategy is straightforward; the model needs to control for as many factors that affect earnings and are associated with gender as possible. This method essentially cuts out the effect from other factors when the model explains the variation in wage using the variation in gender. Hence, if the model controls for enough factors, the estimated effect of gender on wage can be interpreted as a causal effect. The list of controls includes occupation, education, children, province of residence, number of weeks worked last year, and marital status.

The sample was restricted to individuals who were legally receiving payments as employees in 2010. Besides, those who had a negative or zero gross wage in 2010 are not included in the sample because the regression model uses natural logarithmic wage as the dependent variable. Also, it is essential to note that the 2011 Census data contains age groups instead of real ages. Thus, the age variable in the sample was derived by taking the nearest whole number of the median of each age group. Furthermore, $POTENEXP_i$ is an alternative measurement of years of working experience. Its formula is (age – years of schooling – 6), where 6 is the expected age of first-year schooling. "Age" is the rounded median of the age group that each individual belongs. This compromised measurement replaces the actual years of the working experience because the record of working experience is unavailable in the Census data.

Descriptive statistics

Table 1[1] is the summary statistics of all variables in the regression model. Gender, marital status, visible minority and children are treated as binaries; while the province of residence, education and occupation are categorical variables. We can inspect how many categories each categorial variable has by looking at the maximum. The model uses infant, preschooler, school-age and youth dummies instead of a multi-level categorical variable for all types of children. This specification allows the model to capture the effect of having children who are in different age groups. The sample mean of $FEMALE_i$ is close to .5, meaning that the sample contains an approximately equal share of female and male. Besides, the standard deviations of potential experience and all its associated variables are large, because the potential experience takes on same values for individuals who are in the same age group and hold the same highest degree. To be specific, there exist enormous values of the experience variable that are relatively far away from the mean. Further, the range of $\ln WAGES$ is not close to 0, because observations who earn 0 or approximately 0 wages were dropped from the sample, and the lowest wage record in this data set is 1000.

Regression result

Column 1 of table 3 is the complete regression output, where each row represents a parameter. Column 2 is the output of an alternative model which is for later discussion. The model made estimations based on 310,910 observations. The R-squared and adjusted R-squared are .5238 and .5237, respectively. As mentioned before, the interpretation depends on the control variables that interact with the gender dummy in the model. For instance, consider a woman and a man

[1] All tables are to be found in the *Table* section.

whom both have 10 years of working experience, no husband/wife, only school-aged children, and no visible minority identities. The expected wage gap is $-.093 - .009 * 10 + .0001 * 100 - .092 = -.265$. In other words, the woman is expected to earn 26.5% less than the man, holding all other variables constant. To take another example, imagine another pair of woman and man. Suppose they both have 15 years of working experience and children who are infants, they are both married, and both belong to a visible minority group. Their expected wage difference is $-.093 - .009 * 15 + .0001 * 15^2 - .115 - .109 + .140 =$ -.2895. In both examples, women have a lower wage than the men who are similar to them in perspectives besides gender. This is not a surprising result, because most of the estimated coefficients take on negative values, and all of them have significantly large t-values. To illustrate, a woman's wage penalty associated with her gender will increase by 10.9% if she becomes married, assuming all other characteristics remain unchanged. Similarly, if we randomly define values for a list of characteristics, the result is much likely to be that the woman receives a lower wage than the man.

By looking at the estimated coefficient of FEMALE*POTENEXP, we can tell that how does the wage penalty when individuals have more years of working experience. Intuitively, an employee who has more years of working experience ought to earn more salaries than her/his less experienced colleagues. Also, the return to experience should start decreasing at some point as the worker has less need for experience. This expectation matches the estimated coefficient of POTENEXP and its quadratic term (.067 and -.001, respectively). Nevertheless, the expected wage difference does not disappear immediately when women accumulate more years of experience, as the estimated coefficient of $FEMALE * POTENEXP^2$ has only a tiny magnitude.

Specifically, when a pair of male and female workers both gain one more year of working experience, the benefit they gain will be differentiated by [100*(-.009 + .0002 POTENEXP)] %, given the two workers are identical in other perspectives. Likewise, although the wage gap starts to shrink at some point, it would take about 45 years for the shrinkage to begin. It would take even longer for women and men to enjoy the same return to working experience
, supposing other covariates do not change.

Moreover, having children seems always to exacerbate the wage penalty associated with being women, regardless of children's ages. If a woman has, say, an infant and a preschool-aged child, her wages will be 20.3% lower than a woman who is identical except not having any children. The reason for this wage discrimination against mothers is ambiguous. One possible explanation is those female employees who have children need to take care of their children or attend school events. Therefore, they may spend less time on their jobs.

On the other hand, although we cannot ignore the importance of fathers in childcare, male employees who have children do not always earn less. The male employees who are fathers of infants and preschoolers seem to earn more than other identical male employees who are not fathers. Being fathers of school-aged kids yields a positive effect on male employees' wages as well, but this effect is not significant. It is hard to explain why this wage improvement occurs only for certain types of children. Probability, the male workers need to earn extra to support their children only when the children are little. Nevertheless, it does not comprehensively explain the wage penalty against the mothers.

Furthermore, being a member of the visible minority appears to have a considerable impact on wage. If a female employee belongs to a visible minority group, her wage is expected to hike by 14% compared to other identical female workers who are not minorities. However, a male minority employee is expected to earn 27% less than his identical but non-minority colleagues. Unlike most of the other variables in the model, being a minority actually benefits the earnings of the female employees, conditioning on other variables. The reason, in my judgement, is too blurry to see through intuition. It might be the case that the employers want to establish a "fair and unbiased" appearance to the public. Thus they choose to offer minority female workers more generous salaries. But on the other hand, by doing so, the employers are engaging in gender discrimination. Then it could be the case that the majority of people and media do not pay much attention to gender discrimination against men. In any case, it is better not making any excessive speculations when the regression model does not intend to explain the impact of visible minority identities.

Robustness check

The regression model above used the potential experience as both a control and interacting variable. The potential experience may affect the result as it is, after all, an "educated guess". To check the reliability of the regression result, I constructed an alternative model using age instead of the potential experience to form control variables and interaction terms. Since potential experience was derived using age, an alternative model that adopts age instead of potential experience should yield similar interpretations. The model for robustness check is:

$$\begin{aligned}\ln WAGE_i = {} & \alpha' + \beta'_f FEMALE_i + \beta_{f,a} FEMALE * AGE_i + \beta_{f,a2} FEMALE * AGE_i^2 \\ & + \beta'_{f,m} FEMALE * MARRIED_i + \beta'_{f,inf} FEMALE * INFANT_i \\ & + \beta'_{f,psc} FEMALE * PRESCHOOLER_i + \beta'_{f,sch} FEMALE * SCHOOLAGE_i \\ & + \beta'_{f,you} FEMALE * YOUTH_i + \beta'_{f,min} FEMALE * MINORITY_i + \beta_a AGE_i \\ & + \beta_{a2} AGE_i^2 + \beta'_{inf} INFANT_i + \beta'_{psc} PRESCHOOLER_i + \beta'_{sch} SCHOOLAGE_i \\ & + \beta'_{you} YOUTH_i + \beta'_n i.NAICS_i + \beta'_h HDGREE_i + \beta'_{pr} PR_i + \beta'_w WEEKWRK_i \\ & + \beta'_m MARRIED_i + \beta'_{min} MINORITY_i + \varepsilon'_i\end{aligned}$$

Table 2 is the descriptive statistics of the alternative model. This model used the same number of observations as the primary model. The standard deviations associated with age variables are all notably large; the reason is that ages were taken as the rounded median for each age group, which is similar to the reason that potential experience variables have large standard deviations.

The second column of table 3 is the regression output of the alternative model. Although the estimated coefficient of FEMALE is positive, the interpretation does not change dramatically. Indeed, the alternative model suggests a comparable interpretation as the primary model. For instance, controlling for all other variables, a married white female employee who is 45 years old and has school-aged kids is expected to earn approximately 46% less than a male employee. Besides, notice that the estimated coefficient of FEMALE*AGE is slightly larger than FEMALE*POTENEXP, and FEMALE*AGE^2 has the same coefficient as FEMALE*POTENEXP^2 after rounding. The difference in these estimates suggests that it takes longer for the wage gap to shrink according to the alternative model. However, in either model, it takes several decades for the wage gap to start diminishing if other covariates do not change. Therefore, the real-life interpretations of the two models are not much different.

Threats to identification

The result indeed provides us with an insight into the wage gap between female and male employees, being a woman in the workplace appears to have a negative effect on earnings. Notwithstanding, we should not be overconfident to say that being female leads to a lower wage compared to males. Even though the model managed to control for as many factors as possible, there still exists potential omitted factors that are correlated with gender and wage. For example, the working abilities of the employees can have a considerable impact on earnings; employers tend to offer a higher wage to skilled workers in order to stop job-hopping.

Another source of bias is the fact that employers may implement very generous benefits for their employees, including free therapists, free transportations, free insurance, etc. The existence of these benefits may reduce the salaries that an employer can offer. For example, if a company provides pregnant employees with reimbursed health care and childcare service, the wages that those female employees receive are likely to be lower as part of the wages were used to make up for these benefits. In light of these examples, the regression model may not be able to capture the true effects.

Conclusion

The result turned out to be saddening. In a model that controls for a list of variables, gender seems to play an important role in the wage penalty. The magnitude of this wage penalty against women diminishes very slowly when women have more working experience or get older. Also, having children and being married are essential factors which aggravate the wage penalty against women. The minority identity is also a remarkable determinant of the wage gap, though we need further investigation to reveal the source of its effect. Regardless of the accuracy of this result,

the methodology in this study does require some improvements. First of all, the potential experience variables are likely to contain massive measurement errors, given that there are only finite possible combinations of age and years of education due to the nature of these measurements. To consider real life, some people may have started accumulating working experience before finishing schools. Indeed, many college and university students in Canada already have some years of working experience by the time they graduate. Hence, the potential experience we adopted might be significantly different from the actual records, causing the estimates to be biased. Therefore, it would be helpful if we know the actual years of experience of each individual. Secondly, there are some factors that I wish to control for but are not included in the Census data, such as the working ability. If I have access to more information, I would consider adopting the average hours that each individual needs to finish a work assignment as a proxy to the working ability, since the actual working ability is nearly impossible to capture.

Reference page

1. Liaw, K.-L., & Xu, L. (2013). Changes in Wage Distributions of Wage Earners in Canada: 2000-2005, 22 pages. https://doiorg.proxy.lib.sfu.ca/http://socserv.mcmaster.ca/qsep/p/qsep451.pdf

Tables

Table 1

	(1)				
	mean	sd	count	min	max
lnWAGES	10.36	1.042	310910	6.908	13.89
Female	0.500	0.500	310910	0	1
Female*Potential experience	10.55	13.83	310910	0	68
Female*Potential experience^2	302.7	508.4	310910	0	4624
Female*Married	0.299	0.458	310910	0	1
Female*Infant	0.0338	0.181	310910	0	1
Female*Preschooler	0.0609	0.239	310910	0	1
Female*Schoolage	0.128	0.334	310910	0	1
Female*Youth	0.181	0.385	310910	0	1
Female*Minority	0.0903	0.287	310910	0	1
Infant	0.0742	0.262	310910	0	1
preschooler	0.129	0.335	310910	0	1
school-age children	0.251	0.434	310910	0	1
Youth-age children	0.354	0.478	310910	0	1
Married	0.605	0.489	310910	0	1
Potential experience	21.42	12.87	310910	1	68
Potential experience^2	624.3	601.9	310910	1	4624
Labour: Industry sectors (based on the NAICS 2007)	11.01	5.246	310910	1	20
Education: Highest certificate, diploma or degree	5.324	3.396	310910	1	13
Province or territory of current residence (2011)	6.476	2.000	310910	1	11

weeks worked in 2010	43.84	12.50	310910	5	51
Minority	0.180	0.385	310910	0	1

Table 2

	(1)				
	mean	sd	count	min	max
lnWAGES	10.36	1.042	310910	6.908	13.89
Female	0.500	0.500	310910	0	1
Female*Age	20.56	22.42	310910	0	85
Female*Age^2	925.5	1186.4	310910	0	7225
Female*Married	0.299	0.458	310910	0	1
Female*Infant	0.0338	0.181	310910	0	1
Female*Preschooler	0.0609	0.239	310910	0	1
Female*Schoolage	0.128	0.334	310910	0	1
Female*Youth	0.181	0.385	310910	0	1
Female*Minority	0.0903	0.287	310910	0	1
Infant	0.0742	0.262	310910	0	1
preschooler	0.129	0.335	310910	0	1
school-age children	0.251	0.434	310910	0	1
Youth-age children	0.354	0.478	310910	0	1
Married	0.605	0.489	310910	0	1
Age	41.47	12.93	310910	19	85
Age^2	1887.1	1092.6	310910	361	7225
Labour: Industry sectors (based on the NAICS 2007)	11.01	5.246	310910	1	20
Education: Highest certificate, diploma or degree	5.324	3.396	310910	1	13
Province or territory of	6.476	2.000	310910	1	11

current residence (2011)					
weeks worked in 2010	43.84	12.50	310910	5	51
Minority	0.180	0.385	310910	0	1

Table 3

	(1) lnWAGES	(2) lnWAGES
Female	-0.093***	0.078**
	(-11.48)	(2.92)
Female*Potential experience	-0.009***	
	(-11.12)	
Female*Potential experience^2	0.000***	
	(6.75)	
Female*Age		-0.012***
		(-8.75)
Female*Age^2		0.000***
		(6.57)
Female*Married	-0.109***	-0.116***
	(-17.35)	(-18.21)
Female*Infant	-0.115***	-0.110***
	(-10.72)	(-10.29)
Female*Preschooler	-0.088***	-0.080***
	(-10.32)	(-9.42)
Female*Schoolage	-0.092***	-0.085***
	(-14.24)	(-13.29)
Female*Youth	-0.043***	-0.047***
	(-7.53)	(-8.11)
Female*Minority	0.140***	0.144***
	(20.59)	(21.24)
Youth-age children	-0.029***	-0.002
	(-7.05)	(-0.40)
Married	0.115***	0.108***
	(24.44)	(22.80)
Potential experience	0.067***	
	(119.23)	

Potential experience^2	-0.001***	
	(-101.53)	
Age		0.114***
		(117.11)
Age^2		-0.001***
		(-105.66)
Minority	-0.270***	-0.279***
	(-54.60)	(-56.50)
Infant	0.073***	0.060***
	(10.13)	(8.24)
preschooler	0.061***	0.053***
	(10.26)	(8.92)
school-age children	0.005	0.004
	(1.14)	(0.84)
11 Agriculture, forestry, fishing and hunting	0.000	0.000
	(.)	(.)
21 Mining, quarrying, and oil and gas extraction	0.765***	0.767***
	(49.44)	(49.64)
22 Utilities	0.604***	0.609***
	(35.21)	(35.59)
23 Construction	0.275***	0.278***
	(22.23)	(22.53)
31-33 Manufacturing	0.247***	0.248***
	(20.93)	(21.11)
41 Wholesale trade	0.270***	0.273***
	(21.41)	(21.64)
44-45 Retail trade	-0.125***	-0.111***
	(-10.65)	(-9.50)
48-49 Transportation and warehousing	0.175***	0.175***
	(13.89)	(13.91)
51 Information and cultural industries	0.291***	0.295***
	(20.87)	(21.22)
52 Finance and insurance/55 Management of companies and enterprises	0.389***	0.391***
	(30.83)	(31.08)
53 Real estate and rental and leasing	0.071***	0.077***
	(4.57)	(4.98)
54 Professional, scientific and technical services	0.289***	0.289***

	(23.37)	(23.37)
56 Administrative and support, waste management and remediation services	-0.068***	-0.061***
	(-5.24)	(-4.73)
61 Educational services	0.101***	0.110***
	(8.36)	(9.12)
62 Health care and social assistance	0.162***	0.169***
	(13.65)	(14.27)
71 Arts, entertainment and recreation	-0.123***	-0.112***
	(-8.31)	(-7.59)
72 Accommodation and food services	-0.269***	-0.248***
	(-21.78)	(-20.11)
81 Other services (except public administration)	-0.046***	-0.038**
	(-3.54)	(-2.90)
91 Public administration	0.354***	0.360***
	(29.55)	(30.11)
No certificate, diploma or degree	0.000	0.000
	(.)	(.)
High school diploma or equivalent	0.101***	0.108***
	(20.49)	(21.96)
Trades certificate or diploma (other than apprenticeship)	0.161***	0.122***
	(23.65)	(18.04)
Registered Apprenticeship certificate	0.292***	0.254***
	(38.34)	(33.49)
College,CEGEPorothernon-universitycertificatordiplomfromaprgrmof3mnthstlssthn1yr	0.173***	0.160***
	(18.73)	(17.44)
College,CEGEPorothernon-universitycertificateordiplomafromaprogramof1yeart2years	0.257***	0.233***
	(43.32)	(39.68)
College,CEGEPorothernon-universitycertificateordiplomfromaprogramofmorthan2years	0.365***	0.344***
	(59.39)	(56.34)
University certificate or diploma below bachelor level	0.387***	0.366***
	(52.86)	(50.36)
Bachelor's degree	0.581***	0.529***
	(101.99)	(94.63)
University certificate or diploma above bachelor level	0.632***	0.570***

	(68.09)	(61.95)
Degree in medicine, dentistry, veterinary medicine or optometry	0.867***	0.769***
	(37.15)	(33.04)
Master's degree	0.690***	0.616***
	(92.32)	(83.59)
Earned doctorate degree	0.856***	0.787***
	(56.31)	(52.00)
Newfoundland and Labrador	0.000	0.000
	(.)	(.)
Prince Edward Island	-0.038	-0.034
	(-1.72)	(-1.51)
Nova Scotia	-0.050***	-0.043***
	(-3.83)	(-3.34)
New Brunswick	-0.040**	-0.038**
	(-2.99)	(-2.80)
Quebec	-0.009	-0.012
	(-0.86)	(-1.07)
Ontario	0.081***	0.085***
	(7.62)	(8.04)
Manitoba	0.012	0.020
	(0.96)	(1.60)
Saskatchewan	0.097***	0.104***
	(7.49)	(8.05)
Alberta	0.205***	0.213***
	(18.47)	(19.26)
British Columbia	0.073***	0.079***
	(6.63)	(7.22)
Northern Canada	0.266***	0.282***
	(9.07)	(9.62)
weeks worked in 2010	0.035***	0.034***
	(314.31)	(309.71)
Constant	7.755***	6.012***
	(468.81)	(249.70)
Observations	310910	310910

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$