An Empirical Analysis of the Gender Wage Gap in Canada

ECON 5600(M) | Final Research Paper

Written By:

Isanna Biglands (218887273)
ibigland@yorku.ca

Jian Hao Chiah (216031965)
jhaoc98@yorku.ca

April 2022

Table of Contents

1. Introduction

The paper aims to examine the existing gender wage gap in Canada. By utilizing Statistics Canada's Labour Force Survey 2021 data, this paper estimates the value of gender wage gap and its associated gender discrimination. The methodology involves implementing the Ordinary Least Squares (OLS) regressions and the Blinder-Oaxaca decomposition method. Continuous development in this research topic is essential for Canada to progress towards a more equitable labour market and the economy overall.

This paper will first review the prominent theories of wage discrimination found in the existing research literatures. We will then introduce the data sources utilized to perform our empirical analysis. This will be followed by explanations of the empirical approach and methodology employed to estimate the potential gender wage gap. The main results of our statistical analysis will be presented in section five. Our findings from this research identify the existence of a wage gap between men and women in Canada, and at least a portion of the wage gap might be due to gender discrimination in the Canadian labour market. Finally, in section six, we will conclude the analysis conducted and discuss some limitations and opportunities for further research.

2. Literature Review

There are vast economic literature that analyzes gender wage gap in both theoretical and empirical contexts. Two traditional main categories of explanations for the gender wage gap are labour supply and labour demand explanations. We will summarize the research findings concerning the gender wage gap from both the supply-side and demand-side perspectives.

2.1. Supply-Side

The literature on the supply-side says that women's preferences explain their labour market choices and earnings. For example, Hakim (2000) found that gender differences in career path and earnings derive largely from gender differences in work preferences. Work preferences have been found to be shaped by social norms and gender stereotypes which may encourage men and women to choose different types of careers (Correll, 2001, 2004; Ridgeway, 2009; Ridgeway and Correll, 2004; Winslow, 2010). Women and men tend to specialize in different occupations in the United States (Hegewisch et al., 2010) and Europe (Bettio and Verashchagina, 2009). Female-dominated occupations have lower wages than male-dominated occupations despite comparable qualifications (England, 2005; England, 2010; England and Folbre, 2005). According to the Mincer-Polachek hypothesis (1974), gender differences in experience and labour force participation are the key determinants of the wage gap, consistent with the human capital model. Mincer and Polachek (1974) argue that anticipated family responsibilities influence women's decisions to pursue schooling and the length of time they devote to labour market activities. Recently, women have made substantial progress in educational outcomes, and the gender gap in educational attainment has almost completely closed (Blau and Kahn, 2000, 2007; Bobbitt-Zeher, 2007; Fortin, 2019). Unfortunately, despite women's progress in higher education, the belief that women and men are suited for different fields of study and professions still exists (England and Li, 2006; Reskin and Bielby, 2005).

2.2. Demand-Side

Job discrimination against women is suggested as a demand-side explanation for the gender wage gap. Gender discrimination refers to employers' gender-biased decisions on the allocation of individuals across and within occupations (Castagnetti and Rosti, 2013). There are two main theories related to gender discrimination in the literature. The first theory relates to horizontal segregation. This type of discrimination is usually called "taste discrimination," which was defined by Becker (1957) or "statistical discrimination," which was introduced by several authors throughout the 1970s (Arrow, 1972; Stiglitz, 1973; Spence, 1973). The second theory is related to vertical segregation, i.e., the "glass ceiling." Blau and Kahn (1997) coined the term "swimming upstream" to characterize women's pursuit of pay equality in the face of countervailing currents. There is evidence of a growing role of the underrepresentation of women among top earners in explaining the persistence of the gender pay gap in Canada (Fortin, 2019).

From this literature review, we can conclude that both explanations play a role in defining the gender pay gap. Our study complements the other research summarized above to continue analyzing the evolution of women's labour force participation and the gender earnings ratio.

3. Data

We use data from Statistics Canada's Labour Force Surveys (LFS) Public Use Microdata Files (otherwise known as PUMFs) in this research. These data are collected at the individual respondent level for the year 2021. The variables used in this analysis are presented in **Appendix A.** Each variable is important to consider when estimating potential gender wage discrimination as they account for many of the exogenous factors that might impact an individual's wages.

4. Empirical Approach: Oaxaca Decomposition

An often-used methodology to study labour market outcomes by groups (gender, race, and others) is to decompose mean differences in log wages based on regression models in a counterfactual manner. The procedure is known as the Oaxaca decomposition. It divides the wage differential between two groups into a part that is "explained" by group differences in productivity characteristics such as education or work experience and a residual part that cannot be accounted for by such differences in wage determinants. This "unexplained" part is often used to measure discrimination, but it also subsumes the effects of group differences in unobserved predictors. Since this paper is primarily focused on the amount of potential wage discrimination based on gender, the analysis will focus on the unexplained portion of the wage differential described above in this section.

The Oaxaca decomposition technique is explained in detail in **Appendix B**.

5. Main Results

The results of the estimation strategy are presented below.

5.1. OLS Regression Results

To begin, in the OLS regressions performed, the logarithm of the hourly wage rate is the dependent variable. The predictors are described in Section 3 above. The linear regressions are grouped by gender to estimate the earnings function of males and females separately. The results from this analysis can be found in **Appendix C**. Table 1 represents the results for males, and Table 2 contains the results for females. An interesting finding from these results is that individuals in the age category 45 to 49 years earn the highest wages for both genders. This finding is in-line with the lifecycle as most workers reach their peak earning potential at that age. Across the "education" categories, there is a similar trend for both men and women; higher educated individuals earn more in wages. A final observation from the OLS results is that individuals who have immigrated to Canada more recently (within the last ten years) earn lower wages than individuals who are not recent immigrants and those born in Canada. This difference is more pronounced for females than it is for males. This finding might highlight a form of racial discrimination; however, this type of analysis is beyond the scope of this paper.

5.2. Oaxaca Decomposition Results

A threefold Blinder-Oaxaca decomposition between male and female workers was performed using the methodology outlined in Section 4 with Stata. The explained and unexplained variations in the independent variables are shown in Table 3 and Table 4 in **Appendix D**. Here, the values indicate the portion of the total difference in wages between males and females explained by the variables (and therefore, an individual's endowment) or unexplained (representing potential discrimination in the labour force). A more significant "unexplained" value would indicate a higher amount of potential gender discrimination. From Table 4, the output suggests that among the \$0.25 wage gap, \$ 0.07 can be accounted for group differences in endowments, and \$0.14 can be accounted for group differences in coefficients. In contrast, the remaining \$0.04 represents simultaneous effects between group differences in endowments and coefficients.

Table 3 includes our main findings using a two-fold Oaxaca decomposition method in STATA. The results show that the mean of the log wages is approximately \$6.93 for males and \$6.69 for females, yielding a wage gap of \$0.24. This wage gap can be further decomposed to illustrate what portion of the gap is explained compared with the portion that is unexplained and potentially due to discrimination. For instance, we can see in the results in Table 3 that \$0.09 of the wage gap is explained by group differences in the independent variables, and approximately \$0.15 is unexplained and potentially due to gender discrimination and other factors.

To summarize, the three-fold decomposition results are consistent with the two-fold technique summarized above in terms of variables' significance in contribution to wage differential and potential gender discrimination.

6. Conclusion

This research project aimed to measure the gender wage gap in Canada in 2021. In conclusion, our results support the claim that there is potentially gender discrimination against women in the labour market. While there has been a large body of literatures devoted to estimating the amount of gender discrimination that exists in Canada's labour market, the contribution of our analysis is to update these findings to the year 2021.

References

1. Ridgeway, C. L., & Correll, S. J. (2004). Unpacking the Gender System. *Gender & Society*, *18*(4), 510–531. https://doi.org/10.1177/0891243204265269
2. Correll, S. J. (2004). Constraints into Preferences: Gender, Status, and Emerging Career Aspirations. *American Sociological Review*, *69*(1), 93–113. https://www.jstor.org/stable/3593076
3. RIDGEWAY, C. L. (2009). FRAMED BEFORE WE KNOW IT: How Gender Shapes Social Relations. *Gender and Society*, *23*(2), 145–160. https://www.jstor.org/stable/20676769
4. WINSLOW, S. (2010). GENDER INEQUALITY AND TIME ALLOCATIONS AMONG ACADEMIC FACULTY. *Gender and Society*, *24*(6), 769–793. https://www.jstor.org/stable/25789906
5. Hegewisch, A., Liepmann, H., Hayes, J., & Hartmann, H. (2010, September 1). *Separate and Not Equal? Gender Segregation in the Labor Market and the Gender Wage Gap*. IWPR. https://iwpr.org/iwpr-issues/esme/separate-and-not-equal-gender-segregation-in-the-labor-market-and-the-gender-wage-gap/
6. Bettio, F., & Verashchagina, A. (2009). *Gender segregation in the labour market. Root causes, implications and policy responses in the EU*. ResearchGate; unknown. https://www.researchgate.net/publication/303566495_Gender_segregation_in_the_labour_market_Root_causes_implications_and_policy_responses_in_the_EU
7. England, P. (2005). Emerging Theories of Care Work. *Annual Review of Sociology*, *31*, 381–399. https://www.jstor.org/stable/29737725

8. England, P. (2005). THE GENDER REVOLUTION: Uneven and Stalled. *Gender and Society, 24*(2), 149–166. https://www.jstor.org/stable/27809263

9. England, Paula and Nancy Folbre (2005). "Gender and Economic Sociology." Pp. 627-49 in The Handbook of Economic Sociology, edited by N. J. Smelser and R. Swedberg. New York: Russell Sage Foundation.

10. Mincer, J., & Polachek, S. (1978). An Exchange: The Theory of Human Capital and the Earnings of Women: Women's Earnings Reexamined. *The Journal of Human Resources*, *13*(1), 118. https://doi.org/10.2307/145305

11. Blau, F. D., & Kahn, L. M. (2000). Gender Differences in Pay. *Journal of Economic Perspectives*, *14*(4), 75–100. https://doi.org/10.1257/jep.14.4.75

12. Blau, F. D., & Kahn, L. M. (2007). The Gender Pay Gap: Have Women Gone as Far as They Can? *Academy of Management Perspectives*, *21*(1), 7–23. https://www.jstor.org/stable/4166284

13. Fortin, N. M. (2019). *Increasing Earnings Inequality and the Gender Pay Gap in Canada: Prospects for Convergence*. Papers.ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3385014

14. Bobbitt-Zeher, D. (2007). The Gender Income Gap and the Role of Education. *Sociology of Education*, *80*(1), 1–22. https://www.jstor.org/stable/20452692

15. England, P., & Li, S. (2012, March 6). *Desegregation Stalled the Changing Gender Composition of College Majors, 1971-2002*. Papers.ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2017175

16. Reskin, B. F., & Bielby, D. D. (2005). A Sociological Perspective on Gender and Career Outcomes. *The Journal of Economic Perspectives*, *19*(1), 71–86. https://www.jstor.org/stable/4134993

17. CASTAGNETTI, C., & ROSTI, L. (2013). UNFAIR TOURNAMENTS: Gender Stereotyping and Wage Discrimination among Italian Graduates. *Gender and Society*, *27*(5), 630–658. https://www.jstor.org/stable/43669821

18. Gary Stanley Becker. (1957). *The economics of discrimination*. The University Of Chicago Press.

19. Blau, F. D., & Kahn, L. M. (1997). Swimming Upstream: Trends in the Gender Wage Differential in the 1980s. *Journal of Labor Economics*, *15*(1), 1–42. https://www.jstor.org/stable/2535313

20. Hlavac, M. (2014). oaxaca: Blinder-Oaxaca Decomposition in R. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2528391

21. Jann, B. (2008). The Blinder–Oaxaca Decomposition for Linear Regression Models. *The Stata Journal: Promoting Communications on Statistics and Stata*, *8*(4), 453–479. https://doi.org/10.1177/1536867x0800800401

Appendix

Appendix A: Variables

In the analysis, we use the following variables from Statistics Canada's Labour Force Survey:

Variable in Labour Force Survey (2021)	Category	Variable Name	Code
HRLYEARN	Usual hourly wages (Employees only. Includes tips, commission and bonuses. Before taxes and other deductions)		
AGE_12	Five-year age group of respondents	15 to 19 years	1
		20 to 24 years	2
		25 to 29 years	3
		30 to 34 years	4
		35 to 39 years	5
		40 to 44 years	6
		45 to 49 years	7
		50 to 54 years	8
		55 to 59 years	9
		60 to 64 years	10
		65 to 69 years	11
		70+ years	12
MARSTAT	Marital status of respondent	Married	1
		Living in common-law	2
		Widowed	3
		Separated	4
		Divorced	5
		Single, never married	6
EDUC90	Highest educational attainment	0 to 8 years	0
		Some secondary	1
		Some post secondary	3
		Post secondary certificate of diploma	4
		University: bachelors degree	5
		University: graduate degree	6
IMMIG	Immigrant status	Immigrant, landed 10 or less years earlier	1
		Immigrant, landed more than 10 years earlier	2
		Non-immigrant	3
FTPTMAIN	Full- or part-time status at main or only job (Currently employed only)	Full-time	1
		Part-time	2

Variable in Labour Force Survey (2021)	Category	Variable Name	Code
City-Size	Census Metropolitan Areas (CMA) or not	In a CMA	1
		Not in a CMA	0
SEX	Sex of respondent	Male	1
		Female	2
FINALWT	Final individual or family population weight		

Appendix B: Oaxaca Decomposition Detailed Methodology

To determine the potential gender discrimination in Canada, the methodology we employ here is an Oaxaca decomposition. This method was first developed by Blinder (1973) and Oaxaca (1973) and has been used widely in the literature to estimate gender discrimination. The methodology used here follows that which was originally developed by Blinder (1973) and Oaxaca (1973), as described by Jann (2008). The Oaxaca decomposition technique can be explained as follows:

This approach is titled a “decomposition” technique because it decomposes the observed wage gap into two parts: the “endowment” and a “coefficient” component. The “endowment” is the component of the wage differential explained by individual ‘characteristics’ like education, age, and others. The “coefficient” is derived as an unexplained residual and shows the component of wage differentials explained by ‘discrimination.’

In this study, we have two groups – males and females – with an outcome variable, Y (log of hourly wages) and variables like education, age and more. Now the mean outcome difference is computed as:

$$\mathbf{D = E(Y_M) - E(Y_F) \qquad (1)}$$

Where E (Y) denotes the expected value of the outcome variable. We want to estimate how much of the mean outcome difference is accounted for by group differences in the predictors. The linear model is:

$$\mathbf{Y_i = X'_i\beta_i + \varepsilon_i, \qquad E(\varepsilon_i) = 0, \qquad i \in \{F, M\} \qquad (2)}$$

Where, X is a vector containing the predictors and a constant, β contains the slope parameters and the intercept, and ε is the error term. Based on the linear model, the mean outcome difference can be expressed as the difference in the linear prediction at the group-specific means of the regressors. That is,

$$\mathbf{D = E(Y_M) – E(Y_F) = E(X_M)' \beta_M – E(X_F)' \beta_F} \qquad \textbf{(3)}$$

since,

$$\mathbf{E(Y_i) = E(X'_i\beta_i + \varepsilon_i) = E(X'_i\beta_i) + E(\varepsilon_i) = E(X_i)'\beta_i}$$

Where $E(\beta_i) = \beta_i$, and $E(\varepsilon_i) = 0$ by assumption.

To identify the contribution of group differences in predictors to the overall outcome difference, equation (3) can be rearranged, for example, as follows:

$$\mathbf{D = [E(X_M) – E(X_F)]' \beta_F + E(X_F)'(\beta_M - \beta_F) + [E(X_M) – E(X_F)]'(\beta_M - \beta_F)} \qquad \textbf{(4)}$$

This is a “three-fold” decomposition, that is, the outcome difference is divided into three parts:

$$\mathbf{D = E + C + I}$$

The first component, $\mathbf{E = [E(X_M) – E(X_F)]' \beta_F}$, amounts to the part of the differential that is due to group differences in the predictors (the “endowments effect”). The second component $\mathbf{C = E(X_F)'(\beta_M - \beta_F)}$, measures the contribution of differences in the coefficients (including differences in the intercept). The third component $\mathbf{I = [E(X_M) – E(X_F)]'(\beta_M - \beta_F)}$ is the interaction term accounting for the fact that differences in endowments and coefficients exist simultaneously between the two groups.

The “decomposition” in equation (4) is formulated from the viewpoint of females (F). That is, the group differences in the predictors are weighted by the coefficients of females to determine the endowments effect (E). In other words, the E component measures the expected change in females’ mean outcome if females had males’ predictor levels. Similarly, for the second component (C), the differences in coefficients are weighted by females' predictor levels. The second component measures the expected change in females' mean outcome if females had males' coefficients. Naturally, the differential can analogously be expressed from the viewpoint of males, yielding the reverse three-fold decomposition.

Appendix C: OLS Results

Table 1: OLS Results for Males in 2021

Survey: Linear regression

Number of strata	= 1	Number of obs	= 254,645
Number of PSUs	= 254,645	Population size	= 99,127,555
		Design df	= 254,644
		F(26, 254619)	= 4786.80
		Prob > F	= 0.0000
		R-squared	= 0.5798

lnwage	Coefficient	Linearized std. err.	t	P>|t|	[95% conf.	interval]
1.age15to19	-.1691293	.0146091	-11.58	0.000	-.1977628	-.1404958
1.age20to24	-.0186948	.0139355	-1.34	0.180	-.0460081	.0086184
1.age25to29	.1410581	.0137432	10.26	0.000	.1141218	.1679944
1.age30to34	.2566034	.0135175	18.98	0.000	.2301095	.2830974
1.age35to39	.3090593	.0134262	23.02	0.000	.2827443	.3353744
1.age40to44	.3100912	.0134023	23.14	0.000	.283823	.3363594
1.age45to49	.3162781	.0134771	23.47	0.000	.2898633	.3426928
1.age50to54	.3059377	.013458	22.73	0.000	.2795604	.332315
1.age55to59	.2866362	.0134487	21.31	0.000	.2602772	.3129952
1.age60to64	.214979	.0136937	15.70	0.000	.1881396	.2418184
1.age65to69	.1369658	.0154616	8.86	0.000	.1066614	.1672702
1.age70plus	0	(omitted)				
1.married	.1680276	.0037663	44.61	0.000	.1606457	.1754094
1.commonlaw	.1102108	.004081	27.01	0.000	.1022122	.1182095
1.widow	.1100384	.0193676	5.68	0.000	.0720784	.1479984
1.separated	.1103346	.0093321	11.82	0.000	.0920439	.1286253
1.divorced	.1039985	.0080135	12.98	0.000	.0882923	.1197046
1.single	0	(omitted)				
1.grade0to8	-.5278574	.010235	-51.57	0.000	-.5479177	-.5077972
1.somehighschool	-.4573414	.0066249	-69.03	0.000	-.470326	-.4443568
1.highschoolgrad	-.3800324	.005493	-69.19	0.000	-.3907985	-.3692664
1.somepostsecondary	-.3654198	.0073429	-49.77	0.000	-.3798117	-.351028
1.postsecondarygrad	-.2600638	.0050862	-51.13	0.000	-.2700326	-.2500949
1.bachdegree	-.1318292	.0055088	-23.93	0.000	-.1426263	-.1210321
1.abovebach	0	(omitted)				
1.immigrant_less10	-.2156522	.0052008	-41.46	0.000	-.2258457	-.2054587
1.immigrant_more10	-.1011993	.0040101	-25.24	0.000	-.1090591	-.0933395
1.nonimmigrant	0	(omitted)				
1.fulltime	1.131588	.0061304	184.59	0.000	1.119572	1.143603
1.parttime	0	(omitted)				
1.city_size	.0242393	.0024653	9.83	0.000	.0194073	.0290713
_cons	5.900625	.0147858	399.07	0.000	5.871646	5.929605

Table 2: OLS Results for Females in 2021

```
Survey: Linear regression

Number of strata =       1                  Number of obs     =    255,398
Number of PSUs   = 255,398                  Population size = 95,243,191
                                            Design df         =    255,397
                                            F(26, 255372)     =    5693.21
                                            Prob > F          =     0.0000
                                            R-squared         =     0.5956
```

lnwage	Coefficient	Linearized std. err.	t	P>|t|	[95% conf.	interval]
1.age15to19	-.2944529	.0186711	-15.77	0.000	-.3310477	-.2578582
1.age20to24	-.1243701	.0182009	-6.83	0.000	-.1600435	-.0886968
1.age25to29	.025843	.018027	1.43	0.152	-.0094896	.0611755
1.age30to34	.1363018	.0179452	7.60	0.000	.1011297	.1714738
1.age35to39	.1812695	.0178649	10.15	0.000	.1462547	.2162843
1.age40to44	.1943971	.017866	10.88	0.000	.1593802	.229414
1.age45to49	.2024796	.0178588	11.34	0.000	.1674768	.2374824
1.age50to54	.2021749	.017838	11.33	0.000	.167213	.2371368
1.age55to59	.1651815	.0178463	9.26	0.000	.1302032	.2001598
1.age60to64	.130169	.0180876	7.20	0.000	.0947177	.1656203
1.age65to69	.0480554	.0198947	2.42	0.016	.0090624	.0870485
1.age70plus	0	(omitted)				
1.married	.0620869	.0037148	16.71	0.000	.054806	.0693678
1.commonlaw	.0485002	.0040315	12.03	0.000	.0405985	.0564018
1.widow	.033452	.011815	2.83	0.005	.0102949	.0566091
1.separated	.0589362	.0078846	7.47	0.000	.0434826	.0743898
1.divorced	.0661085	.0067772	9.75	0.000	.0528254	.0793917
1.single	0	(omitted)				
1.grade0to8	-.7073227	.0143423	-49.32	0.000	-.7354332	-.6792122
1.somehighschool	-.6246269	.0073619	-84.85	0.000	-.639056	-.6101977
1.highschoolgrad	-.4913545	.0053762	-91.39	0.000	-.5018918	-.4808172
1.somepostsecondary	-.4724909	.0070973	-66.57	0.000	-.4864014	-.4585805
1.postsecondarygrad	-.3714678	.0046096	-80.59	0.000	-.3805025	-.3624332
1.bachdegree	-.147591	.0048853	-30.21	0.000	-.1571661	-.138016
1.abovebach	0	(omitted)				
1.immigrant less10	-.2443295	.0054672	-44.69	0.000	-.255045	-.2336139
1.immigrant_more10	-.1148584	.0039862	-28.81	0.000	-.1226713	-.1070454
1.nonimmigrant	0	(omitted)				
1.fulltime	.9605792	.0043999	218.32	0.000	.9519555	.9692029
1.parttime	0	(omitted)				
1.city_size	.0571822	.0025476	22.45	0.000	.052189	.0621753
_cons	6.122728	.0184492	331.87	0.000	6.086568	6.158888

Appendix D: Oaxaca Decomposition Results

Table 3: Two-fold Oaxaca Decomposition Results (2021)

```
Blinder-Oaxaca decomposition

Number of strata =       1                 Number of obs    =     510,043
Number of PSUs   = 510,043                 Population size = 194,370,746
                                           Design df        =     510,042
                                          Model              =     linear
Group 1: SEX = 1                          N of obs 1         =     254645
Group 2: SEX = 2                          N of obs 2         =     255398
```

lnwage	Coefficient	Linearized std. err.	t	P>|t|	[95% conf.	interval]
overall						
group 1	6.93237	.0019251	3601.04	0.000	6.928597	6.936143
group 2	6.689728	.0019885	3364.16	0.000	6.685831	6.693626
difference	.2426413	.0027677	87.67	0.000	.2372167	.248066
explained	.0888486	.0021253	41.81	0.000	.0846831	.0930141
unexplained	.1537927	.001846	83.31	0.000	.1501746	.1574108
explained						
AGE 12	.0007816	.0002586	3.02	0.003	.0002747	.0012884
MARSTAT	-.0033315	.0003315	-10.05	0.000	-.0039812	-.0026817
EDUC	-.0284972	.0007004	-40.69	0.000	-.0298699	-.0271245
IMMIG	.0001287	.0002772	0.46	0.642	-.0004146	.000672
FTPTMAIN	.1201966	.001687	71.25	0.000	.11689	.1235031
city_size	-.0004295	.0000906	-4.74	0.000	-.0006071	-.0002519
unexplained						
AGE 12	-.0346928	.0043313	-8.01	0.000	-.0431821	-.0262036
MARSTAT	-.0585769	.0028259	-20.73	0.000	-.0641155	-.0530384
EDUC	-.1553498	.0051288	-30.29	0.000	-.165402	-.1452976
IMMIG	-.0303426	.0088249	-3.44	0.001	-.0476393	-.013046
FTPTMAIN	-.2567258	.0081676	-31.43	0.000	-.272734	-.2407176
city size	-.0197362	.0020124	-9.81	0.000	-.0236804	-.0157921
_cons	.709217	.0154388	45.94	0.000	.6789573	.7394766

Table 4: Threefold Oaxaca Decomposition Results (2021)

Blinder-Oaxaca decomposition	Number of obs = 510,043
	Model = linear
Group 1: SEX = 1	N of obs 1 = 254645
Group 2: SEX = 2	N of obs 2 = 255398

lnwage	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
overall						
group_1	6.922326	.0013548	5109.32	0.000	6.919671	6.924982
group_2	6.669057	.0014256	4678.08	0.000	6.666263	6.671852
difference	.2532691	.0019667	128.78	0.000	.2494144	.2571237
endowments	.1144882	.0016331	70.10	0.000	.1112873	.1176891
coefficients	.1803296	.0013678	131.84	0.000	.1776488	.1830104
interaction	-.0415488	.0006276	-66.20	0.000	-.0427788	-.0403187
endowments						
AGE_12	.0001738	.0001421	1.22	0.221	-.0001048	.0004524
MARSTAT	-.0014304	.0002914	-4.91	0.000	-.0020014	-.0008593
EDUC	-.0274258	.0004309	-63.65	0.000	-.0282703	-.0265812
IMMIG	-.0004953	.0001573	-3.15	0.002	-.0008037	-.0001869
FTPTMAIN	.1437428	.0013717	104.79	0.000	.1410544	.1464312
city_size	-.000077	.0000397	-1.94	0.052	-.0001548	8.32e-07
coefficients						
AGE_12	-.0250875	.0031338	-8.01	0.000	-.0312296	-.0189454
MARSTAT	-.063045	.0019658	-32.07	0.000	-.0668978	-.0591922
EDUC	-.174929	.0033391	-52.39	0.000	-.1814735	-.1683846
IMMIG	-.0282168	.0066675	-4.23	0.000	-.0412849	-.0151487
FTPTMAIN	-.276804	.0041717	-66.35	0.000	-.2849803	-.2686276
city_size	-.0105564	.000907	-11.64	0.000	-.012334	-.0087788
_cons	.7589682	.0103012	73.68	0.000	.7387783	.7791581
interaction						
AGE_12	.0000401	.0000331	1.21	0.227	-.0000249	.0001051
MARSTAT	.0006618	.0001362	4.86	0.000	.0003949	.0009288
EDUC	-.014115	.0003357	-42.05	0.000	-.0147729	-.013457
IMMIG	-.0000509	.0000201	-2.53	0.011	-.0000904	-.0000114
FTPTMAIN	-.0279983	.000495	-56.57	0.000	-.0289684	-.0270282
city_size	-.0000865	.0000449	-1.93	0.054	-.0001744	1.40e-06

Appendix E: Stata do-file

Stata do-file for OLS Regression analysis & Blinder-Oaxaca Decomposition: page 1

Note: Please access drive for materials such as Stata, R outputs and LFS datasets codebook.

LFS_2021 (Apr 11 22) - Printed on 4/20/2022 9:55:32 AM

```
cd "D:\Desktop\ECON5600\PROC A\Data"

*Import LFS 2021 Datast into Stata
import excel "D:\Desktop\ECON5600\PROC A\Data\2021 LFS.xlsx", sheet("2021 LFS") firstrow

*Data cleaning/ manipulation
drop if REC_NUM ==.
drop if FTPTMAIN==.
drop if HRLYEARN==.

ssc install oaxaca, replace

generate hrlywage = HRLYEARN/100
generate hours = UTOTHRS/10
generate average_weekly_wage = hrlywage*hours
generate lnwage = log(average_weekly_wage)

generate city_size = 0 if CMA == 0
replace city_size=1 if (CMA==1 | CMA==2 | CMA==3 | CMA==4 | CMA==5 | CMA==6 | CMA==7 | CMA==8 | CMA==9)


*********************************************
*********Generate dummy variables**********
*********************************************

** Age **
gen age15to19=(AGE_12==1)
gen age20to24=(AGE_12==2)
gen age25to29=(AGE_12==3)
gen age30to34=(AGE_12==4)
gen age35to39=(AGE_12==5)
gen age40to44=(AGE_12==6)
gen age45to49=(AGE_12==7)
gen age50to54=(AGE_12==8)
gen age55to59=(AGE_12==9)
gen age60to64=(AGE_12==10)
gen age65to69=(AGE_12==11)
gen age70plus=(AGE_12==12)

** Gender **
gen male=(SEX==1)

** Marital Status **
gen married=(MARSTAT==1)
gen commonlaw=(MARSTAT==2)
gen widow=(MARSTAT==3)
gen separated=(MARSTAT==4)
gen divorced=(MARSTAT==5)
gen single=(MARSTAT==6)

** Education **
gen grade0to8=(EDUC==0)
gen somehighschool=(EDUC==1)
gen highschoolgrad=(EDUC==2)
gen somepostsecondary=(EDUC==3)
gen postsecondarygrad=(EDUC==4)
gen bachdegree=(EDUC==5)
gen abovebach=(EDUC==6)

** Immigrant Status **
gen immigrant_less10=(IMMIG==1)
gen immigrant_more10=(IMMIG==2)
```

Stata do-file for OLS Regression analysis & Blinder-Oaxaca Decomposition: page 2

Note: Please access drive for materials such as Stata, R outputs and LFS datasets codebook.


```
gen nonimmigrant=(IMMIG==3)

** Full- and Part-Time **
gen fulltime=(FTPTMAIN==1)
gen parttime=(FTPTMAIN==2)

****************************************
******** add population weights ********
****************************************
svyset [pweight=FINALWT]


****************************************
********* What is r-squared TEST *******
****************************************
svy: reg lnwage i.age15to19 i.age20to24 i.age25to29 i.age30to34 i.age35to39 i.age40to44 i.age45to49 i
.age50to54 i.age55to59 i.age60to64 i.age65to69 i.age70plus i.married i.commonlaw i.widow i.separated
i.divorced i.single i.grade0to8 i.somehighschool i.highschoolgrad i.somepostsecondary i.
postsecondarygrad i.bachdegree i.abovebach i.immigrant_less10 i.immigrant_more10 i.nonimmigrant i.
fulltime i.parttime i.city_size i.male

*****************************************
******* MALES - OLS Regressions *********
*****************************************
preserve

keep if SEX == 1

svy: reg lnwage i.age15to19 i.age20to24 i.age25to29 i.age30to34 i.age35to39 i.age40to44 i.age45to49 i
.age50to54 i.age55to59 i.age60to64 i.age65to69 i.age70plus i.married i.commonlaw i.widow i.separated
i.divorced i.single i.grade0to8 i.somehighschool i.highschoolgrad i.somepostsecondary i.
postsecondarygrad i.bachdegree i.abovebach i.immigrant_less10 i.immigrant_more10 i.nonimmigrant i.
fulltime i.parttime i.city_size

restore

*****************************************
****** FEMALES - OLS Regressions ********
*****************************************
preserve

keep if SEX == 2

svy: reg lnwage i.age15to19 i.age20to24 i.age25to29 i.age30to34 i.age35to39 i.age40to44 i.age45to49 i
.age50to54 i.age55to59 i.age60to64 i.age65to69 i.age70plus i.married i.commonlaw i.widow i.separated
i.divorced i.single i.grade0to8 i.somehighschool i.highschoolgrad i.somepostsecondary i.
postsecondarygrad i.bachdegree i.abovebach i.immigrant_less10 i.immigrant_more10 i.nonimmigrant i.
fulltime i.parttime i.city_size

restore

*****************************************
********** Oaxaca Decomposition *********
*****************************************
**three-fold
oaxaca lnwage AGE_12 MARSTAT EDUC IMMIG FTPTMAIN city_size, by(SEX) threefold(reverse)

**two_fold
oaxaca lnwage AGE_12 MARSTAT EDUC IMMIG FTPTMAIN city_size, by(SEX) pooled svy

```