2024

The Effect of Gender-Based Occupational Segregation on Women's Earnings in the United States and Canada

Jadyn Grace Milius
University of Northern Iowa

THE EFFECTS OF GENDER-BASED OCCUPATIONAL SEGREGATION ON WOMEN'S EARNINGS IN THE UNITED STATES AND CANADA

A Thesis Submitted

in Partial Fulfillment

of the Requirements for the Designation

University Honors

Jadyn Grace Milius

University of Northern Iowa

May 2024

This Study by: Jadyn Milius

Entitled: The Effects of Gender-Based Occupational Segregation on Women's Earnings in the United States and Canada

has been approved as meeting the thesis or project requirement for the Designation University Honors

Approved by:

Shahina Amin, Honors Thesis Advisor

Dr. Jessica Moon Asa, Director, University Honors Program

Abstract

The gender wage gap continues to prove a contentious topic with some going as far as to question its existence. However, scholarly works continue to prove it is real and impactful. More fruitful debates have arisen about the source of this disparity in wages. This paper seeks to analyze one possible explanation for the pay gap: gender-based occupational segregation. I used OLS regressions based on women working full-time between the ages of sixteen and sixty-five to measure this. I find a negative effect of segregated occupations on women's earnings across countries and years. On a similar note, the percentage of females in an occupation correlates with lower earnings across the years in Canada and the United States.

I. Introduction

The gender pay gap persists to be a significant issue affecting women's earnings. While there has been some evidence of a decrease in the difference between men's and women's earnings in the United States, progress appears to be slowing down. In the 1980s, the United States made substantial advancements, but this growth slowly declined and became inconsistent (Blau & Kahn, 2017). Although this may not be the ideal progression, progress has still been made. Before the 1980s, women consistently earned approximately 60% of what their male counterparts made; however, by 2014, this percentage had increased by nearly twenty points, to 79% (Blau & Kahn, 2017). Despite this improvement, a pay ratio of 79% is still far from ideal, particularly considering the advancements in women's educational attainment since the 1980s. As of 2011, women have surpassed men in years of schooling completed and the percentage holding advanced degrees. Although women have not surpassed men in years of full-time experience, they have reduced the difference by over five years (Blau & Kahn, 2017). Overall, women are making great strides to try to reduce the gap, yet it remains.

This story is not unique to the United States either, Canada is facing a similar problem. The gender pay ratio in Canada has stayed consistent at around 0.73 from 2009 to 2017 (Moyser, 2019). This means that for every Canadian dollar a man earns, a woman earns 0.73 of that. This is despite women becoming more educated. Over 40% of women aged 25 to 34 in Canada have at least a bachelor's degree compared to a little over 29% of men in the same age range (Moyser, 2019). Unfortunately, different studies have yielded varying results, most likely due to the specific populations included. As Drolet and Amini (2023) note, the difference in the pay gap when studying Canada depends on whether one includes Canadian-born, indigenous, and immigrants into the population. When comparing Canadian-born women and men, the women earn 9.2% less (Drolet & Amini, 2023). When one compares Canadian-born men to indigenous women, the women earn over 20% less in 2022 (Drolet & Amini, 2023). This leads to numerous calculations of the pay gap in Canada, yet most calculations agree that women still get paid less. Again, just as with the US, women in Canada have been making great strides to acquire comparable labor market qualifications like education and experience (Drolet & Amini, 2023). This leads many to wonder what causes the difference in wages in the US and Canada.

This paper aimed to explore one possible explanation for this persistent discrepancy using gender-based occupational segregation. Gender-based occupational segregation is the tendency for certain jobs to have a largely female or male workforce (Gauchat et al., 2012). To measure this segregation, I used the Otis Dudley Duncan and Beverly Duncan index. I then used regression analysis to test for the significant variables that affect a woman's annual earnings, with a specific focus on whether the variable representing the segregation index of the occupation group is statistically significant. I used regression analysis to further study if the type of segregation, specifically the proportion of females in an occupation is statistically significant. This was a comparative study between the United States and Canada in 2011 and then a separate analysis of the United States in 2022. The independent analysis provided a more updated view of this issue.

The regressions focused on women between the ages of sixteen and sixty-five working at least thirty-five hours a week. I find that a highly segregated occupation negatively impacts a woman's annual earnings in the US and Canada. Additionally, I find evidence that women working in occupations with a higher proportion of females have lower earnings in both countries in the year 2011 and in the United States in 2022.

II. Literature Review

In a 2015 study based in the United Kingdom, a regression model was constructed to examine variables influencing hourly earnings. The study investigated factors such as age, employment status, gender, sector, occupation, region, tenure, size of organization, and various interaction variables. The results of the regression identified sector, job tenure, age, and gender as key variables affecting wages (Foxton % Massey, 2015). Another study based in Poland by Kompa and Witkowska (2018) looks at factors affecting wages. They use twelve variables to run ordered logit models with net income categorized into intervals as the dependent variable. The study finds that gender, age, education, size of firm, and occupation significantly affect wages. Kompa and Witkowska (2018) hypothesize that the gender pay gap in Poland could be attributed to occupational segregation, with women concentrated in low-wage jobs. They also suggest inter-industry pay differences, which are common across European countries, could be a cause.

The studies in Poland and the United Kingdom illustrate the significance of many variables in estimating earnings in European countries. Storrie, Lee, and Matzel (2023) analyze the disparity of weekly wages between men and women from 2000 to 2020 in the US. They use the natural log of weekly wages as the dependent variable and many of the same independent variables as the previous two studies, along with a few additional factors. The results found age, education, race, marital status, number of children, and metropolitan status to be statistically significant at the 1% significance level in the years between 2000 and 2020 (Storrie et al., 2023). Although not for the same purpose as this paper, Lemieux (1993) ran an OLS regression on the logarithm of hourly wages in Canada and the United States. The purpose of this research was to look at how unionization plays a role in determining wages. This paper confirmed many of the above factors to be important when studying the United States and Canada like age, education, marital status, race, occupation, industry, and region (Lemieux, 1993). These studies demonstrate that wage-determining characteristics remain relatively unchanged between countries.

The previous studies have highlighted the importance of education, experience, and other factors in individual wage determination. However, when looking to compare wage determination between the sexes, their usefulness over time has diminished. As a result, researchers have turned towards occupational segregation as a potential cause for the difference between men's and women's wages. Gauchat et al. (2012), define occupation segregation as "the systematic concentration of groups of workers (women) in particular jobs". Hegewisch et al. (2010) reported that in 2009, nearly 40% of women in the United States workforce were employed in occupations that were at minimum 75% female and only 5.5% of women worked in jobs that were at least 75% male. This idea of gender-based or sex-based occupational segregation appears to be a trend that is evident in countries all over the world (Allen, 2002). This is significant because evidence suggests a negative correlation between the proportion of females in an occupation and the wages for that occupation (Meara et al., 2019).

One effort to try to see the effect of occupation segregation on wages was conducted in 2021 by Jane Nyawira Maina. Maina (2021) ran a regression where the dependent variable is the logarithm of the monthly earnings of individuals in Kenya. She conducted separate regressions for men's earnings and women's earnings. In this regression, there were several variables to control for factors such as experience, education, marital status, hours worked, union status, and industry (Maina, 2021). To account for occupational segregation, Maina (2021) added a variable that accounted for the proportion of women and men in the measured occupation. The results of this regression showed that the proportion of women in an occupation and female wages have a negative relationship while the proportion of men in an occupation has a positive relationship with female wages (Maina, 2021). Consequently, this paper provides a method of accounting for occupational segregation on wages.

There are other ways to look at occupational segregation besides using the proportion of women in an occupation. One way to measure the degree of segregation is using Otis Dudley Duncan and Beverly Duncan's index of dissimilarity. This measure gives a value between zero and one and tells what percent of the labor force would have to change their occupation to achieve a balanced gender makeup (Hegewisch et al., 2010). Zero would mean no segregation, and one would indicate that it is perfectly segregated. This can be used for specific occupations or the labor force as a whole (Hegewisch et al., 2010). As mentioned above, this plays a part in wages because some studies have found that occupations with more females have lower median wages compared to those with fewer women (Wrohlich, 2017). The Duncan-Duncan index adds another option for a variable to measure gender-based occupational segregation.

III. Data

A. United States

To start this section, I will discuss the variables included in the United States regressions for both years, 2022 and 2011. For the regressions, I retrieved data from the Integrated Public Use Microdata Series (IPUMS). I used American Community Survey (ACS) data from 2011 and 2022. The sample size, after adjusting for ages between sixteen and sixty-five and at least thirty-five hours worked per week, was 532,679 for 2022 and 329,415 for 2011 (Ruggles, 2024). The year 2011 was chosen as that is the most updated information available about Canada found on IPUMS.

All variables included in the models were inspired by previous literature and theory in some way. The dependent variable for all US regressions was the natural logarithm of an individual's pre-tax salary in dollars for that specific year. This does not include money from personal businesses or farms. Further, to avoid working in intervals and making estimations, the dependent variable for all regressions was reported as an annual figure rather than weekly or hourly.

The independent variables were a combination of personal, family, and job-related characteristics. Personal characteristics incorporated in the model include education, sex, race, metropolitan status, region, age, and age squared. Education was measured as educational attainment or the highest year of school completed, categorized into four groups. The baseline category was those with less than a grade 12 education, while the other three categories were those who had completed grade 12, one to four years of college, and five or more years of college, respectively. The independent variable representing metropolitan status was a dummy variable, where zero represents individuals in metropolitan areas and one signifies those not in a metropolitan area or those in a mixed area. Concerning race, a value of zero was assigned to those who identify as white. Other races were categorized with one representing those who identify as African American or Black, two for American Indians or Alaskan Natives, three for Asians, four for other races not mentioned above, and five for anyone who identifies as more than one major race. A region variable was also added. This variable splits the United States into four groups with the base category representing the Northeastern part of the US. For this paper, the characterization of certain states into regions followed IPUMS categorizations. Moreover, age and age squared were measured in years and functioned as proxy variables for experience. Including age squared accounts for the non-linear relationship between wage and age, addressing any potential leveling off of wage increases in the later stages of one's career.

Familial and job characteristics were also included. The number of children in the home will be measured as the number of children in the woman's care under the age of five. Marital status was expressed as a dummy variable with the baseline category being never married, followed by categories for those married and those separated, divorced, or widowed, respectively. A job-related independent variable was industry. The industry variable was divided into 13 categories with professional and business services as the base category with a value of zero (Blau &Winkler, 2022).

Finally, the key variables of interest revolve around occupational segregation: "index" and "female." To make these variables, all occupations were categorized into 10 major groups (Blau & Winkler, 2022). These 10 categories align well with categories from the United States Bureau of Labor Statistics (BLS). The BLS provides data about the total number of people employed in each of these major categories and the number of men and women employed in each of these categories. This was then used to calculate the percentage of females and males in each category. Given that information, the Duncan-Duncan Segregation Index was used to find 10 separate segregation index values for each occupation in the respective years. This value ranges from zero to one, with a higher number meaning more segregation. To allow for a more intuitive interpretation of regression results, the value calculated was multiplied by 100%. As a result, the values will range from 0% to 100%. The "female" variable is simply the proportion of females in each major occupation. The specific values can be seen in the appendix.

A further note is that the variables "index" and "female" will not be included in the same regression. These variables essentially measure the same thing, occupational segregation, but provide a little different insight. Therefore, for each year and country, a regression will be run including all the other independent variables and "index" and then all the other independent variables and "female." The segregation index is derived as:

$$\textbf{Segregation Index} = \frac{1}{2}\Sigma|M_i - F_i|$$

M_i= Percentage of males in the labor force in i occupation
F_i= Percentage of females in the labor force in i occupation

Table 1: US 2022 Summary Statistics

VARIABLE	MEAN	STD. DEV.	MIN	MAX
REGION	1.656101	1.020228	0	3
METRO	.2246156	.4173293	0	1
NCHLT5	.1363804	.4194162	0	7
AGE	42.47759	12.71406	16	65
AGESQ	1965.993	1087.779	256	4225
MARST	.8712151	.6655795	0	2
RACE	1.093602	1.773178	0	5
EDUC	1.866349	.7610505	0	3
INDEX	3.293591	1.863009	.2559	5.12
IND	6.744197	3.575204	0	12
UHRSWORK	42.32589	6.99164	35	99
INCWAGE	62418.23	62841.64	4	791000
LNINCWAGE	4.651745	.379435	.60206	5.898176
FEMALE	.5363991	.1228798	.0420359	.7188121

Table 2: US 2011 Summary Statistics

VARIABLE	MEAN	STD. DEV.	MIN	MAX
REGION	1.506443	1.039168	0	3
METRO	.369549	.4826834	0	1
NCHLT5	.1437913	.4257623	0	5
AGE	43.07275	12.25627	16	65
AGESQ	2005.477	1040.328	256	4225
MARST	.9547713	.6596513	0	2
RACE	.5139171	1.15192	0	5
EDUC	1.736688	.7625561	0	3
INDEX	3.816276	2.102907	.363648	6.751475
IND	6.982026	3.431362	0	12
UHRSWORK	42.0799	6.746267	35	99
INCWAGE	42664.81	38944.81	4	607000
LNINCWAGE	4.493498	.3781165	.60206	5.783189
FEMALE	.5513387	.1382224	.0228772	.7340438

B. Canada

As noted above, the regression for Canada was based on the year 2011. The data I collected was sourced from IPUMS International. The sample size after making the same adjustments as above was 133,697 women.

All variables included in the models were inspired by previous literature and theory in some way. The dependent variable was the natural logarithm of a woman's earnings for that year including money made from a self-owned business or a farm in Canadian dollars based on the year 2011. This was a slightly different dependent variable due to IPUMS not having the stand-alone pre-tax wage earned from an employer available for Canada like it does the US.

In terms of independent variables, the Canada regression also included a combination of personal, family, and job-related factors. Personal variables included in the model were education, sex, indigenous status, age, and age squared. Just like the United States' regressions, the education variable was measured as the highest level attained by the woman. Education will similarly be categorized into four groups with the baseline category representing less than a high school diploma, one representing those with either a high school diploma, trades certification, or apprenticeship certification, two representing those with a college degree equivalent to or less than a bachelor's degree, and three representing those with anything higher than a bachelor's degree. Race was excluded from the Canada regression as IPUMS did not have data on it in 2011. To try to account for some of this difference a variable representing if one belongs to an indigenous group was added. For this variable, the baseline category with a value of zero represented non-indigenous people and one represented indigenous people. Age and age squared were measured and added to the regression for the same reasons as mentioned above. Other missing variables include a regional variable and a metropolitan status variable. Again, IPUMS did not have this information and so, unfortunately, must be excluded. The familial characteristics of the number of children and marital status for this regression were measured the same as the regressions above.

Lastly, there were the key variables "index" and "female," which were also included in the Canada regression. Again, to form these variables occupations were grouped into 10 major categories, this time based on the 2016 National Occupational Classification system for Canada. This allowed the percentages of total employment by gender in 2011 based on these 10 major categories to be found easily. Given that information, the Duncan and Duncan Segregation Index was used to find 10 separate segregation index values for 2011. Similarly, the "female" variable was the proportion of females in each of the 10 major occupation groups.

Table 3: Canada 2011 Summary Statistics

VARIABLE	MEAN	STD. DEV.	MIN	MAX
NCHLT5	.1205188	.3887314	0	4
AGE	41.72195	11.46097	16	65
AGESQ	1872.074	950.8118	256	4225
MARST	.9034085	.5863187	0	2
INDIG	.0281831	.1654964	0	1
EDUC	1.684518	.7548327	0	3
INDEX	4.951372	2.101546	.3514	12.3932
FEMALE	.5862683	.1782352	.0559457	.8133512
IND	6.742081	3.691199	0	12
HRSACTUAL1	40.97645	6.891189	35	80
INCEARN	46652.16	35083.28	1	269200
LNINCEARN	4.537095	.4158852	0	5.430075

IV. Econometric Model and Methodology

With the given dependent and independent variables, ordinary least squares (OLS) regression models were constructed. This was used to test the statistical significance of each of the independent variables on the dependent variable, holding all other variables constant. I also tested for multicollinearity and heteroskedasticity. The following equation was used in both United States' regressions.

$$\ln INCWAGE = b_0 + b_1\mathit{EDUC} + b_2\mathit{RACE} + b_3\mathit{AGE} + b_4\mathit{AGESQ} + b_5\mathit{NCHLT5} + b_6\mathit{MARST} + b_7\mathit{IND} + b_8\mathit{METRO} + b_9\mathit{UHRSWORK} + b_{10}\mathit{REGION} + b_{11}\mathit{INDEX} \text{ OR } \mathit{FEMALE} + \varepsilon$$

Where:
lnINCWAGE = Logarithm of Individual’s annual pre-tax wage or salary income
EDUC = Highest year in school completed
RACE = Self-reported race
AGE = Age in years
AGESQ = Age squared
NCHLT5 = Number of children under the age of five in the household
MARST = Current marital status
IND = The industry an individual performs their occupation
METRO = The location of an individual's household about a metropolitan area
UHRSWORK= Usual number of hours worked per week
REGION= Current US region the individual resides in
INDEX = value between 0% and 100% calculated using the Duncan-Duncan Segregation Index
FEMALE= Proportion of females in an occupation
ε = stochastic random error term

The following regression was used in Canada’s regression due to the availability, or lack thereof, of variables in IPUMS.

$$\ln INCEARN = b_0 + b_1\mathit{EDUC} + b_2\mathit{AGE} + b_3\mathit{AGESQ} + b_4\mathit{NCHLT5} + b_5\mathit{MARST} + b_6\mathit{IND} + b_7\mathit{INDIG} + b_8\mathit{HRSACTUAL1} + b_9\mathit{INDEX} \text{ OR } \mathit{FEMALE} + \varepsilon$$

Where:
lnINCEARN = Logarithm of Individual’s annual earned income
EDUC = Highest attained certification/degree
AGE = Age in years
AGESQ = Age squared
NCHLT5 = Number of children under the age of five in the household
MARST = Current marital status
IND = The industry an individual performs their occupation
INDIG = One’s identification with an indigenous group
HRSACTUAL1= Actual number of hours worked per week
INDEX = value between 0% and 100% calculated using the Duncan-Duncan Segregation Index
FEMALE= Proportion of females in an occupation
ε = stochastic random error term

V. Regression Results

In the model, nearly all variables were found to be statistically significant at the 95% confidence level, with many matching theoretical expectations. The variable that was of key interest in this regression was the index variable. As expected, the regression produced a negative coefficient. This means that the more segregated an occupation is by gender, the less a woman is likely to earn. The specific coefficient can be interpreted as meaning for a 1% increase in the segregation index, women's wages are estimated to fall by about 1.8%.

To test for near-perfect multicollinearity among variables, I ran the Variance Inflation Factor (VIF) test resulting in an overall mean of 5.72 for the US 2022 model. There are a couple of variables, age and age squared, with mean values above 10. Both age and age squared have high VIF values since they are directly correlated to each other. However, this was needed to account for the non-linear relationship with wages. Overall, the model did not exhibit significant signs of near-perfect multicollinearity.

Furthermore, to test for heteroskedasticity in the model, I ran a Breusch-Pagan test. It resulted in a chi-squared value of 0.0000 showing the model suffered from heteroskedasticity. To correct for the effect of heteroskedasticity on the standard errors in the regression, I have run the regression using robust standard errors.

Table 4: United States 2022

LNINCWAGE	COEF.	STD. ERR.	T	P>|T|
INDEX	-.0179946	.0002477	-72.65	0.000
REGION				
1	-.0477025	.0013759	-34.67	0.000
2	-.056855	.001258	-45.20	0.000
3	-.0000171	.0014186	-0.01	0.990
1.METRO	-.0802497	.0010764	-74.55	0.000
NCHLT5	.0204325	.0011448	17.85	0.000

AGE	.0384437	.000301	127.74	0.000
AGESQ	-.0003849	3.40e-06	-113.20	0.000
MARST				
1	.0304741	.001201	25.37	0.000
2	-.0077361	.0015368	-5.03	0.000
RACE				
1	-.0758491	.0016462	-46.08	0.000
2	-.0755857	.004509	-16.76	0.000
3	.0106761	.0018764	5.69	0.000
4	-.0756043	.0020907	-36.16	0.000
5	-.0413424	.0015132	-27.32	0.000
EDUC				
1	.1183193	.0030151	39.24	0.000
2	.2626691	.0029899	87.85	0.000
3	.4291547	.0031105	137.97	0.000
IND				
1	.1032666	.0100237	10.30	0.000
2	-.0386303	.0037171	-10.39	0.000
3	-.0273962	.0021058	-13.01	0.000
4	-.1304633	.0020258	-64.40	0.000
5	-.0476971	.0028818	-16.55	0.000
6	.0338202	.0037508	9.02	0.000
7	.0130822	.0019811	6.60	0.000
8	-.2629423	.0068272	-38.51	0.000
9	-.0809855	.0015596	-51.93	0.000
10	-.2104488	.0025581	-82.27	0.000
11	-.1536908	.0029255	-52.54	0.000
12	-.0114879	.0020777	-5.53	0.000
UHRSWORK	.005797	.0000836	69.31	0.000
_CONS	3.455694	.0077087	448.28	0.000

Number of Observations	=	532,679
F(31, 532,647)	=	6132.81
Prob > F	=	0.0000
R-Squared	=	0.2962

Table 5: United States 2022 FEMALE

	COEF.	STD. ERR.	T	P>|T|

FEMALE	-.1377257	.0038732	-35.56	0.000

When replacing the index variable with the female variable, most of the results stayed consistent. A few individual coefficients within the industry dummy variable changed signs, but all other coefficients remained unchanged. For the full regression results, view the appendix. The female variable yielded statistically significant results at the 5% significance level. This model produced a negative coefficient indicating that holding all other factors equal, the proportion of females in an occupation is negatively correlated with women's wages. More specifically, for a one percentage increase in the proportion of females in an occupation a woman will earn 13.77% less than before. In this updated model, no significant signs of near-perfect multicollinearity were found, and robust standard errors were used to correct for any heteroskedasticity.

Table 6: United States 2011

LNINCWAGE	COEF.	STD. ERR.	T	P>|T|
INDEX	-.0110364	.0002677	-41.22	0.000
REGION				
1	-.0458786	.0016145	-28.42	0.000
2	-.0464647	.0014983	-31.01	0.000
3	-.0179387	.0017804	-10.08	0.000
1.METRO	-.1019946	.0012078	-84.44	0.000
NCHLT5	.0255355	.0015097	16.91	0.000
AGE	.0470526	.0003967	118.62	0.000
AGESQ	-.0004724	4.48e-06	-105.51	0.000
MARST				
1	.0274591	.0015695	17.50	0.000
2	-.00224	.0018766	-1.19	0.233
RACE				
1	-.0612615	.0018193	-33.67	0.000
2	-.0592733	.0059332	-9.99	0.000
3	-.0085582	.0027544	-3.11	0.002
4	-.0674959	.0034284	-19.69	0.000

5	-.0524889	.0045569	-11.52	0.000
EDUC				
1	.1569996	.003079	50.99	0.000
2	.2995383	.003064	97.76	0.000
3	.4751864	.0033078	143.66	0.000
IND				
1	.1111736	.0119428	9.31	0.000
2	-.0093997	.0057666	-1.63	0.103
3	-.0042998	.0028328	-1.52	0.129
4	-.0815917	.0027364	-29.82	0.000
5	.0375223	.0036989	10.14	0.000
6	.0417043	.0044146	9.45	0.000
7	.0485328	.0026905	18.04	0.000
8	-.2268076	.0091911	-24.68	0.000
9	-.0300031	.0022308	-13.45	0.000
10	-.1777599	.0033388	-53.24	0.000
11	-.1264665	.0039074	-32.37	0.000
12	.0611197	.0028047	21.79	0.000
UHRSWORK	.0059656	.0001082	55.16	0.000
_CONS	3.027846	.0099412	304.58	0.000

Number of Observations	=	329,415
F(31, 329383)	=	4019.42
Prob > F	=	0.0000
R-Squared	=	0.3188

For the United States 2011 model, the key variable, index, was found to have a negative coefficient. Again, this means that the more segregated an occupation is by gender, the less a woman is likely to earn in wages, all else being equal. When including the female variable in place of the index variable, the model produced statistically significant results at the 5% significance level. Again, the negative coefficient indicates a negative relationship between the proportion of females in an occupation and wages. A VIF test was run resulting in means less than 10 for both US 2011 regressions which demonstrates that neither model suffers from near-perfect multicollinearity. Robust standard errors were also used in both instances to correct for any heteroskedasticity present.

Table 7: United States 2011 FEMALE

	COEF.	STD. ERR	T	P>|T|
FEMALE	-.084025	.0043981	-19.10	0.000

Table 8: Canada 2011

LNINCEARN	COEF.	STD. ERR.	T	P>|T|
INDEX	-.0074126	.0005467	-13.56	0.000
NCHLT5	-.0690745	.0031681	-21.80	0.000
AGE	.0427614	.0007065	60.52	0.000
AGESQ	-.0004304	8.39e-06	-51.29	0.000
MARST				
1	.017368	.002813	6.17	0.000
2	.0092618	.003851	2.41	0.016
1.INDIG	-.0422548	.0065637	-6.44	0.000
EDUC				
1	.0693018	.0044794	15.47	0.000
2	.1924259	.0044541	43.20	0.000
3	.3071418	.005498	55.86	0.000
IND				
1	.2192117	.0137243	15.97	0.000
2	.028219	.0090893	3.10	0.002
3	-.0062518	.0053737	-1.16	0.245
4	-.0381278	.0045583	-8.36	0.000
5	.076156	.0064635	11.78	0.000
6	.074168	.0071636	10.35	0.000
7	.09081	.004666	19.46	0.000
8	-.2581717	.0129444	-19.94	0.000
9	.0285655	.0039462	7.24	0.000
10	-.1580642	.0055903	-28.27	0.000
11	-.1212914	.006326	-19.17	0.000
12	.1224591	.0045121	27.14	0.000
HRSACTUAL1	.0002298	.0001865	1.23	0.218
_CONS	3.417399	.0168646	202.64	0.000

Number of Observations	=	133,697
F(23, 133673)	=	1198.66
Prob > F	=	0.0000
R-Squared	=	0.1732

When shifting the focus away from the United States to Canada many variables remained significant at the 5% significance level. Importantly, the index variable resulted in statistically significant results and a negative coefficient. For a one percent increase in the Duncan-Duncan Segregation Index women's earnings decrease by approximately .74%. In the regression with the female variable, there is again a negative relationship between the proportion of women in an occupation and women's earnings. More specifically, for a one percent increase in the proportion of women, women's wages decrease by approximately 7.4%. In both Canada regression models, the VIF test was run resulting in means less than 10. This demonstrates that neither model suffers from near-perfect multicollinearity. Robust standard errors were also used in both instances to correct for any heteroskedasticity.

Table 9: Canada 2011 FEMALE

	COEF.	**STD. ERR**	**T**	**P>|T|**
FEMALE	-.0743464	.006758	-11.00	0.000

VI. Conclusions

Women in the United States and Canada continue to earn less than their male counterparts making it imperative to continue research on what could be the cause. The above research aimed to test the effect of gender-based occupational segregation on women's earnings to see if it is significant. The results of the regression analysis provided evidence supporting the statistical significance of nearly all variables in relation to annual earnings. Concerning the specific objectives of this regression, the individual outcomes for the segregation index were as expected. Statistically speaking, in the United States, the degree of gender-based segregation of a woman's occupation proves to be relevant to their annual earnings. The more segregated an occupation, the less a woman should expect to earn, all else equal. Regression results also provided evidence that the proportion of females in an occupation is negatively correlated to one's wages in the United States in 2011 and 2022. In Canada a similar story emerges; there is a negative relationship between the degree of segregation in an occupation and women's earnings. In terms of the proportion of females in an occupation and women's earnings, a negative relationship is seen again.

Based on these results, more research on occupational segregation is recommended. Specifically, I recommend diving deeper into specific occupations rather than looking at a broad 10 categories. This may provide more information that could be used to help address the issue if a solution is desired. Further, continued research into why the effects of segregation are different across countries is recommended. Lastly, I recommend looking deeper into the coefficient on the proportion of females in an occupation is negative across countries and years. This could include research into the devaluation of female labor or specific job characteristics not measured in this study.

This regression has many limitations that, if addressed, could provide better results. The first of these limitations is the restriction of data to only those whose sex is female. The complexities of gender identity were not able to be added to these regressions which could provide valuable insight. Secondly, there is a limitation with categorizing occupations into only 10 major categories. Doing this makes many generalizations that if avoided could add valuable information to this research. For example, this may lead some highly segregated occupations being grouped in with integrated jobs which may overshadow their effect on an individual's earnings. Additionally, not all variables relevant to one's earnings were included in the above regressions. There were many variables that were excluded for various reasons including availability, time, and energy like unionization status, private versus public sector, and one's language. Adding these variables could provide more accurate results. The last limitation necessary to mention is not having exact matching variables when comparing the United States and Canada. IPUMS did not have the exact matching variables needed for this paper. This leads to imperfect comparisons. To be more precise in the analysis, coordinating variables would be preferred.

Consequently, this paper adds further research into the many factors that can determine wages, especially for women. It models occupational segregation in two ways, demonstrating how two different variables can add new dimensions to the research. The Duncan-Duncan index provides an overall view, while the female proportion variable adds a gender-specific perspective. It is evident that occupations continue to be segregated by gender leading to lower earnings for women. To achieve fair compensation for women both in the United States and Canada, it is imperative that occupational segregation is understood and considered.

References

Allen, J., & Sanders, K. (2002). Gender Gap in Earnings at the Industry Level. *The European Journal of Women's Studies* 9(2): 163-180. Retrieved December 12, 2023, from https://library.fes.de/libalt/journals/swetsfulltext/13529010.pdf

Blau, F.D., & Winkler, A. E. (2022). *The Economics of Women, Men, and Work* (9thed.). Oxford University Press.

Blau F. D., & Khan, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, 55(3), 789-865. DOI: 10.1257/jel.20160995

Drolet, M., & Amini, M. M. (2023). Intersectional Perspective on the Canadian Gender Wage Gap. *Statistics Canada*. Retrieved December 12, 2023, from https://www150.statcan.gc.ca/n1/pub/45-20-0002/452000022023002-eng.htm

Foxton, F., & Massey, F. (2015). Analysis of Factors Affecting Earnings Using Annual Survey of Hours and Earnings. *Office for National Statistics.* Retrieved December 12, 2023, from https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/articles/analysisoffactorsaffectingearningsusingannualsurveyofhoursandearnings/2016

Gauchat, G., et al. (2012). Occupational Gender Segregation, Globalization, and Gender Earnings Inequality in U.S. Metropolitan Areas. *Gender & Society,* 26(5), 718-747. https://doi.org/10.1177/0891243212453647

Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *American Economic Review,* 104(4): 1091-1119. DOI: 10.1257/aer.104.4.1091

Hegewisch, A., et al. (2010). *Separate and Not Equal? Gender Segregation in the Labor Market and the Gender Wage Gap.* Institute for Women's Policy Research. Retrieved December

12, 2023, from https://iwpr.org/separate-and-not-equal-gender-segregation-in-the-labor-market-and-the-gender-wage-gap/#:~:text=And%20occupational%20segregation%20is%20a,such%20as%20medicine%20or%20law.

Kompa, K., & Witkowska, D. (2017). Factors Affecting Men's and Women's Earnings in Poland. Economic Research, 31, 252-269 https://doi.org/10.1080/1331677X.2018.1426480

Lemieux, T. (1993). *Unions and Wage Inequality in Canada and the United States.* University of Chicago Press 69-108. https://core.ac.uk/download/pdf/6778269.pdf

Maina , J. (2021).*The Effects of Occupational Segregation on Gender Wage Gap in Kenya*. Uon Digital Repository. Retrieved December 12, 2023, from http://erepository.uonbi.ac.ke/bitstream/handle/11295/160152/Maina_The%20Effects%20of%20Occupational%20Segregation%20on%20Gender%20Wage%20Gap%20in%20Kenya.pdf?sequence=3.

Meara, K., et al. (2019). *The Gender Pay Gap in the USA: A Matching Study*. Journal of Population Economics, 33, 271-305. Retrieved December 12, 2023, from https://link.springer.com/article/10.1007/s00148-019-00743-8

Moyser, M. (2019). *Measuring and Analyzing the Gender Pay Gap: A Conceptual and Methodological Overview.* Center for Gender, Diversity, and Inclusion Statistics, Statistics Canada. Retrieved December 12, 2023, from https://epe.lac-bac.gc.ca/100/201/301/weekly_acquisitions_list-ef/2019/19-35/publications.gc.ca/collections/collection_2019/statcan/452000022019001-eng.pdf

Statistics Canada. Table 14-10-0335-02 Proportion of women and men employed in occupations, annual, inactive. https://doi.org/10.25318/1410033501-eng

Ruggles, S., et al. IPUMS USA: Version 15.0 [dataset]. Minneapolis, MN: IPUMS, 2024. https://doi.org/10.18128/D010.V15.0

Storrie, C. L., et al. (2023). *The Gender Wage Gap in the New Millennium An Analysis of the United States 2000-2020.* Journal of Business Diversity, 23(2). https://doi.org/10.33423/jbd.v23i2.6165

U.S. Bureau of Labor Statistics. (2012). *Employed persons by occupation, sex, and age*. U.S. Bureau of Labor Statistics. https://www.bls.gov/cps/aa2011/cpsaat09.htm

U.S. Bureau of Labor Statistics. (2023). *Employed persons by occupation, sex, and age*. U.S. Bureau of Labor Statistics. https://www.bls.gov/cps/aa2022/cpsaat09.htm

Wrohlich, K. (2017). *Gender Pay Gap Varies Greatly by Occupation*. DIW Economic Bulletin 7(43): 429-435. Retrieved December 12, 2023, from https://www.econstor.eu/bitstream/10419/170734/1/1002315573.pdf

Appendix

Occupation Groupings, Index & Female Values for United States (rounded values)

Occupations	Segregation Index 2022	Female Proportion 2022	Segregation Index 2011	Female Proportion 2011
Management, Business, and Financial Operations Occupations	0.7299	0.4484	1.1602	0.4314
Professional and Related Occupations	4.8012	0.5657	4.5448	0.5711
Service Occupations	3.2943	0.5701	3.2102	0.5591
Sales and Related Occupations	0.48	0.4945	0.5876	0.4956
Office and Administrative Support Occupations	5.12	0.7188	6.7515	0.7340
Farming, Fishing, and Forestry Occupations	0.2559	0.2622	0.3636	0.2158
Construction and Extraction Occupations	4.5511	0.0424	4.5614	0.0229
Installation, Maintenance, and Repair Occupations	2.623	0.0420	3.0381	0.0354
Production Occupations	1.7828	0.2978	2.1553	0.2845
Transportation and Material Moving Occupations	3.7523	0.2180	3.8155	0.1493

Occupation Groupings, Index, & Female Values for Canada (rounded values)

Occupations	Segregation Index	Female Proportion
Management Occupations	2.2788	0.3570
Business, Finance, and Administration Occupations	7.2317	0.6921

Natural and Applied Sciences and Related Occupations	3.6775	0.2208
Health Occupations	4.5678	0.8135
Occupations in Education, Law, and Social, Community and Government Services	4.4488	0.6817
Occupations in Art, Culture, Recreation and Sport	0.3514	0.5331
Sales and Service Occupations	4.7994	0.5725
Trades, Transport, and Equipment Operators and Related Occupations	12.3932	0.0559
Natural Resources, Agriculture, and Related Production Occupations	1.2723	0.1645
Occupations in Manufacturing and Utilities	1.7778	0.2911

US 2022 Female Results

LNINCWAGE	COEF.	STD. ERR.	T	P>|T|
REGION				
1	-.048629	.0013807	-35.22	0.000
2	-.0572866	.0012621	-45.39	0.000
3	.0000435	.0014234	0.03	0.976
1.METRO	-.0819989	.0010797	-75.94	0.000
NCHLT5	.0209863	.0011473	18.29	0.000
AGE	.0389701	.0003015	129.25	0.000
AGESQ	-.00039	3.41e-06	-114.44	0.000
MARST				
1	.0314585	.0012047	26.11	0.000
2	-.0075814	.0015425	-4.91	0.000
RACE				
1	-.0766378	.0016509	-46.42	0.000
2	-.0765951	.0045309	-16.91	0.000
3	.0083189	.0018737	4.44	0.000
4	-.078021	.0020986	-37.18	0.000
5	-.0419154	.0015184	-27.60	0.000
EDUC				
1	.1212234	.0030314	39.99	0.000
2	.2675495	.003006	89.01	0.000
3	.4302998	.0031198	137.92	0.000
IND				

1	.0957857	.0101604	9.43	0.000
2	-.050431	.0037778	-13.35	0.000
3	-.0315442	.0021444	-14.71	0.000
4	-.1195097	.0020285	-58.91	0.000
5	-.0628449	.0029174	-21.54	0.000
6	.0324396	.0037813	8.58	0.000
7	.0263704	.001982	13.31	0.000
8	-.2543187	.0068146	-37.32	0.000
9	-.0948788	.0015496	-61.23	0.000
10	-.2021204	.002565	-78.80	0.000
11	-.1517648	.0029442	-51.55	0.000
12	-.0171572	.0020843	-8.23	0.000
UHRSWORK	.0059976	.0000842	71.26	0.000
FEMALE	-.1377257	.0038732	-35.56	0.000
_CONS	3.450114	.0079067	436.35	0.000

Number of Observations	=	532,679
F(31, 532,647)	=	5957.74
Prob > F	=	0.0000
R-Squared	=	0.2915

US 2011 FEMALE Results

LNINCWAGE	COEF.	STD. ERR.	T	P>|T|
FEMALE	-.084025	.0043981	-19.10	0.000
REGION				
1	-.0463073	.0016178	-28.62	0.000
2	-.0465556	.0015012	-31.01	0.000
3	-.0179468	.0017842	-10.06	0.000
1.METRO	-.1029691	.0012096	-85.13	0.000
NCHLT5	.0258595	.0015114	17.11	0.000
AGE	.0471577	.0003973	118.70	0.000
AGESQ	-.0004737	4.48e-06	-105.64	0.000
MARST				
1	.0273493	.0015721	17.40	0.000
2	-.0022836	.0018791	-1.22	0.224
RACE				
1	-.0610124	.0018211	-33.50	0.000
2	-.059261	.0059325	-9.99	0.000
3	-.0080838	.0027553	-2.93	0.003
4	-.0681057	.0034304	-19.85	0.000
5	-.0526066	.0045624	-11.53	0.000

EDUC				
1	.1554345	.0030881	50.33	0.000
2	.300315	.0030749	97.67	0.000
3	.4765305	.0033118	143.89	0.000
IND				
1	.1035538	.0120224	8.61	0.000
2	-.0203215	.0058137	-3.50	0.000
3	-.0064536	.002899	-2.23	0.026
4	-.0688802	.0027199	-25.32	0.000
5	.0290924	.0037187	7.82	0.000
6	.0427137	.0044434	9.61	0.000
7	.0539328	.0027011	19.97	0.000
8	-.2210123	.0091952	-24.04	0.000
9	-.0319402	.0022366	-14.28	0.000
10	-.1679643	.0033323	-50.40	0.000
11	-.1229387	.0039158	-31.40	0.000
12	.0591149	.0028185	20.97	0.000
UHRSWORK	.0061645	.0001087	56.73	0.000
_CONS	3.020911	.0102496	294.73	0.000

Number of Observations	=	329,415
F(31, 329383)	=	3969.49
Prob > F	=	0.0000
R-Squared	=	0.3163

Canada 2011 FEMALE Results

LNINCEARN	COEF.	STD. ERR.	T	P>|T|
FEMALE	-.0743464	.006758	-11.00	0.000
NCHLT5	-.0689382	.0031694	-21.75	0.000
AGE	.0427293	.000707	60.44	0.000
AGESQ	-.0004302	8.40e-06	-51.24	0.000
MARST				
1	.0177823	.0028125	6.32	0.000
2	.0096574	.0038536	2.51	0.012
1.INDIG	-.0427927	.0065603	-6.52	0.000
EDUC				
1	.0698995	.0044785	15.61	0.000
2	.1962147	.0044583	44.01	0.000

3	.3124959	.0054788	57.04	0.000
IND				
1	.2095632	.0137561	15.23	0.000
2	.0100538	.0091273	1.10	0.271
3	-.0076876	.0054185	-1.42	0.156
4	-.0353899	.0045409	-7.79	0.000
5	.0617411	.0064897	9.51	0.000
6	.0787633	.0071248	11.05	0.000
7	.0906299	.0046635	19.43	0.000
8	-.2566998	.0129498	-19.82	0.000
9	.0409856	.0040386	10.15	0.000
10	-.1520997	.0055616	-27.35	0.000
11	-.1174398	.0063073	-18.62	0.000
12	.1224772	.0045129	27.14	0.000
HRSACTUAL1	.0002706	.0001863	1.45	0.146
_CONS	3.416649	.0170329	200.59	0.000

Number of Observations	=	133,697
F(23, 133673)	=	1196.20
Prob > F	=	0.0000
R-Squared	=	0.1727