An Econometric Analysis on the Gender Wage Gap

Judy Chen

Stat 2123

Professor Benjamin Williams

May 14, 2019

I. Introduction

Since the Equal Pay Act passed 1963, it has been illegal for women to be paid less than their male counterparts. However, as of 2019, women make $0.79 for every dollar earned by a man (Chamberlain, 2019). This disparity is not explained by discriminatory wages that put women at disadvantage, but rather the difference in career choices between men and women. Many institutional barriers, unconscious bias being one of the most prominent, exist to prevent women from entering or advancing in many professions dominated by men. The "leaky pipeline" phenomenon was documented by a study conducted by McKinsey & Company (McKinsey & Co., 2015) in 2015, which demonstrate that women make up 45% of entry-level professionals but only contributes 17% to C-suite level positions.

This paper examines the relationship between gender and income during the time period 1980 to 2000. This paper aims to answer these questions: does gender have a significant effect on wage at the state level? In other words, after controlling for as much selection bias as possible, do states with higher proportion of women in the employed population have lower income per-capita? Do women tend to end up in lower-paying jobs? It is important to understand the effect gender has a on individuals' level of compensation. If the effect of gender on wage is significant, the U.S. government should be incentivized to implement policies to encourage diversity in the workplace and help women break the glass ceiling. This also means it is more important for women to make sound personal finance decisions as their life-time earnings are likely less than that of their male counterparts.

II. Econometric Model

A. Assumptions for Multiple Regressions

The analysis uses a regression model where the natural log of income per-capita of each U.S. state (*log_income_percap*) is the dependent variable, and the percentage of females in the employed population (*pct_female_employed)* is the regressor of interest. There are three assumptions that multiple regression models should satisfy. First, the zero conditional means assumption states that the covariance between the error term and the regressor of interest should be zero, meaning the average effect of other determinants on income per-capita does not depend on the regressor of interest. Second, the sample must be independently and identically distributed. Third, there is no perfect multicollinearity between independent variables, meaning no one independent variable is a perfect linear function of any other independent variables. Finally, large outliers are unlikely.

The zero conditional means assumption is likely violated if the regression model only contains *log_income_percap* on *pct_female_employed*. Many factors that have a significant effect on per-capita income are correlated with the percent of females in the employed population. Therefore, control variables are needed to minimize omitted variable bias (OVB).

B. The Baseline Model

There are many reasons why states where more women are employed may have higher per-capita income, contrary to what is expected in the presence of the gender wage gap. The two most evident omitted variables are education level and year of survey. States where more women are employed may be ones where women are more educated. As society becomes more progressive overtime, surveys taken in a later year may record more women contribute to the employed population. Excluding these two variables cause OVB as they are correlated with the regressor and likely have significant effect on the dependent variable as well. States where women have higher education level likely have higher income as more people have higher-paying jobs. With inflation, nominal income rises every year. The base line regression model regresses *log_income_percap* on *pct_female_employed* while controlling for year of the survey with dummy variables *year_1980* and *year_1990*, and *percent_female_college_grad,* which indicates the percentage of women above age 25 whose education level is bachelor's degree or above. The coefficient on *pct_female_employed* in the baseline model is still potentially biased as other necessary control variables may be omitted.

C. Models with More Control Variables

Potentially omitted variables that were not accounted for in the baseline regression include a state's count of crimes for both violent and property crimes relative to a state's population (*total_crime_rate*), unemployment rate (*unemployment_rate*), percentage of population between ages 25 to 64 (*pct_working_ages*), percentage of African American in the population (*pct_black*) and percentage of population that lives in urban areas (*popmetro_pct*). Omitting the crime rate variable present OVB. A study has found that men are more likely to commit crimes than woman, especially violent crimes (Choy, 2017). This means states with higher crime rates could have lower female population than male, thus lower female contribution to the employed contribution. States with high crime rates also tend to be poorer and thus have lower per-capita income. Excluding unemployment rate as a variable may lead to OVB as well. Due to existing sexism in many professions, it is likely easier for men to find work than women, especially in labor-intensive or male-dominant jobs where many women are rejected due to unconscious bias. States with high unemployment rate could have low percent of women in the employed workforce. High unemployment rate also signals poverty, which would explain low per-capita income for a state. States with higher proportion of women in the workforce could also be younger and contain more people within working ages, therefore higher income per-capita than states with high proportion of retirees and children in the population. Due to existing discrimination related to race and intersectionality of gender and race when it comes to unconscious bias, states where fewer women are employed could also have high African American population, which potentially means lower per-capita income. States that has more women in the workforce could also be more progressive and contain high population living in urban areas. States with large metropolitan cities could have higher per-capita income because high-paying jobs tend to be concentrated in these cities.

III. Description of Data and Results

A. Description of Data

The data used in this paper is from state-level censuses on 50 states and District of Columbia in years 1980, 1990 and 2000. Table 1 contains summary statistics of the variables used in this econometric analysis rounded to 4 decimals.

Table 1: Summary Statistics of Variables

Variable name	Description	Type of Variable	Sample Mean	Sample Standard Deviation	Minimum Value	Maximum Value
Log_percap _income	Natural Log of income per-capita in USD	Dependent Variable	9.4370	0.4706	8.5531	10.267
Pct_female _employed	Percentage of females in the employed population	Regressor of Interest	45.1961	2.4988	37.5976	51.9760
Year_1980	Dummy variable that indicates if census was from 1980	Control Variable	0.3333	0.4730	0	1

Year_1990	Dummy variable that indicates if census was from 1990	Control Variable	0.3333	0.4730	0	1
Pct_female _college_grad	Percentage of women above age 25 whose education level is bachelor's degree or higher	Control Variable	17.8185	5.4484	8.7686	36.8177
Total_crime_rate	Total crime counts divided by total state population, in percentage points	Control Variable	4.9775	1.5163	2.2881	10.7743
Pct_working_age	Percentage of population between ages 25 to 64	Control Variable	49.9018	3.0279	40.7524	54.9824
Pct_black	Percentage of population that is African American	Control Variable	10.62	12.0549	0.2209	70.2408
Popmetro_pct	Percentage of population that lives in urban areas	Control Variable	54.1175	21.2571	14.9628	100
Unemployment _rate	Percentage of unemployed people in the labor force	Control Variable	6.0739	1.4295	3.5175	10.9512

B. Results of Regression Models

Table 2 provides the coefficient estimate on *pct_female_employed* of the baseline model specified in Part II Section B.

Table 2: Effect of Percentage of Women in Employed Workforce in Baseline Model

Variable	Coefficient Estimate	Robust Standard Error	t-statistic	Is the estimate statistically significant at the 5% level?
pct_female_employed	0.0009	0.0055	0.16	No
pct_female_college_grad	0.0288	0.0027	10.82	Yes
year_1980	-0.8	0.0308	-26.01	Yes
year_1990	-0.2739	0.0235	-11.67	Yes

The coefficient on *female_employed* is 0.0009. Holding the control variables constant, for every additional percentage of women in the employed population, the state's per-capita income increases 0.09%. The percentage of women in employed population has a positive, but statistically insignificant on per-capita income. Next, Table 3 presents Comprehensive Model I, which includes variables that were potentially omitted in the baseline regression.

Table 3: Result of Comprehensive Model I

Variable	Coefficient Estimate	Robust Standard Error	t-statistic	Is the estimate statistically significant at the 5% level?
pct_female_employed	-0.009	0.0055	-1.65	No
pct_female_college_grad	0.0155	0.002	7.79	Yes
year_1980	-0.7637	0.0272	-28.07	Yes
year_1990	-0.2816	0.0153	-18.46	Yes
Total_crime_rate	-0.0078	0.0046	-1.68	No
Pct_working_age	0.033	0.0025	13.30	Yes
Pct_black	-0.0009	0.0005	-1.74	No
popmetro_pct	0.0031	0.0003	8.82	Yes
Unemployment_rate	-0.0084	0.0057	-1.50	No

The coefficient on *female_employed* is -0.009. This number can be interpreted as a 0.9% decrease in state per-capita income for every additional 1% of women in the employed population. This estimate is not statistically significant at the 5% level but is statistically significant at the 10% level. This means percentage of women in employed population does not have a statistically significant effect on per-capita income. Since the coefficient on *total_crime_rate*, *pct_black* and *unemployment_rate* is statistically insignificant as well, these variables are removed from the regression model. Table 4 presents coefficient results in

Comprehensive Model II, where *unemployment_rate, pct_black* and *unemployment_rate* are removed.

Table 4: Result of Comprehensive Model II

Variable	Coefficient	Robust Standard Error	t-statistic	Is the estimate statistically significant at the 5% level?
pct_female_employed	-0.0133	0.0044	-3.00	Yes
pct_female_college_grad	0.0167	0.0019	8.75	Yes
year_1980	-0.7986	0.0237	-33.69	Yes
year_1990	-0.2972	0.0151	-19.47	Yes
Pct_working_age	0.0315	0.0025	12.35	Yes
popmetro_pct	0.0026	0.0003	9.87	Yes

The coefficient on *pct_female_employed* decreased dramatically from -0.009 to -0.0133. As per Comprehensive Model II, every additional 1% of women in the employed population results in a 1.33% decrease in a state's per-capita income. We reach a different conclusion than Comprehensive Model I. There is a statistically significant difference in per-capita income for states where women make up more of the employed population. Thus, controlling for year of survey, women's education level, percentage of working age population and percentage of urban population, percentage of women in the employed population has a statistically significant effect on state's per-capita income.

IV. Conclusion

Using data from the state-level census, this study is able to conclude, at a 95% confidence level, that the gender wage gap is real. The result indicates that percentage of women in the employed population has a statistically significant effect on per-capita income of a state. However, there is a chance that the effect of 1.33% decrease for every additional 1% of women in the employed population is not fully representative of reality. The real gender wage gap is likely larger, but the effect may not have been captured by Comprehensive Model II because it is still flawed after adding the control variables. To better understand the difference in income between gender, a better dependent variable would be the difference in states' per-capita income between males and females. The ideal dependent variable would be the state-level average income by gender for each profession. Including profession dummy variables as control variables minimizes OVB, as type of profession has a significant effect on income and is correlated with gender. Women are more likely to end up in certain professions than men due to institutional barriers.

The result of the Comprehensive Model II also indicates that women's level of education, measured by the percentage of women above age 25 whose education level is bachelor's degree or above, has a statistically significant effect on a state's per-capita income. Controlling for percentage of women in the employed workforce, year of survey, percentage of working-age and urban population, for every 1% increase in the proportion of women above 25 whose education level is bachelor's degree or above, state's per-capita income increases by 1.67%. This effect is statistically significant. Having access to per-capita income for both male and female would further help examine the effect of education on gender wage gap. I would first use a t-test to determine if the difference in per-capita income between the two genders is statistically significant. If yes, I would regress the difference in per-capita income between gender on the difference in percentage of population age 25 and above who have completed a bachelor's degree or higher between men and women. This will help me understand if the difference in education level between men and women has a statistically significant effect on the difference in per-capita income between genders. Therefore, the result will shed some light on whether the gender wage gap gets narrower as men and women are closer in average education level. States where the education gap between men and woman are wide are more likely to have higher poverty rates. In the new regression, I would control for crime rate, unemployment rate, percent of African American population and percent of population that live in urban areas. These control variables are correlated with a state's level of poverty and likely have a significant effect on gender wage gap. The year dummy variables should also be included as control variables because poverty in a state can change significantly in an increment of ten years. The gender wage gap can also widen overtime. With more comprehensive data that segregates male and female per-capita income for each state, I am hopeful to develop a regression model that can provide more details on the magnitude of the gender wage gap, as well as examine more nuanced thesis within this topic, such as the effect of education on closing the gap.

Works Cited:

Chamberlain, Andrew, et al. “Progress on the Gender Pay Gap: 2019.” *Glassdoor*, 2019, www.glassdoor.com/research/app/uploads/sites/2/2019/02/Gender-Pay-Gap-2019-Research-Report.pdf.

“Women in the Workplace.” *McKinsey & Company*, www.mckinsey.com/business-functions/organization/our-insights/women-in-the-workplace.

Choy, Olivia; et al. “Explaining the Gender Gap in Crime: The Role of Heart Rate,” Criminology, 2017. doi: 10.1111/1745-9125.12138.