Preliminary Econometric Estimates

I. Data Selection

All data is collected in the American Community Survey (ACS) facilitated and provided by the US Census Bureau. Our data is specifically from the 2018 IPUMS dataset. We first began by removing NA values from the income variable (PINCP) as it is our dependent variable. We also removed NA values from the commute time and incomes below $0 per year. The latter allows us to take the natural log of income in our model, a common practice to present the model in terms of percentage points in the dependent variable rather than the absolute values. The original dataset had 76,225 rows. Our cleaned dataset has 32,888 rows. This did significantly reduce the size of our dataset but we still have enough observations to make statistical inferences.

We then conducted an exploration into population subsets to see if they should be removed. We considered removing rows with an income value under $6240 as that is the annual income of a person making the minimum wage for Washington ($12) and working 10 hours per week. After analyzing that subset we found that that population made up 5.5% of the dataset. The average age of the cleaned dataset was 42.5 years, the average of the subset was 28.9 years, a difference of 13.6 years. The average commute time for the cleaned dataset is 28.4, the average for the subset is 21.5, a difference of 6.9 minutes.

There were no unknown values for the age variable (AGEP) in the dataset so no null values needed to be removed. There was also no need to filter out individuals under the legal working age (16) as they were already captured as part of the NA population for the income variable. One population we considered removing was those over the age of 65. The subset makes up 4.6% of the dataset. The average income of the cleaned population is $66,280, the average for those over 65 is $80,988, a difference of 14,708.

We also considered removing those under 18. This subset makes up only .97% of the dataset. The average income of the subset was only $5,126, $61,154 below the population average.

We would have removed the population categorized as 8 or 9 for the type of employer field (COW) but there were no rows in our cleaned dataset with either of those values.

Variables:

Female	Respondents who selected Female gender
PrivateF	Females working in private sector
PublicF	Females working in public sector
NotForProfitF	Females working in not for profit sector
JWMNP	Travel time to work in minutes
AGEP	Age in years
Nonwhite	Not white by race
HSgrad	Individuals with H.S. Diploma
BAdegree	Individuals with Bachelor's Degree
AdvanceDegree	Individuals with a Master's Degree or Higher
ENGnonProficient	Individuals who selected they did not speak English well or at all
CommuterF	Female whose commute time to work is greater than 0
PERNP	Total Person's Earnings in USD

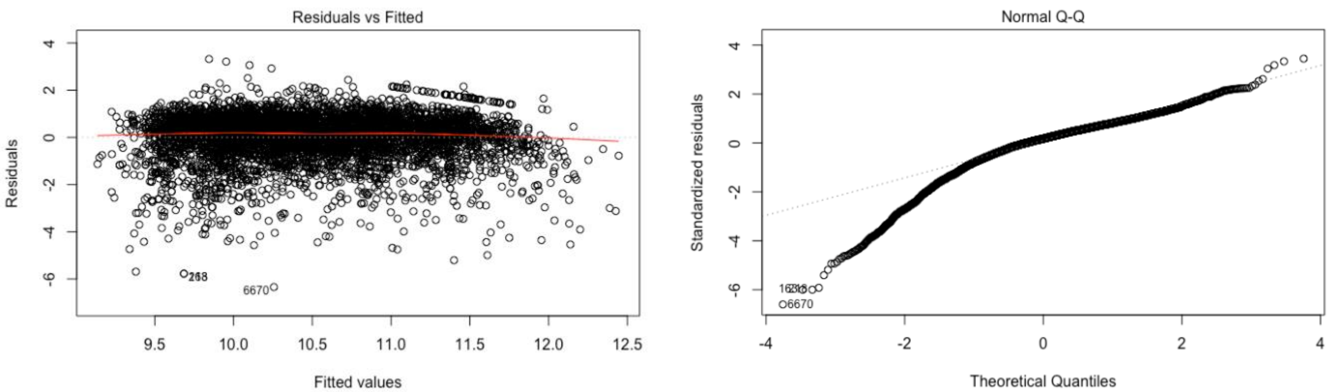
II. Research Question 1: "Can women expect a smaller magnitude pay gap in the public sector in Washington State?"

$$Income = \beta 0 + \beta 1(Sex) + \beta 2(Age) + \beta 3(Race) + \beta 4(Education) + \beta 5(English) + \varepsilon$$

	Dependent variable:	
	log(PERNP)	
	(1)	(2)
Female	-0.4480***	
	(0.0252)	
PrivateF		-0.3603***
		(0.0281)
PublicF		-0.5068***
		(0.0538)
NotForProfitF		-0.3953***
		(0.0610)
JWMNP	0.0040***	0.0041***
	(0.0006)	(0.0006)
AGEP	0.0192***	0.0187***
	(0.0010)	(0.0010)
Nonwhite	0.0640**	0.0620**
	(0.0259)	(0.0261)
HSgrad	-0.0573	-0.0553
	(0.0362)	(0.0365)
BAdegree	0.6274***	0.6325***
	(0.0344)	(0.0346)
AdvanceDegree	1.0870***	1.0976***
	(0.0360)	(0.0363)
ENGnonProficient	-0.2765***	-0.2839***
	(0.0371)	(0.0375)
Constant	9.4714***	9.4414***
	(0.0486)	(0.0491)
Observations	5,820	5,820
R2	0.2890	0.2793
Adjusted R2	0.2880	0.2781
Residual Std. Error	0.9566 (df = 5811)	0.9632 (df = 5809)
F Statistic	295.1797*** (df = 8; 5811)	225.1278*** (df = 10; 5809)
Note:		*p<0.1; **p<0.05; ***p<0.01

In our first model, we have a basic income model without evaluating the gender pay gap by sector. We can see that all of our independent variables are highly significant with the exception of High School graduates (which appears to have no significant impact on income when compared to non-high school graduates 18 years or older).Most of our coefficients have the expected signs with commute distance, age and education producing a positive sign. While non-English proficient and being female produced negative signs. However, what is interesting is that Nonwhite is producing a positive sign. This would indicate that in the state of Washington non-white workers can expect higher pay than their non-white counterparts (holding all else equal of course).

In our second model, we created categorical variables for female workers in the private, public and not for profit sectors. The coefficients on our control variables remained relatively the same in sign and magnitude. However, our variables of interest produced unexpected results as well. The gender pay gap in Washington state is highest in the public sector with not for profit and private the next highest respectively. We can interpret this as the following: A female worker in the public sector can expect to earn roughly 50% less than her male counterpart in Washington state. While a female worker in the private sector can expect to earn 36% less than her male counterpart in Washington state.



III. Research Question 2: “What is the relationship of commute time and income with respect to gender in the state of Washington?”

$$Income = \beta 0 + \beta 1(Commute) + \beta 2(Age) + \beta 3(Race) + \beta 4(Sex) + \beta 5(Education) + \beta 6(English) + \varepsilon$$

	Dependent variable:	
	log(PERNP)	
	(1)	(2)
JWMNP	0.0040***	
	(0.0006)	
CommuterF		-0.4588***
		(0.0253)
AGEP	0.0192***	0.0196***
	(0.0010)	(0.0010)
Nonwhite	0.0626**	0.0638**
	(0.0259)	(0.0260)
Female	-0.4474***	
	(0.0252)	
BAdegree	0.6432***	0.6609***
	(0.0329)	(0.0329)
AdvanceDegree	1.1029***	1.1188***
	(0.0345)	(0.0346)
ENGnonProficient	-0.2746***	-0.2771***
	(0.0371)	(0.0373)
Constant	9.4545***	9.5539***
	(0.0475)	(0.0455)
Observations	5,820	5,820
R2	0.2886	0.2825
Adjusted R2	0.2878	0.2817
Residual Std. Error	0.9567 (df = 5812)	0.9608 (df = 5813)
F Statistic	336.9031*** (df = 7; 5812)	381.4321*** (df = 6; 5813)
Note:		*p<0.1; **p<0.05; ***p<0.01

Here we turn to our second model as our first has already been discussed in the previous research question. In our second model, we have pulled out commuters that our female to compare them with commuters in general. As we can see for commuters in general in Washington state, there is a positive relationship with income and commute distance. However, for female commuters there is a statistically significant negative coefficient with a much larger magnitude than what is illustrated in model 1. Admittingly, the causal explanation of this is difficult. Our results indicate that being a female commuter will reduce income by 45% compared to male counterparts. However, we did not set bounds yet on our commuter variable. In our next iteration of this model, we will likely create either out right categories (short vs long commutes) or, at the very least, upper and lower bounds.