Problem Set 1 ECON8825

Federico Rodari

October 1, 2022

1. The following table provides standard summary statistics about the variables in the data. The main

	Male	Female		Male	Female		Male	Female		Male	Female
Mean Lnw	2.909	2.648	**Mean Hsd911**	0.028	0.015	**Mean We**	0.208	0.188	**Mean Married**	0.742	0.657
Std Lnw	0.668	0.625	**Std Hsd911**	0.164	0.121	**Std We**	0.406	0.391	**Std Married**	0.438	0.475
Min Lnw	-7.470	-7.247	**Min Hsd911**	0.000	0.000	**Min We**	0	0	**Min Married**	0	0
Max Lnw	5.971	5.876	**Max Hsd911**	1.000	1.000	**Max We**	1	1	**Max Married**	1	1
Skewness Lnw	-0.882	-1.064	**Skewness Hsd911**	5.751	8.045	**Skewness We**	1.439	1.593	**Skewness Married**	-1.105	-0.664
Mean Widowed	0.003	0.014	**Mean Hsg**	0.270	0.217	**Mean Exp1**	18.910	18.553	**Mean Ne**	0.227	0.225
Std Widowed	0.056	0.119	**Std Hsg**	0.444	0.412	**Std Exp1**	8.608	8.971	**Std Ne**	0.419	0.417
Min Widowed	0	0	**Min Hsg**	0	0	**Min Exp1**	0	0	**Min Ne**	0	0
Max Widowed	1	1	**Max Hsg**	1	1	**Max Exp1**	43.500	42.500	**Max Ne**	1	1
Skewness Widowed	17.661	8.161	**Skewness Hsg**	1.038	1.370	**Skewness Exp1**	-0.053	-0.087	**Skewness Ne**	1.305	1.320
Mean Divorced	0.088	0.147	**Mean Cg**	0.275	0.294	**Mean Exp2**	4.317	4.247	**Mean Sc**	0.276	0.302
Std Divorced	0.283	0.354	**Std Cg**	0.447	0.456	**Std Exp2**	3.301	3.349	**Std Sc**	0.447	0.459
Min Divorced	0	0	**Min Cg**	0	0	**Min Exp2**	0	0	**Min Sc**	0	0
Max Divorced	1	1	**Max Cg**	1	1	**Max Exp2**	18.923	18.063	**Max Sc**	1	1
Skewness Divorced	2.911	1.990	**Skewness Cg**	1.005	0.904	**Skewness Exp2**	0.600	0.560	**Skewness Sc**	1.003	0.861
Mean Separated	0.013	0.021	**Mean Ad**	0.146	0.169	**Mean Exp3**	10.931	10.802	**GroupCount**	16690	12527
Std Separated	0.115	0.143	**Std Ad**	0.353	0.375	**Std Exp3**	11.119	11.126			
Min Separated	0	0	**Min Ad**	0	0	**Min Exp3**	0	0			
Max Separated	1	1	**Max Ad**	1	1	**Max Exp3**	82.313	76.766			
Skewness Separated	8.457	6.710	**Skewness Ad**	2.006	1.767	**Skewness Exp3**	1.112	1.064			
Mean Nevermarried	0.154	0.160	**Mean Mw**	0.285	0.301	**Mean Exp4**	29.529	29.249			
Std Nevermarried	0.361	0.367	**Std Mw**	0.451	0.459	**Std Exp4**	36.701	36.395			
Min Nevermarried	0	0	**Min Mw**	0	0	**Min Exp4**	0	0			
Max Nevermarried	1	1	**Max Mw**	1	1	**Max Exp4**	358.061	326.254			
Skewness Nevermarried	1.921	1.855	**Skewness Mw**	0.955	0.868	**Skewness Exp4**	1.593	1.534			
Mean Hsd08	0.005	0.003	**Mean So**	0.281	0.286	**Mean Weight**	1508.918	1520.404			
Std Hsd08	0.072	0.050	**Std So**	0.449	0.452	**Std Weight**	1006.623	1014.047			
Min Hsd08	0	0	**Min So**	0	0	**Min Weight**	106.790	125.180			
Max Hsd08	1	1	**Max So**	1	1	**Max Weight**	6444.150	5939.720			
Skewness Hsd08	13.663	19.710	**Skewness So**	0.977	0.947	**Skewness Weight**	0.734	0.629			

Table 1: Summary Statistics

takeaways are:

- Men have higher salaries on average.
- There are not substantial differences across the whole spectrum of educational attainment and experience.
- Males are relatively more represented in the sample.

2. The moments conditions are

$$
\begin{aligned}
F_j(y_{ij} \mid x_{ij}) &= \Phi(x_{ij}\beta_{jk}) \\
\mathbb{E}\left[1(y_{ij} \leq \tau) - \Phi(x_{ij}'\beta_{jk}) \mid x_{ij}\right] &= 0 \\
\mathbb{E}\left[1(y_{ij} \leq \tau) - \Phi(x_{ij}'\beta_{jk}) m(x_{ij})\right] &= 0 \\
\mathbb{E}\left[\frac{\phi(x_{ij}'\beta_{jk})}{\Phi(x_{ij}'\beta_{jk})\left[1-\Phi(x_{ij}'\beta_{jk})\right]}\left(1(y_{ij} \leq \tau) - \Phi(x_{ij}'\beta_{jk})\right)x_{ij}\right] &= 0_{K\times 1} \\
\mathbb{E}\left[s(y_{ij}, x_{ij}) \mid \beta_{jk}\right] &= 0_{K\times 1}
\end{aligned}
$$

The unconditional moment condition is equivalent to the identification condition for MLE. It is computationally convenient to use MLE to estimate the Probit model.
Let K be the number of regressors, J the number of genders and T the number of quantiles at which the model is estimated. We have $K \times J \times T$ moments conditions to estimate $K \times J \times T$ parameters, therefore the model is exactly identified. On top of that, given the equivalence between GMM and MLE and given that MLE attains the Cramer-Rao lower bound, the moments conditions are efficient.

3. I report the difference of each coefficient x quantile between males and females

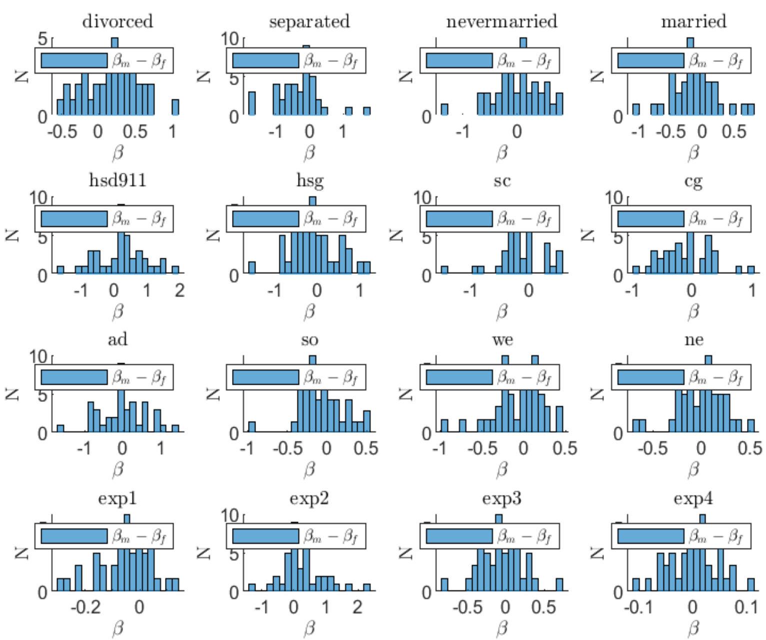

Figure 1: Distribution of $\Delta(\beta)$ Between Males and Females

By looking at the range of each coefficients, it is possible to see that there is substantial variation for each variable. At a first glance, there is a relatively symmetric distribution of the difference for each predictor.

4. We can use the following approximation

$$\begin{aligned}F(y)_{m|w} &= \int_{x_w} F(y \mid x_w) dF_w(x_w) \\ &= \int_{x_w} \Phi(x'_{iw}\beta_{my}) dF_w(x_w) \\ &= \int_{x_w} \Phi(x'_{iw}\beta_{my}) f_w(x_w) dx_w \\ &\approx \frac{1}{n_w} \sum_{x_w} \Phi(x'_{iw}\beta_{my})\end{aligned}$$

and similarly for $m|m$ and $w|w$, the unconditional

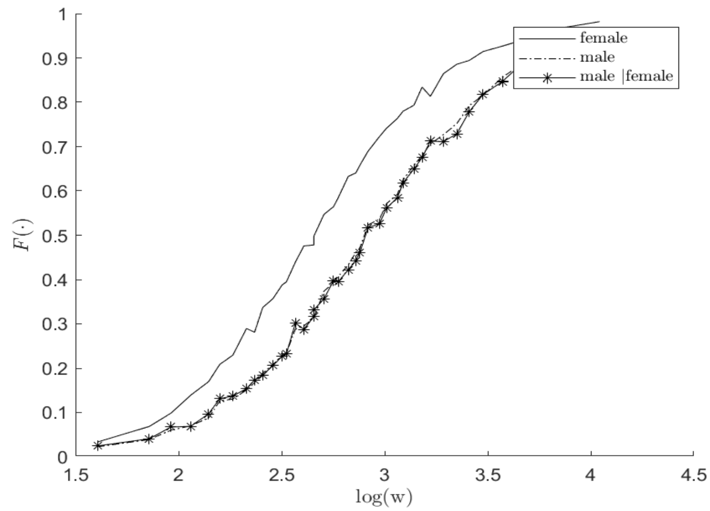

Figure 2: Counterfactual Distributions

5. The estimated price and composition effects are

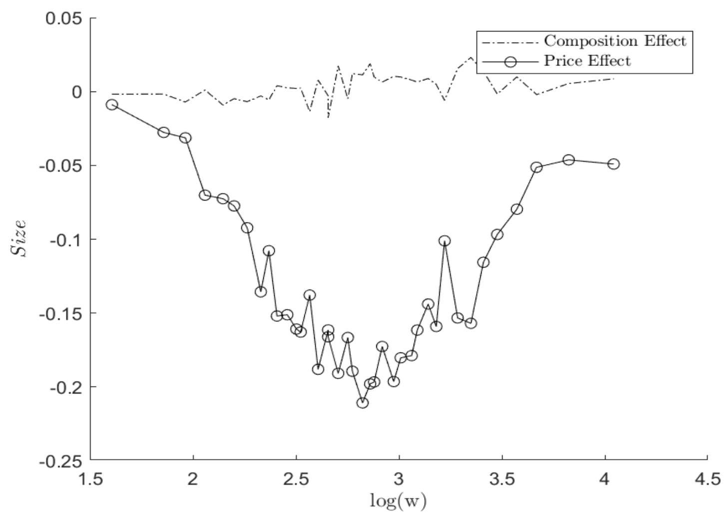

Figure 3: Price and Composition Effects

We can see that the main driver is the price effect, namely the coefficients drive most of the variation between male and female wages.

6. I chose $B = 100$ to speed-up on computational time.

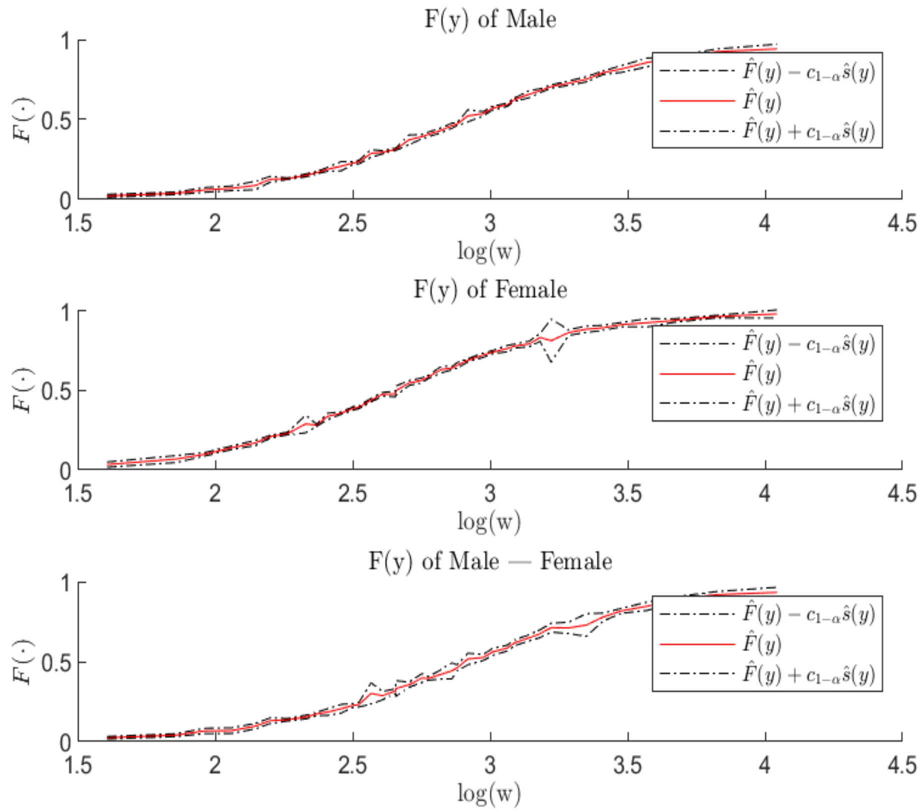

Figure 4: Bootstrap CDFs

7. The confidence bands around the price and the composition effects look like the following

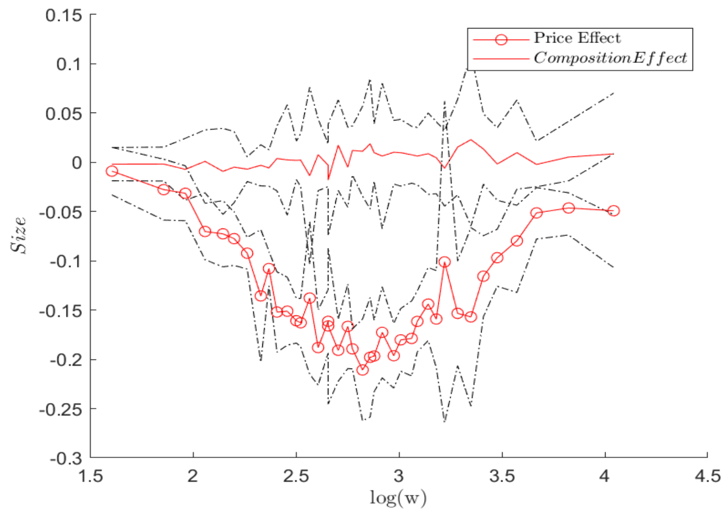

Figure 5: Bootstrap CDFs

Given the small size of bootstrap replications, we might still observe a lot of noise in the confidence bands. From these results we can conclude that for low and high wages we cannot disentangle the price and the composition effect and moreover we cannot infer there is actually a statistical difference between male and females. The decomposition looks robust for wages in the middle of the distribution.