

Econometrics Methods Project: Figures

Mason Veileux

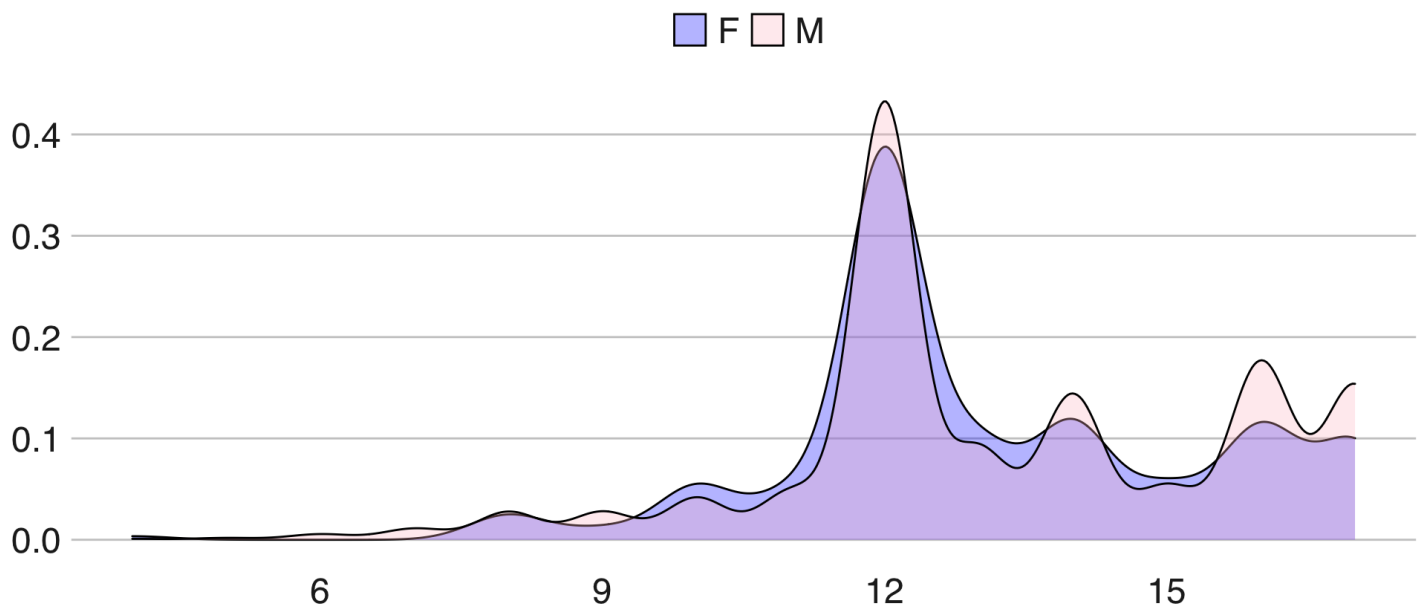
20/1/2022

This file provides summary statistics and figures related to the data and regression analysis for the Econometrics Methods project.

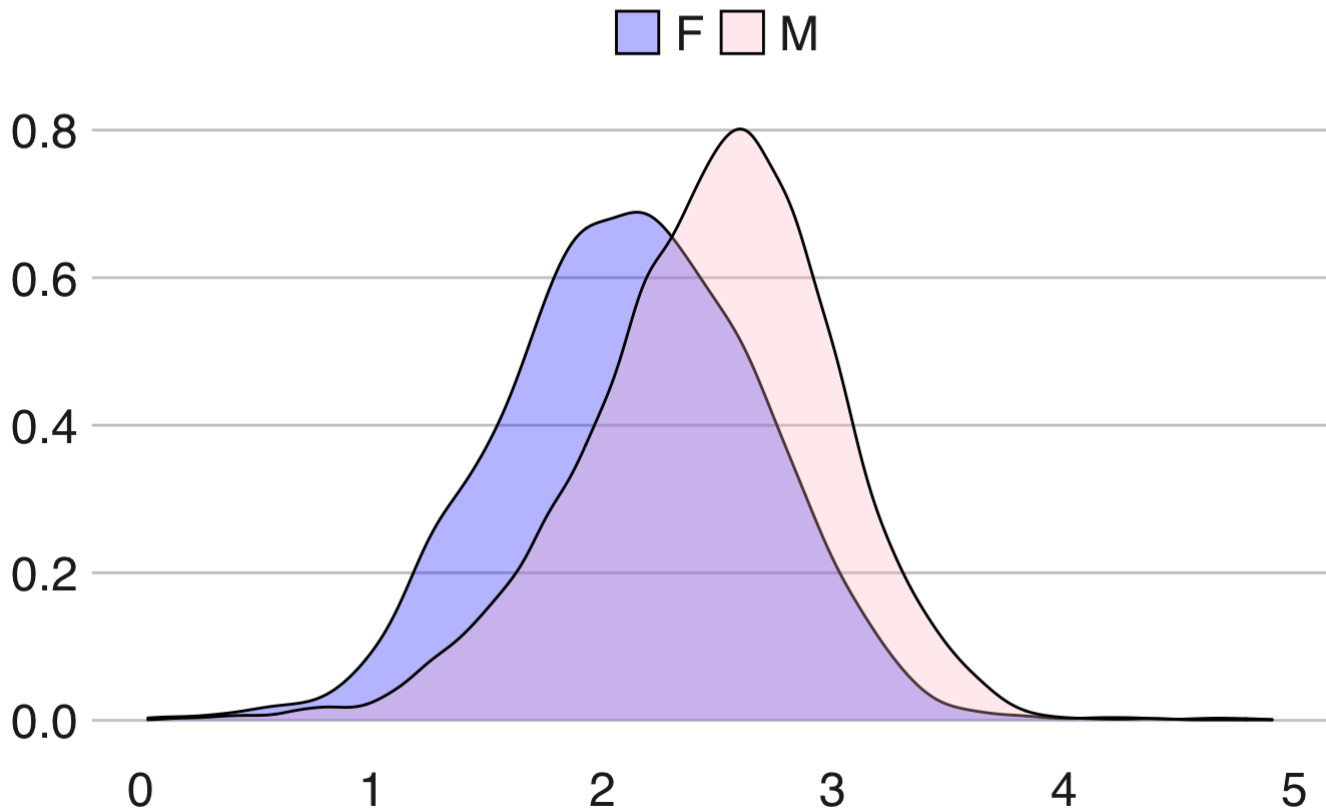
Summary Statistics

These are distributions broken down by gender

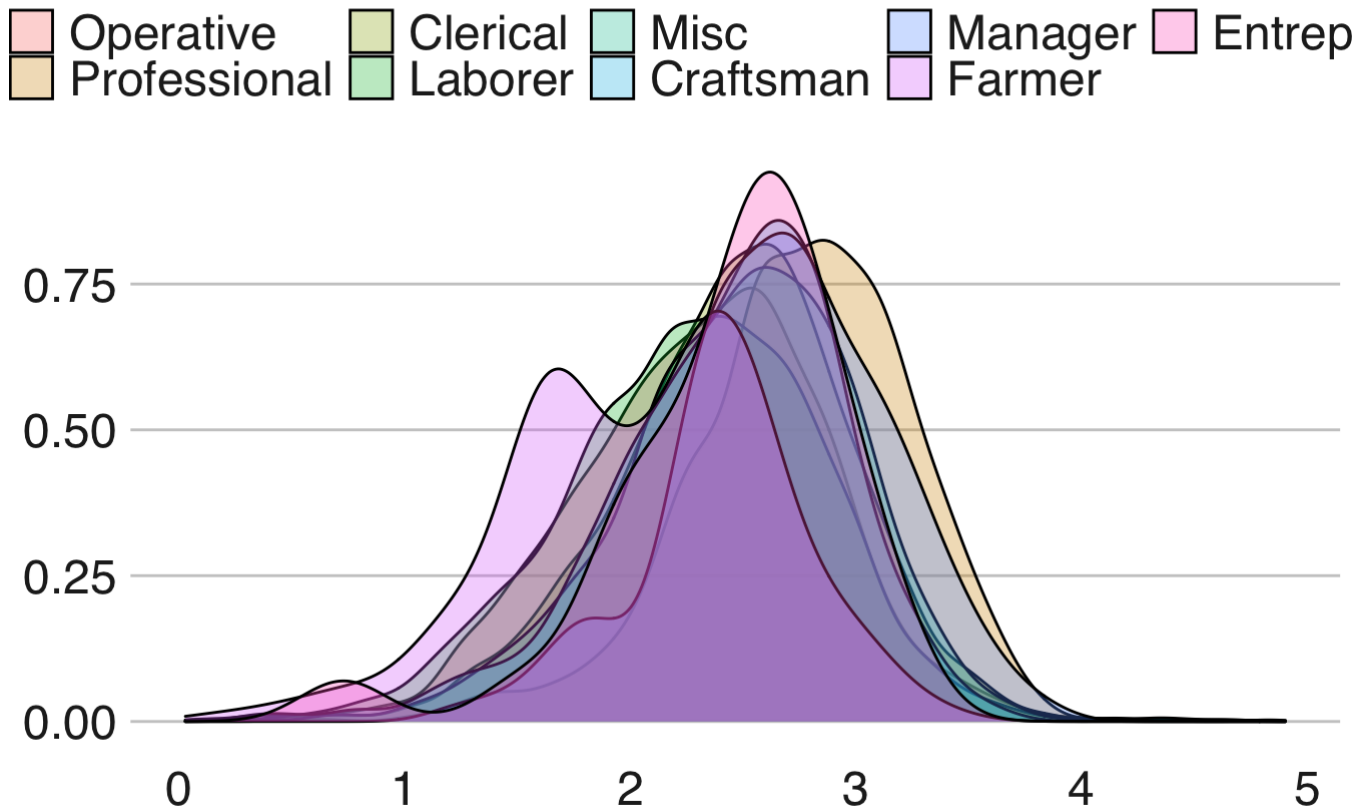
Density of Education Completed in 1985 by Gender



Density of Log Wage by Gender

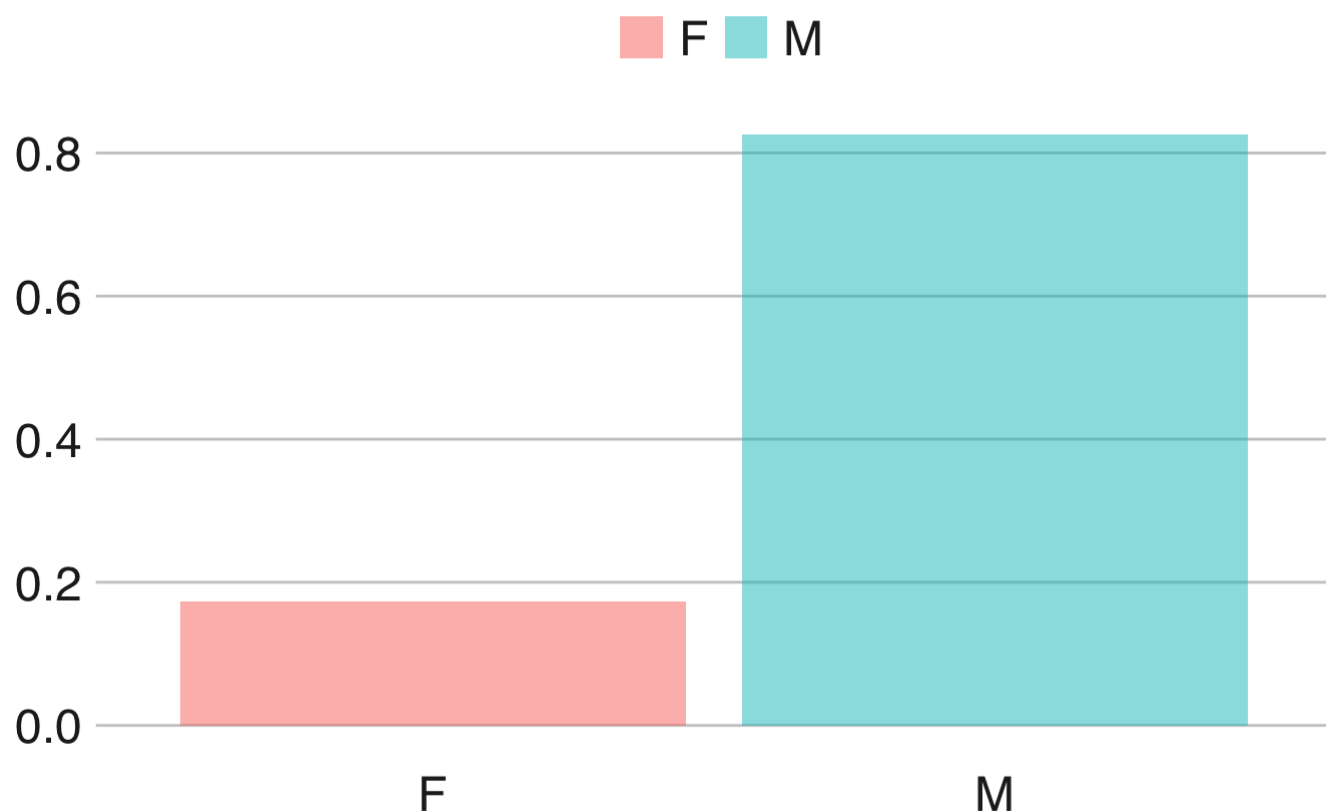


Density of log(Wage) Occupation



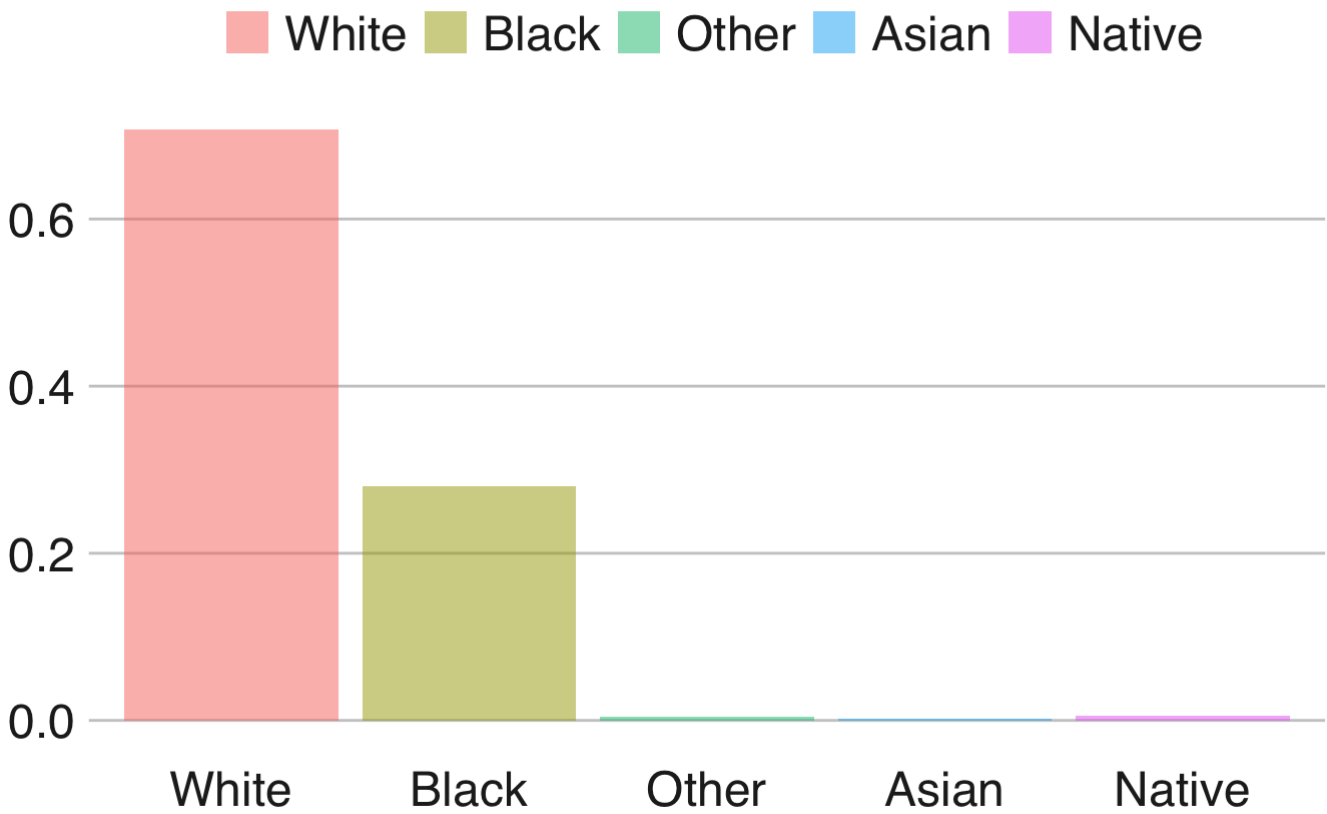
Clearly, men are over-represented in this sample

Sex Bar Plot



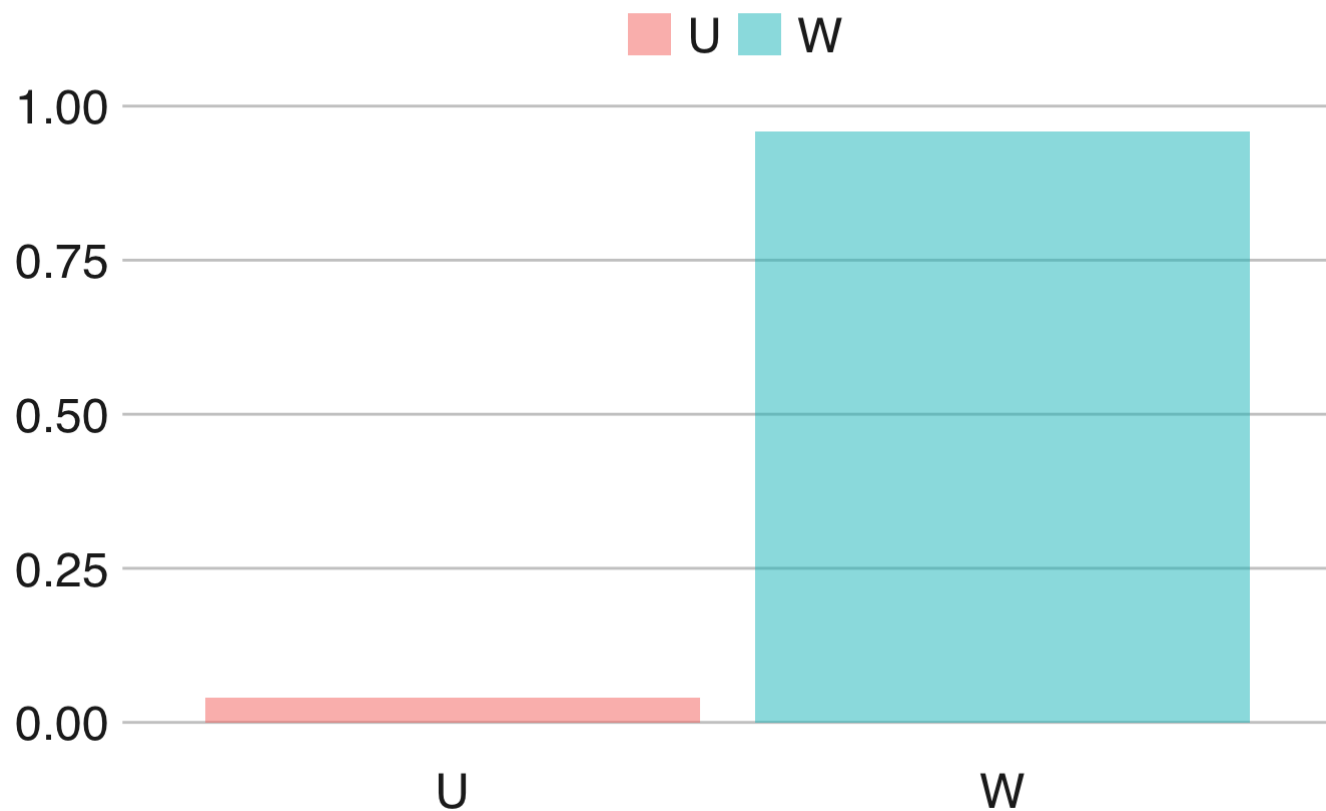
This distribution of white and African-Americans looks similar to the population during this time (1981-1990). Hispanics and Asians are greatly under-represented maybe because the majority of immigration of these groups happened after the PSID began. I believe there is a Latino sample attached to the PSID but did not seem important given the nature of the project.

Race Bar Plot



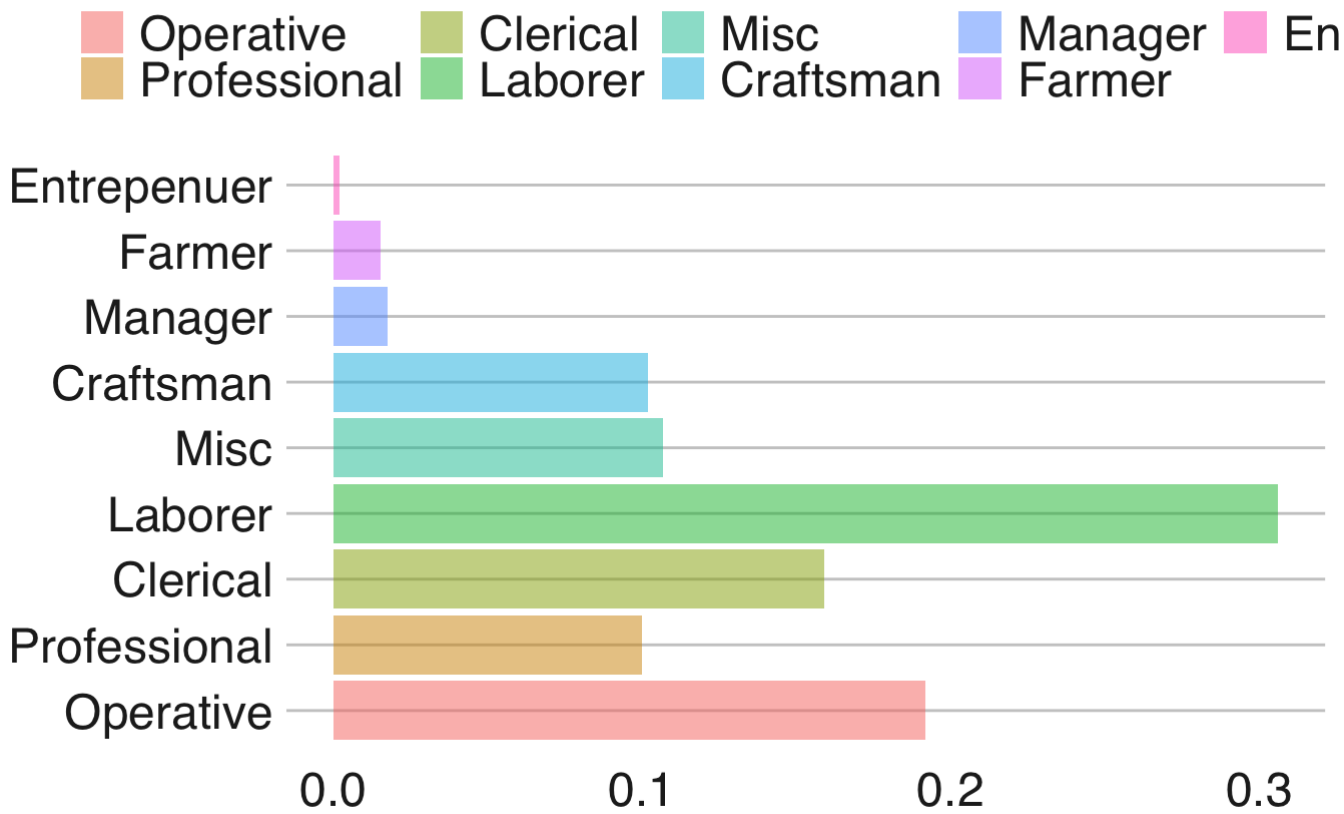
U is for Unemployed, W is for Working. Most people in the sample were working at one time or another.

Employment Last Year Bar Plot



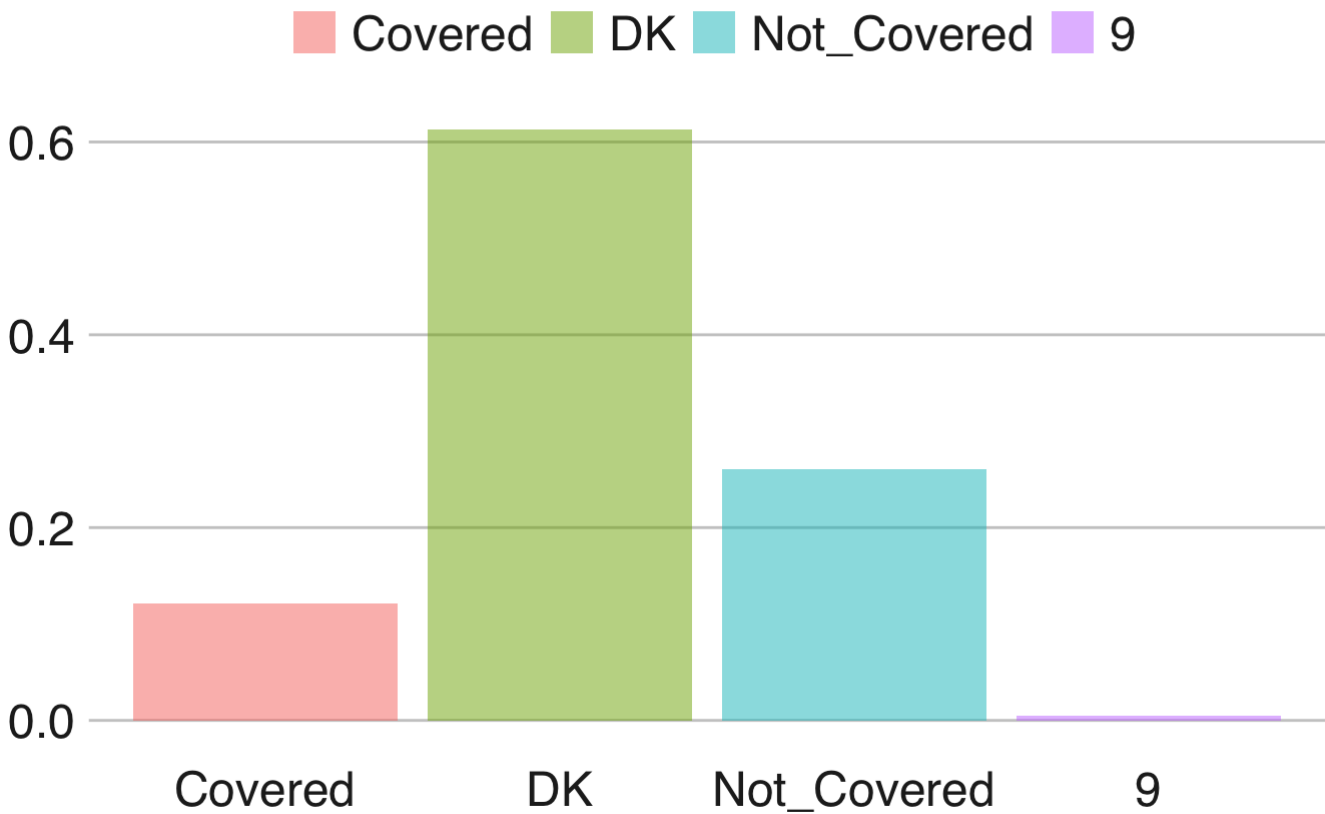
Most workers are laborers. Likely representative of the population

Occupation Bar Plot



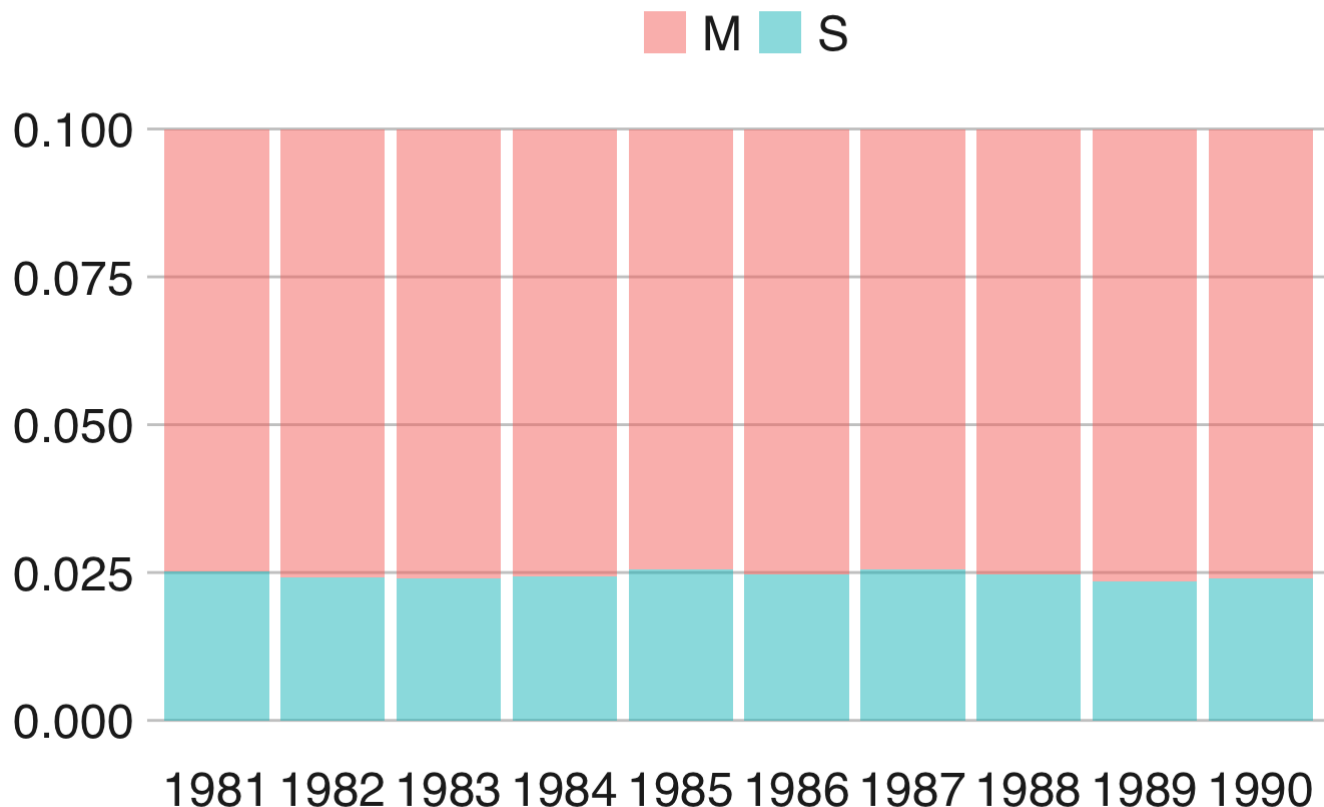
DK means ‘Don’t Know’ if wages are covered by union. There is a 9 in there meaning I did not account for this when cleaning. I am writing this at the end of my paper and cleaning the regression table so I do not wish to go back and clean this up. Doubt it is consequential.

Union Coverage Bar Plot



Most individuals are married over time.

Married Bar Plot



	ind_ef~t	idiosync~t	lnwage	comp_~85	hrs_ann	exp	exp2	race85	occ85	union~85	region85	mar	sex
ind_effect~t	1.0000												
idiosync_e~t	0.0078	1.0000											
lnwage	0.0920	0.4522	1.0000										
comp_edu85	-0.0491	-0.0004	0.3936	1.0000									
hrs_ann	-0.0133	-0.0001	-0.0852	0.0661	1.0000								
exp	0.0023	-0.0000	0.0351	-0.4057	-0.0221	1.0000							
exp2	0.0031	-0.0000	0.0076	-0.4147	-0.0294	0.9807	1.0000						
race85	0.0000	0.0000	-0.2275	-0.2301	-0.1008	0.0770	0.0752	1.0000					
occ85	-0.9743	-0.0076	-0.0334	-0.0687	0.0396	-0.0219	-0.0198	-0.0182	1.0000				
union_cov85	-0.0000	0.0000	0.1111	-0.1125	-0.1216	0.0348	0.0304	0.0780	0.0063	1.0000			
region85	0.0000	-0.0000	0.1789	0.1763	-0.0257	-0.0612	-0.0598	-0.2385	0.0221	0.0340	1.0000		
mar	0.0005	0.0000	-0.1943	0.0030	-0.1555	0.0218	0.0351	0.1348	-0.0781	-0.0079	-0.0300	1.0000	
sex	0.0000	0.0000	0.2315	0.0311	0.2066	-0.0718	-0.0772	-0.1395	0.1032	0.0052	0.0336	-0.6590	1.0000

Correlations

This is the output using the 'cor' command in Stata

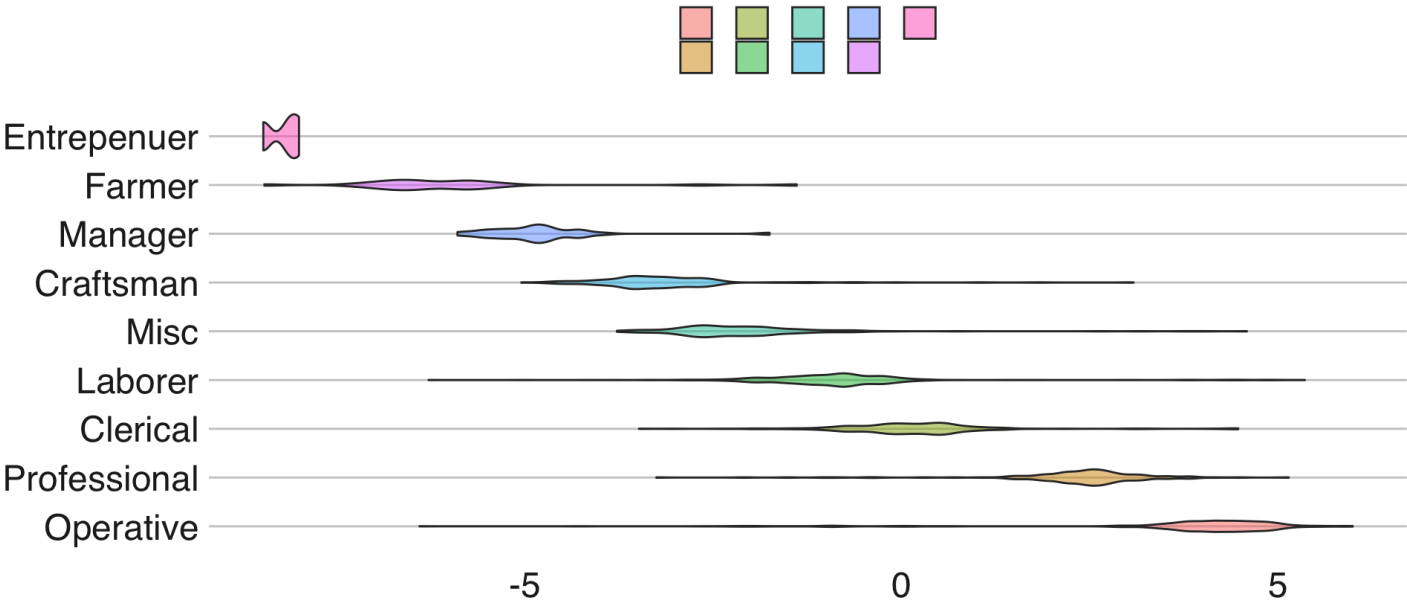
Regression Statistics

This is additional EDA from the Hausman Taylor regression. I made the individual effects error term a vector then categorized the residuals by occupation. What I think this shows is the distribution of ability within and between occupations. We see that professionals have a high level of ability while farmers, managers, and craftsman have lower levels. There are so few entrepreneurs in the survey that we should disregard this interpretation.

Lastly, note the differences in the tails between groups. Laborers have a long tail on both sides while groups with higher mean ability have long left-tails (and vice versa for those with lower mean ability).

0 is the mean ability of the sample.

Individual Effects by Occupation



This plot shows the correlations and distributions of the continuous variables in the analysis. The density is reports the variables in the columns. The row values are on the y-axis. u_{it} is the idiosyncratic error term and α_i is the individual error term.

ln(Wage)	Education	Experience	u_it	alpha_i	
					ln(Wage)
	Corr: 0.394***	Corr: 0.035***	Corr: 0.403***	Corr: 0.074***	
					Education
		Corr: -0.406***	Corr: -0.000	Corr: -0.048***	
					Experience
			Corr: 0.001	Corr: 0.004	
					u_it
				Corr: 0.008	
					alpha_i