ECON 695 Final Project

University of Wisconsin–Madison

Gender Wage Gaps and Coworker Wages

Alejandro De La Torre

Vibhu Mithal

Ankit Modi

Rachel Williams

December 2025

Latest version and replication files

AI Disclaimer

AI tools (specifically ChatGPT) were used solely to assist with minor troubleshooting and formatting tasks, including resolving LaTeX and Python issues, and drafting descriptive tables for improved readability. All analysis, model specifications, interpretations, and the final written narrative are our own.

Contents

Introduction

Data

We use administrative data provided to us by Professor Alice Wu for the ECON 695 final project that follow men and women who are observed for several consecutive years across exactly two jobs. The dataset (projectdata.csv) contains one record per individual, with each record summarizing the worker's demographic characteristics, education, potential labor market experience, and a sequence of annual wage observations surrounding a job transition. In total, the dataset includes 16,969 individuals—10,575 men and 6,394 women—who satisfy the sampling requirement of having at least three years of observations in their first job and three years in their second job. The timing convention indexes the first year on the second job as period 0, with years -3, -2, and -1 representing the three years prior to the move, and years 1 and 2 representing subsequent years on the second job.

A distinctive feature of the dataset is the inclusion of coworker wage measures: for each job spell, we observe a summary measure of the mean log wage of all other workers employed at the same firm. These coworker wage variables are recorded separately for the first job, where $owage1$ is defined as the average coworker wage across periods -1 and -2, and for the second job, where $owage2$ is defined as the average coworker wage across periods 0, 1, and 2. This structure enables us to investigate how the wage distribution among a worker's peers may influence their own wage outcomes. This richness makes the dataset particularly suitable for studying gender differences in wage determination and the role of workplace environments.

Sample Size and Structure

The dataset contains a total of 16,969 individuals, of whom 10,575 are men and 6,394 are women. Thus, approximately 38% of the sample consists of women. The sample is further restricted to workers who move between jobs where there are at least four coworkers at each job, so that coworker wage measures are well defined. Each individual record includes demographic characteristics, education, experience, and a sequence of wage observations around a job transition.

The timing convention follows the project instructions: period 0 is the first year on the second job, periods 1 and 2 are the subsequent years, and periods -1 through -3 correspond to the final three years on the first job. For each job spell, the dataset also includes a measure of the mean log wage of coworkers, which is central to our later analysis of coworker wage effects.

Variables

The dataset includes key demographic and human capital measures—age, years of education (taking values 6, 9, 12, or 16), gender (a binary indicator `female`), and potential experience constructed from age and schooling. Wage information is recorded as log hourly wages across multiple periods relative to a job transition: y denotes the wage in period 0 (the first year on the second job), $yp1$ and $yp2$ correspond to years 1 and 2 on the second job, and $yl1$, $yl2$, and $yl3$ capture wages in the final three years on the first job (periods -1 through -3). These coworker wage variables are summarized as $owage1$ (averaged over periods -1 and -2 for the first job) and $owage2$ (averaged over periods 0, 1, and 2 for the second job). Together, these variables allow us to link individual wage dynamics with characteristics of the worker's surrounding wage environment.

Section 1: Overview: Female vs. Male Workers

Female and male workers in our sample differ systematically in both pay and observed characteristics. Table 1 reports mean values for age, education, wages, and coworker wages for all workers and separately by gender, along with t–statistics comparing men and women at period 0 (the first year on the second job). Figure 1 then visualizes the distribution of log hourly wages by gender.

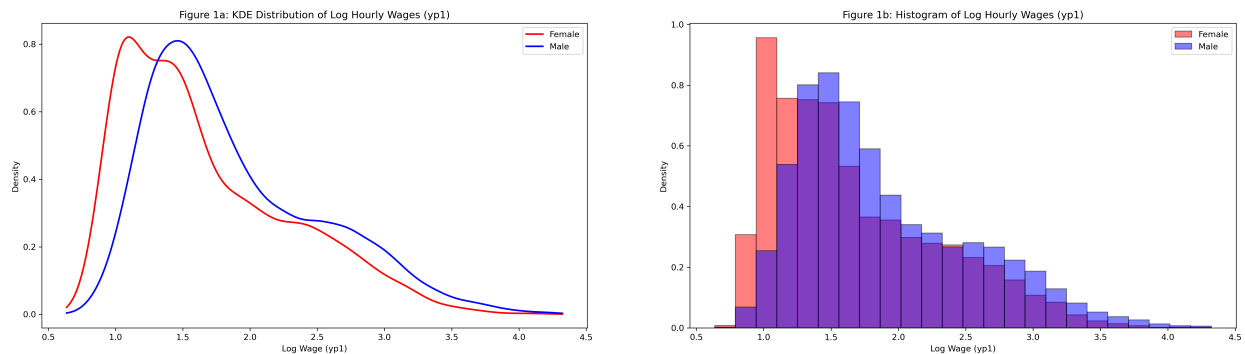## Summary Statistics

Table 1: Summary Statistics by Gender

	All	Female	Male	t–stat
Age (mean)	33.5589	33.5344	33.5737	-0.4358
Log wage y (mean)	1.7879	1.6580	1.8664	-20.8163
Coworker wage owage2 (mean)	1.6921	1.6378	1.7249	-11.6761
Educ = 6 yrs (fraction)	0.2741	0.2487	0.2895	-5.8460
Educ = 9 yrs (fraction)	0.2274	0.1949	0.2471	-8.0450
Educ = 12 yrs (fraction)	0.2958	0.3286	0.2759	7.2054
Educ = 16 yrs (fraction)	0.2027	0.2279	0.1875	6.2317

Women comprise approximately 38% of the sample. As shown in Table 1, average age is nearly identical across genders, but women tend to have slightly higher educational attain-

ment: they are more likely to have completed 12 or 16 years of schooling, whereas men are more concentrated in the 6- and 9-year categories. These differences in educational composition are large and statistically significant, with t–statistics in the range of about 6 to 8 in absolute value.

Despite women's modest advantage in formal schooling, men earn substantially higher wages at period 0. The mean log hourly wage is 1.866 for men compared to 1.658 for women, corresponding to a large and statistically significant gender gap (the t–statistic is about -20.8). Coworker wages show a similar pattern: men are more likely to work alongside higher-paid coworkers (mean owage2 of 1.725 for men versus 1.638 for women), suggesting systematic gender differences in the types of jobs or firms where workers are employed.

Wage Distributions

(a) Kernel density of log hourly wages ($yp1$) by gender.

(b) Histogram of log hourly wages ($yp1$) by gender.

Figure 1: Distribution of log hourly wages in period 1 on the second job ($yp1$) for women and men. Panels (a) and (b) show kernel densities and histograms, respectively.

Figure 1 provides additional detail on the distribution of wages by gender. In panel (a), the kernel density curves indicate that the female wage distribution lies to the left of the male distribution, with women clustering more heavily around log wages of roughly 1.0–1.4. Panel (b) shows a similar pattern in the histograms: women are more prevalent at lower wage levels, while men are overrepresented in the upper tail, particularly above a log wage of 2.0. The male distribution also exhibits greater spread, consistent with higher dispersion in men's wage outcomes.

Taken together, Table 1 and Figure 1 document a sizable raw gender wage gap and meaningful differences in educational attainment and coworker wage exposure, even before controlling for other observable characteristics. These descriptive patterns motivate the regression analysis and Oaxaca decompositions in Section ??.

Section 2: Gender Wage Gaps

2.1 Wage Regressions and Oaxaca Decomposition

2.2 Gender Difference in Experience Profiles

Section 3: Gender Wage Gaps Conditional on Coworker Wages

Section 4: Event Study - Wage Changes around Moves

Section 5 (Bonus): Shrinkage

Conclusion

Appendix