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

Acknowledgements. We thank Professor Alice Wu for guidance and comments. All errors are our own.

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

AI Disclaimer . . . 2
Introduction . . . 4
Data . . . 4
  Sample Size and Structure . . . 4
  Variables . . . 5
Section 1: Overview: Female vs. Male Workers . . . 5
  Summary Statistics . . . 5
  Wage Distributions . . . 6
Section 2: Gender Wage Gaps . . . 7
  2.1 Standard Models and Oaxaca Decomposition . . . 7
  2.2 Gender Difference in Experience Profiles . . . 10
Section 3: Gender Wage Gaps Conditional on Coworker Wages . . . 10
Section 4: Event Study - Wage Changes around Moves . . . 10
Section 5 (Bonus): Shrinkage . . . 10
Appendix . . . 10

References **11**

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

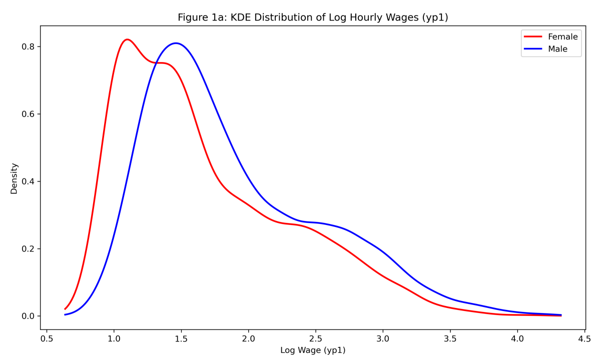
	All	Female	Male	t–stat
Age (mean)	33.5589	33.5344	33.5737	-0.4358
Log wage y (mean)	1.7879	1.6580	1.8664	-20.8163
Coworker wage owage2 (mean)	1.6921	1.6378	1.7249	-11.6761
Educ = 6 yrs (fraction)	0.2741	0.2487	0.2895	-5.8460
Educ = 9 yrs (fraction)	0.2274	0.1949	0.2471	-8.0450
Educ = 12 yrs (fraction)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

Notes: Each column reports sample means for the listed variables across all workers, and separately by gender. Education is coded as categorical dummies for 6, 9, 12, and 16 years of schooling. The final column reports t–statistics from independent two-sample tests comparing male and female means. All values are measured at period 0 (first year on the second job).

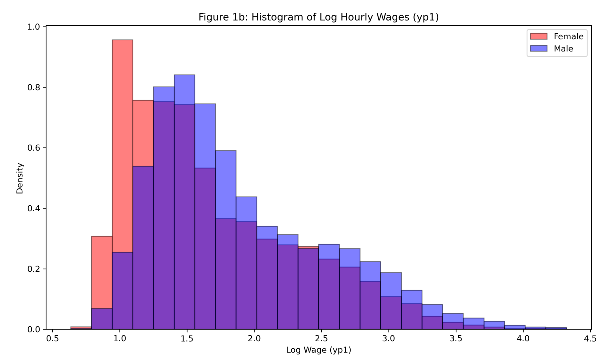
Women comprise approximately 38% of the sample. As shown in Table 1, average age is nearly identical across genders, but women tend to have slightly higher educational attainment: they are more likely to have completed 12 or 16 years of schooling, whereas men are more concentrated in the 6- and 9-year 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 2: Gender Wage Gaps.

Section 2: Gender Wage Gaps

From our descriptive analysis, it is self evident that there exists substantial raw differences in wages, education, and coworker wage environments between men and women. These gaps raise a natural question: to what extent do observable characteristics—such as education and potential experience—account for the wage differences we see, and how much remains unexplained by these factors? Here we begin to attempt to formally investigate these questions by estimating standard wage models and applying the Oaxaca decomposition to separate the portion of the gender wage gap attributable to differences in characteristics from the portion attributable to differences in returns to those characteristics.

2.1 Standard Models and Oaxaca Decomposition

Table 2: Wage Models and Oaxaca Decomposition of the Gender Wage Gap

	(1) Female Dummy	(2) Full Model	(3) Men	(4) Women
Intercept	1.866	0.557	0.491	0.327
	(0.006)	(0.056)	(0.078)	(0.080)
Female	-0.208	-0.271		
	(0.010)	(0.007)		
C(educ)[T.9]		0.275	0.259	0.314
		(0.009)	(0.012)	(0.015)
C(educ)[T.12]		0.680	0.692	0.667
		(0.009)	(0.012)	(0.014)
C(educ)[T.16]		1.515	1.508	1.525
		(0.011)	(0.014)	(0.016)
exp		0.067	0.072	0.069
		(0.011)	(0.015)	(0.015)
exp2		-0.001	-0.001	-0.001
		(0.001)	(0.001)	(0.001)
exp3		-0.000	-0.000	0.000
		(0.000)	(0.000)	(0.000)
R-squared	0.024	0.588	0.556	0.618
Adj. R^2	0.024	0.588	0.556	0.618
N	16,969	16,969	10,575	6,394

Notes: Columns (1)–(2) report pooled OLS models; Columns (3)–(4) estimate the same specification separately by gender. Standard errors in parentheses. Oaxaca decomposition uses gender-specific models.

Across the pooled models, the coefficient on the female indicator is negative and statistically significant, indicating that women earn lower wages on average even before conditioning on observable characteristics. In the baseline model that includes only a constant and the female dummy, the coefficient of approximately -0.208 suggests that women earn roughly 20% lower wages than men at period 0. After adding categorical controls for education and a cubic polynomial in experience, the wage gap widens to about 27%, reflecting that women in our sample tend to have slightly higher educational attainment and similar experience levels. Thus, once controlling for these favorable characteristics, the remaining wage gap attributable to gender becomes larger.

The separate regressions by gender show broadly similar returns to education and experience for men and women. Returns to schooling rise with each additional educational category, while experience exhibits diminishing marginal returns. The largest difference lies in the intercepts: men have a higher baseline log wage than women, implying that even at comparable education and zero experience, men start from a higher predicted wage level.

Yet this baseline gap is muddied between differences in the composition of the characteristics of each group (e.g., average education levels) and differences in the returns to those characteristics (e.g., how much additional education translates to wage increases). In other words, while these regressions quantify the wage gap and its relationship to observable traits, part of the wage gap may be due to differences in the average education and experience levels (composition effects), while another being due to differences each group being paid differently for the same education or experience, or actual gender discrimination, which is the object of our study.

To disentangle these two effects, we apply the Oaxaca decomposition from Oaxaca (1973) using the gender-specific models from Columns (3) and (4) to formally separate the composition-driven (between) component from the differences in returns (within) component. To model this, let y_{gi} denote log hourly wages in period 0 for individual i in group $g \in \{m, f\}$ (men, women), and let X_{gi} be the corresponding row vector of observable characteristics (education dummies and a cubic polynomial in potential experience).[1] The group–specific wage equations are

$$y_{gi} = \alpha_g + X'_{gi}\beta_g + \varepsilon_{gi}, \qquad g \in \{m, f\}. \tag{1}$$

Taking expectations and using sample means,

$$\bar{y}_g = \hat{\alpha}_g + \bar{X}'_g\hat{\beta}_g,$$

where $\bar{y}_g$ and $\bar{X}_g$ are the sample means for group g and $(\hat{\alpha}_g, \hat{\beta}_g)$ are the OLS estimates from

[1] The intercept is kept separate and not included in X_{gi}.

the gender–specific regressions.

We define the (raw) gender wage gap as the difference in mean log wages between men and women,

$$\Delta \equiv \bar{y}_m - \bar{y}_f.$$

Using women as the reference group for the "non–discriminatory" wage structure, the Oaxaca (1973) decomposition writes this gap as

$$\Delta = (\bar{X}_m - \bar{X}_f)'\hat{\beta}_f \; + \; \bar{X}_m'(\hat{\beta}_m - \hat{\beta}_f). \tag{2}$$

The first term,

$$(\bar{X}_m - \bar{X}_f)'\hat{\beta}_f,$$

is the *explained* or *between–group* component: it captures how differences in average characteristics (education and experience) between men and women would translate into a wage gap if both groups were paid according to the same returns $\hat{\beta}_f$ estimated for women. The second term,

$$\bar{X}_m'(\hat{\beta}_m - \hat{\beta}_f),$$

is the *unexplained* or *within–group* component: it reflects differences in the returns to characteristics between men and women, evaluated at men's average characteristics. In our implementation, the explained and unexplained components in equation (2) correspond exactly to the two pieces reported in the decomposition based on the separate male and female regressions used for Table 2.

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

References

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review 14*(3), 693–709.