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 were used sparingly and solely to assist with minor troubleshooting and formatting tasks, including resolving typesetting and Python issues, and formatting descriptive tables for improved readability. All analysis, model specifications, interpretations, and the final written narrative are our own.

Contents

Introduction

Understanding why men and women continue to earn different wages remains one of the central questions in labor economics. While the raw gender pay gap is well documented, much less is known about how workplace environments, job transitions, and the characteristics of coworkers shape these disparities (Blau and Kahn, 2017; Brick et al., 2023). Our dataset offers a rare opportunity to study these mechanisms directly: we observe workers across consecutive jobs, along with detailed information on their coworkers' average wages. This structure allows us to examine not only individual wage differences, but also the role of sorting, workplace quality, and the returns workers receive in different environments.

We combine descriptive evidence, standard wage regressions, Oaxaca decomposition, event-study analyses of job moves, and shrinkage methods such as Ridge and Lasso to build a more complete picture of the gender wage gap. Each approach highlights a different mechanism: differences in characteristics, differences in returns, mobility patterns, and model robustness—and together they help us understand how gender shapes the wage trajectories workers experience. Our goal is not to resolve every dimension of the pay gap, but to make progress in identifying which factors matter most and how they interact in practice.

Data

We use administrative data provided to us by Professor Alice Wu for the ECON 695 final project Wu (2025a) that follow men and women who are observed for several consecutive years across exactly two jobs (Wu, 2025b). 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. The richness of the dataset makes it particularly suitable for studying gender differences in wage determination and the role of workplace environments.

Sample Size and Structure

As described above, the dataset contains a total of 16,969 individuals, of whom 10,575 are men and 6,394 are women. Thus, this sample is limited in the fact that approximately 38% consists of women. It 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).

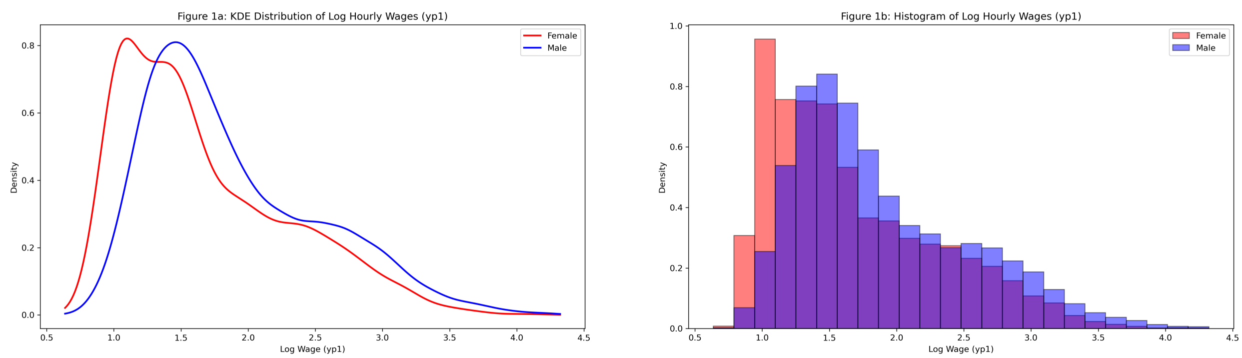
As shown in Table 1, average age is nearly identical across genders, but women tend to have slightly higher educational attainment: they are more likely to have completed 12 or 16 years of schooling, whereas men are more concentrated in the 6- and 9-year categories. These differences in educational composition are large and statistically significant, with t–statistics in the range of about 6 to 8 in absolute value.

Overall, the summary statistics show men earn higher average wages, exhibit greater wage dispersion, and are disproportionately represented among high earners, whereas women are more concentrated in lower-wage ranges despite having slightly higher average educational attainment. According to Table 1, men earn substantially higher wages on average, even before controlling for other variables as reflected in the large magnitude of the t-statistic. Furthermore, Table 1 reveals that although men earn higher wages, women are more likely to have 12 to 16 years of education while men tend to have a lower education of 6 to 9 years.

Figure 1 supports Table 1's narrative by displaying the female wage distribution lying consistently to the left of the male distribution and the female distribution peaking before the male distribution (around 1.2-1.3). In other words, Figure 1 shows men are disproportionately represented among very high-wage earners. Complementing this, Figure 1 suggests that women cluster heavily at lower log wages (around 1.0–1.4), whereas men dominate the range above 2.0 and exhibit a more dispersed, right-shifted distribution in general.

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 compared to 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.

From Figure 1, we find a clear difference in the distribution of wages by gender. In panel 1a, 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 1b 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. Of course this only begins to unpack the mere existence of a gap in the sample data. To begin to identify the reasons behind the gap, we conduct regression analysis and Oaxaca decomposition, which we discuss in Section 2: Gender Wage Gaps.

Figure 1 supports Table 1's narrative by displaying the female wage distribution lying consistently to the left of the male distribution and the female distribution peaking before the male distribution (around 1.2-1.3). In other words, Figure 1 shows men are disproportionately represented among very high-wage earners. Complementing this, Figure 1b suggests that women cluster heavily at lower log wages (around 1.0-1.4), whereas men dominate the range above 2.0 and exhibit a more dispersed, right-shifted distribution in general.

Section 2: Gender Wage Gaps

From our descriptive analysis, it is 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? We begin to attempt to 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.

In the pooled regression that includes a constant and only a female dummy, the coefficient on $female = -0.2084$ indicates that, holding no other variables constant, women earn about 20.8% lower wages than men on average. This is representative of the fundamental gender wage gap. On adding categorical education and a cubic polynomial in experience as controls to our pooled regression, the coefficient on $female = -0.2706$ which reveals that on controlling for education and experience, women earn approximately 27.06% less than men on average. Since the coefficient becomes increasingly negative on adding the respective controls, we interpret that women in our sample are on average more educated and experienced than men. Thus controlling for these variables widened the wage gap.

In the separate by gender model, the coefficients on education split categorically and on experience are positive and significant for men and women alike. As education levels increase, returns to education on wages rise sharply, while the returns to experience on wages seem to show diminishing returns with increase in experience. On comparing the two groups, we conclude that men have much higher intercept values at 0.4906 compared to 0.3270 for women. This indicates that the predicted log wage for men with 6 years of education and no experience is 0.4906 whereas it is 0.3270 for women. While the returns to education are quite similar for both men and women, men have a marginally higher return to experience. This means that controlling for other variables, for a unit increase in years of experience, log wage rose by 0.0723 for men and 0.0686 for women.

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 separate the portion of the wage gap arising from differences in average characteristics from the portion arising from differences in how those characteristics are rewarded in the labor market, we apply the Oaxaca decomposition of Oaxaca (1973) using the gender-specific regressions in Columns (3) and (4). Let y_{gi} denote log hourly wages for individual i in group $g \in \{m, f\}$ and let X_{gi} be the vector of observable characteristics (education dummies and a cubic polynomial in potential experience).[1] We model wages as

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

Taking sample means yields

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

The raw gender wage gap is defined as

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

Using women as the reference group, the Oaxaca decomposition expresses Δ 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 reflects how differences in average education and experience across men and women would translate into a wage gap if both groups received the same returns estimated for women.

[1] The intercept is excluded fromX_{gi} and treated separately.

The second term,

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

is the *unexplained* or *within-group* component: it captures differences in the returns to these characteristics, evaluated at men's mean characteristics.

The Oaxaca Decomposition result shows the total wage gap is approximately 0.045 log points, meaning men earn about 4.5% higher wages than women on average in period 0. Based on the included education and experience characteristics, the *explained component* (the between component) = –0.0623 indicates that women should earn higher wages than men. That is to say, if the labor market rewarded these characteristics equally for both genders, the wage gap would favor women. This is intuitive as women in the sample have slightly higher educational attainment and comparable experience as displayed above. However, given the unexplained component reflects differences in the returns to characteristics by definition, women are not favored in this case since the *unexplained component* (the within component) = 0.1071 suggests that, even though women possess characteristics associated with higher wages, they do not receive the same returns as men. Ultimately, the Table 2 results show a positive wage gap favoring men.

2.2 Gender Difference in Experience Profiles

The existence of a wage gap we identified in 2.1 Standard Models and Oaxaca Decomposition is not monotone as behaviors, attitudes, lawsuits, macro-trends, and many other of the world's moving parts can influence changes in any such gap we have identified and maybe too effects people differently. As such, it is necessary to ask how it evolves over the life cycle. Here, we explore gender differences in the relationship between log wages and potential experience, focusing first on workers with 12 years of education and then extending the comparison across the full education distribution.

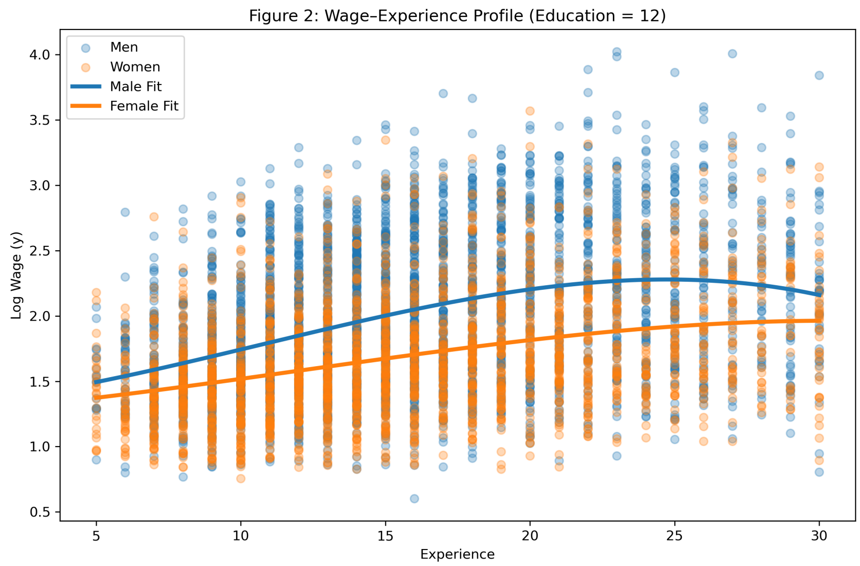
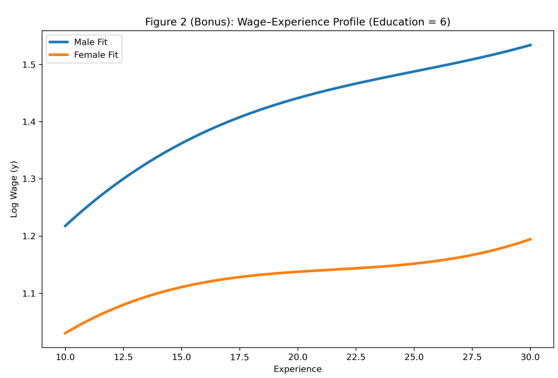
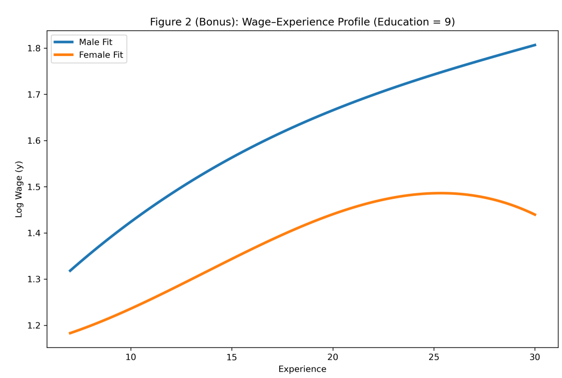


Figure 2: Experience–wage profiles by gender for workers with 12 years of education. Notes: Figure plots log hourly wages against potential experience (5–30 years) and overlays the fitted values from a cubic polynomial in experience, estimated separately for men and women.

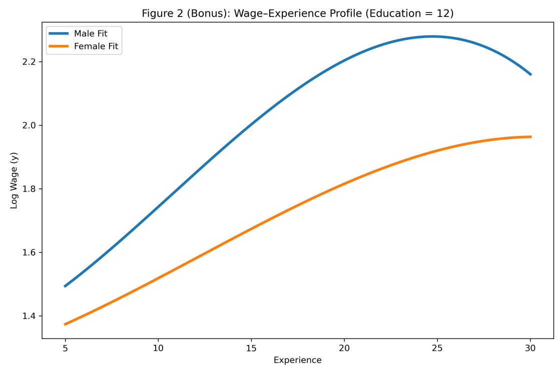
In the primary Figure 2 plot (education = 12), male wages start above female wages even at low experience levels, and the gap widens as experience increases. The male curve shows a sharper upward slope and peaks at a higher level than the female profile. In contrast, the female curve is flatter which suggests that women receive smaller marginal returns to experience. Both curves show wages grow quickly early in a career and then level off. Notably, concurrent to Figure 1, this effect is stronger for men as the women's profile peaks earlier and at a substantially lower predicted wage. Together, these patterns visually reinforce the Oaxaca decomposition result that women have similar and even more favorable education levels than men, but the returns to those characteristics differ considerably.



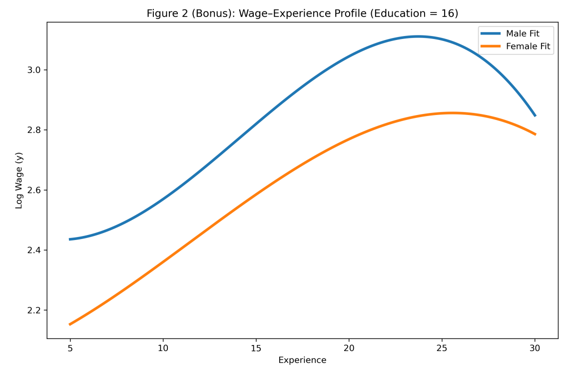
(a) Educ = 6



(b) Educ = 9



(c) Educ = 12



(d) Educ = 16

Bonus Figure: Experience–wage profiles by gender and education level. Each panel plots log hourly wages against potential experience for the indicated education category and overlays gender-specific cubic fits. This bonus panel is intentionally unnumbered and does not contribute to the main sequence of figures.

The bonus plots for education levels 6, 9, 12, and 16 reveal that the pattern holds across the entire education distribution. At all education levels, male predicted wages exceed female predicted wages for the same experience. Moreover, male returns to experience are consistently larger as the slope of the male curve is steeper in the early career years and male profiles peak at higher levels of log wages which strengthens the results from Figure 1and Table 2. Thus, even at higher education levels (e.g., 16 years), although both men and women earn more overall, the gender gap persists and grows with experience. In other words, these panels visually reinforce that women begin their careers at a lower baseline and appear to receive systematically weaker returns to experience, even when education is held fixed.

Section 3: Gender Wage Gaps Conditional on Coworker Wages

So far, we have shown that differences in education and experience explain only part of the gender wage gap. In particular, up until this point, we know that a substantial portion of the gender wage gap reflects differences in the *returns* to education and experience rather than differences in characteristics themselves. However, wages are not determined in isolation. They are shaped by the environments people work in (**?**). In particular, we ask: does the wage level of the people you work with affect your own wage and does this differ by gender?

Specifically, we ask whether working alongside higher-paid coworkers boosts individual wages, and whether this effect differs by gender. Namely, we consider if men tend to work in higher-paying environments, and if those environments raise wages, then part of the gender gap may reflect where people work, not just who they are.

To attempt to address this, it is necessary to examine the average wage of a worker's coworker (`owage2`) by including it into our regression models to then examine how it interacts with gender. This way, it becomes possible to test whether men and women benefit equally from high-wage peer groups and whether differences in coworker exposure help explain the residual wage gap.

Therefore, we extend our baseline wage models by incorporating coworker wages directly into the regression framework. By comparing pooled and gender-specific specifications, we evaluate both *how strongly coworker wages predict individual wages* and *whether the returns to working with higher-paid coworkers differ by gender*. These estimates then feed into a new decomposition that asks: to what extent do gender differences in coworker wage exposure explain the remaining wage gap after controlling for education and experience?

We begin by estimating a set of pooled and gender-specific regressions that progressively incorporate `owage2` and its interaction with gender as shown below in Table 3.

Table 3: Wage Models with Coworker Wages and Gender Interactions

	(1) Pooled: female + owage2	(2) Pooled: full + owage2	(3) Pooled: + female×owage2	(4) Men	(5) Women
Intercept	0.131	-0.199	-0.244	-0.297	-0.284
	(0.013)	(0.047)	(0.047)	(0.064)	(0.066)
female	-0.121	-0.189	-0.071		
	(0.007)	(0.005)	(0.020)		
owage2	1.006	0.634	0.662	0.660	0.598
	(0.007)	(0.007)	(0.008)	(0.009)	(0.010)
C(educ)[T.9]		0.153	0.154	0.150	0.165
		(0.008)	(0.008)	(0.010)	(0.013)
C(educ)[T.12]		0.377	0.377	0.394	0.349
		(0.008)	(0.008)	(0.011)	(0.013)
C(educ)[T.16]		1.011	1.013	1.030	0.978
		(0.010)	(0.010)	(0.013)	(0.016)
exp		0.056	0.055	0.056	0.062
		(0.009)	(0.009)	(0.012)	(0.013)
exp2		-0.001	-0.001	-0.001	-0.002
		(0.001)	(0.001)	(0.001)	(0.001)
exp3		-0.000	-0.000	-0.000	0.000
		(0.000)	(0.000)	(0.000)	(0.000)
female:owage2			-0.070		
			(0.011)		
R-squared	0.549	0.727	0.727	0.707	0.748
Adj. R^2	0.548	0.727	0.727	0.707	0.748
N	16,969	16,969	16,969	10,575	6,394

Notes: Columns (1)–(3) report pooled OLS models of log wage y that include coworker wages (owage2). Column (1) includes only a female dummy and coworker wage. Column (2) adds categorical education and a cubic polynomial in experience. Column (3) further interacts coworker wages with the female dummy. Columns (4)–(5) estimate the same specification as Column (2) separately for men and women. Standard errors are in parentheses. Oaxaca decompositions based on these gender-specific models are reported and discussed in the text rather than in the table.

Table 3 shows that coworker wages are strongly correlated with individual wages. Across all pooled specifications, the coefficient on `owage2` is large and precisely estimated, and the R^2 rises sharply relative to the models in Table 2, indicating that coworker wage environments explain a substantial share of wage variation beyond individual education and experience.

Column (3) introduces an interaction between coworker wages and gender and reveals meaningful heterogeneity in returns: the estimated coefficient on `female` $\times$ `owage2` is approximately -0.07, implying that women receive a significantly smaller marginal wage benefit from working alongside higher-paid coworkers than men do. This pattern is reinforced in Columns (4) and (5), which estimate the full specification separately by gender. While coworker wages are strongly associated with higher wages for both groups, the estimated return to `owage2` is modestly larger for men than for women.

Taken together, these results suggest that coworker wage environments matter for wage determination, but in a gendered way: men both tend to work with higher-paid coworkers and appear to receive larger returns from those environments. However, these regression results alone do not indicate whether coworker wages explain the gender wage gap primarily through differential sorting into workplaces or through differences in how similar environments are rewarded. To disentangle these channels, we next extend the Oaxaca decomposition framework to incorporate coworker wages directly.

For each group $g \in \{m, f\}$, let y_{gi} denote log hourly wages in period 0 and let X_{gi} collect the observable characteristics. Relative to the earlier decomposition, X_{gi} now includes education dummies, a cubic polynomial in potential experience, *and* coworker wages `owage2` (the intercept is again treated separately). The group–specific wage equations estimated in Columns (4) and (5) of Table 3 can be written as

$$y_{gi} = \alpha_g + X_{gi}'\beta_g + \varepsilon_{gi}, \qquad g \in \{m, f\},$$

with corresponding sample means

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

The raw gender wage gap is $\Delta \equiv \bar{y}_m - \bar{y}_f$, and, using women as the reference group, the Oaxaca decomposition becomes

$$\Delta = (\bar{X}_m - \bar{X}_f)'\hat{\beta}_f \ + \ \bar{X}_m'(\hat{\beta}_m - \hat{\beta}_f). \tag{3}$$

As before, the first term $(\bar{X}_m - \bar{X}_f)'\hat{\beta}_f$ is the *explained* or *between–group* component, now capturing differences in average education, experience, and coworker wages evaluated at the female returns $\hat{\beta}_f$. The second term $\bar{X}_m'(\hat{\beta}_m - \hat{\beta}_f)$ is the *unexplained* or *within–group* component, reflecting differences in how these characteristics, including coworker wages, are rewarded for men and women.

Applying equation (3) to the gender–specific estimates in Columns (4) and (5) of Table 3, we obtain an explained component of about 0.0168 log points and an unexplained component of about 0.2043 log points, yielding a total gap of $\Delta \approx 0.2211$. Hence, once we account for education, experience, and coworker wages, differences in observable characteristics, including the fact that men tend to work with somewhat higher-paid coworkers, explain only a small fraction of the gender wage gap, while the majority of the gap is still attributed to differences in returns.

Interpreting the Role of Coworker Wages The introduction of coworker wages into the wage models opens a new window into how workplace environments shape wage inequality. What stands out immediately from Table 3 is the magnitude of the `owage2` coefficient: workers who are surrounded by higher-paid peers earn markedly higher wages themselves. This relationship is large and remarkably stable across pooled and gender-specific specifications. The gender pattern layered on top is equally striking. In every relevant model, women receive smaller marginal returns to high-wage coworker environments than men. Even after conditioning on education and experience, the wage benefits associated with working alongside highly paid coworkers appear systematically muted for women.

This raises a deeper question: why does working with higher-paid coworkers raise wages, and why does this return differ by gender? To discipline the narrative, we thought through the two economic hypotheses laid out in the assignment, each of which implies a different reading of what our decomposition captures.

Model 1: Coworker wages reflect luck, networks, or access to better jobs. Under this view, sorting into high-wage coworker environments is not mainly about productivity but about opportunity. Workers may reach these jobs through referrals, informal hiring networks, or good fortune. If men have stronger networks, or search more aggressively, they may disproportionately end up in higher-wage peer groups. In this interpretation, the explained component of the decomposition—about 0.0168 log points, or roughly 8% of the total 0.2211 gap—captures an inequality in job access, not differences in productivity. Men are more likely to land in workplaces with higher-paid coworkers, and those environments boost wages. Thus, men benefit twice: first from better job sorting, and second from stronger returns to those environments. Model 1 treats the large unexplained component ($\approx$ 92%) as a reflection of unobserved variables, such as discrimination or ability. If job access is partly random or network-driven, then the structural gap in returns to coworker wages may signal discriminatory pay-setting or organizational norms that reward men more for the same workplace exposure.

Model 2: Coworker wages proxy for unobserved productivity or ambition. A different interpretation emerges if coworker wages simply reflect who ends up working with whom. High-ability workers may seek out (or be recruited into) high-skill, high-paying teams. Through this lens, `owage2` acts as a measurable proxy for something we cannot otherwise observe, including cognitive skills, ambition, or complementary abilities. If men and women differ in these unobserved traits—or if women face barriers that suppress the labor market rewards to those traits—the decomposition shifts meaningfully.

Under Model 2, the explained portion of the gap still reflects differences in coworker wage exposure, but this exposure is now interpreted as an indirect measure of latent productivity. The fact that only 8% of the gap is explained would then suggest that men and women with similar observed characteristics differ substantially in unobserved characteristics, or that employers reward unobserved male traits disproportionately. Meanwhile, the unexplained component captures both these latent differences and potential structural forces, including bias, norms, and bargaining disparities, that influence how unobserved skills are monetized.

How the decomposition informs these competing interpretations. The decomposition result reinforces that coworker wage exposure is relevant but on its own does not account for the bulk of the gender wage gap. Even after controlling for education, experience, and coworkers' wages, sorting differences explain only a small portion of wage inequality. Most of the gap arises from differences in returns, not from systematic differences in observed characteristics or environments.

The direction of the results also matters. The explained component is *positive*: men sort into higher-wage coworker groups than women, and this sorting pushes their wages upward relative to women. Nevertheless, the explained share is small, the narrative is dominated by the unexplained component. This feature is consistent with both models, but its implications differ:

- Under Model 1, the large unexplained component suggests that unequal returns to workplace exposure, potentially driven by discrimination or gendered organizational dynamics, are a major driver of the gap.
- Under Model 2, the unexplained component would reflect gender differences in unobserved productivity, or differences in how the labor market rewards those traits.

Both mechanisms are plausible, and the decomposition alone cannot adjudicate between them. However, two findings stand out: (i) men benefit more from high-wage peers than women do, and (ii) men sort into those environments more frequently. Taken together, these patterns underscore that workplace environments are not gender-neutral, and that the processes determining sorting and wage-setting differ meaningfully by gender.

We have yet to uncover which of these stories best captures the role of coworker wages in practice. If coworker wage exposure truly reflects unobserved ability (Model 2), then we would expect workers with strong underlying skills or ambition to remain in, or quickly return to, high-wage peer groups even as their careers evolve. If instead coworker wages primarily reflect access, luck, or networks (Model 1), then movements across jobs could generate astute changes in coworkers' wage profiles that are vaguely tied to individual productivity. This ambiguity in interpretation motivates our next step.

Section 4: Event Study - Wage Changes around Moves

Observing how workers move between different coworker environments over time probes whether coworker wages are acting as structural inputs to wage determination or as signals of deeper, unobserved differences across workers.

By comparing wage paths for workers who transition across coworker-wage terciles, we can observe whether changes in coworker environments are associated with systematic changes in workers' own wages, and whether these patterns align more closely with a sorting-based or ability-based explanation.

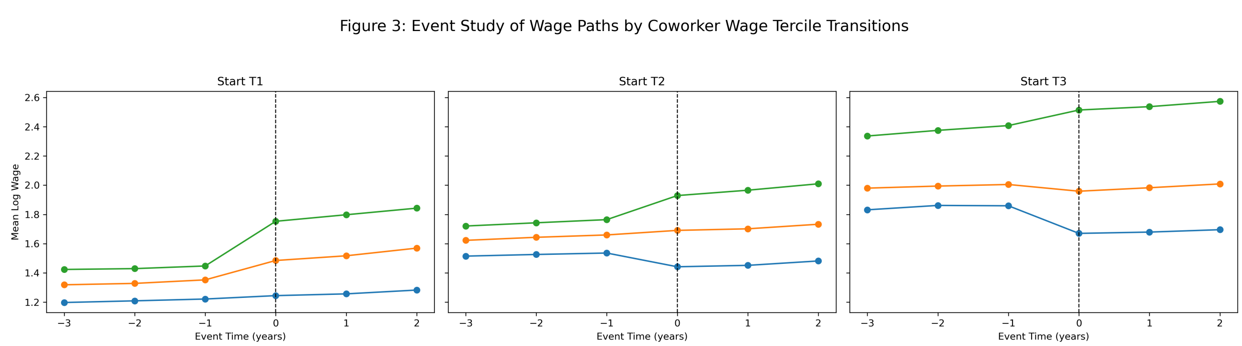


Figure 3: Event study of mean log wages for job changers, plotted from three years before to two years after the move. Each panel corresponds to the worker's initial coworker-wage tercile (T1, T2, T3). Within each panel, lines indicate the tercile of the coworker-wage environment at the *second* job. The vertical dashed line marks the job transition.

Figure 3 compares wage trajectories for movers, grouped by where they start in the coworker-wage distribution (`owage1` tercile) and where they land after moving (`owage2` tercile). Across the panels where workers move up to higher coworker-wage terciles, there is a clear upward shift in mean log wages at event time 0 (the job change) followed by a gradual upward trajectory afterward, especially for moves to End T3. Conversely, moves that stay within the same tercile show gradual wage growth with no large break at the transition and moves that progress down to lower coworker-wage terciles show a drop at event time 0.

These patterns provide more support for Model 1 than Model 2. Given the strongest change happens right at the move, the graphs suggest that switching into a higher-coworker-wage workplace is a consequence of shorter-term mechanisms such as luck, connections, or search, rather than a long-running upward trajectory. It is expected that if Model 2 were the main mechanism, movers who end up with high-paid coworkers already display steeper wage growth before the move since model 2 correlates higher pay with higher skills and ambition. However, the graphs do not rule out Model 2 entirely as selection on unobserved ability could still matter if higher-skill workers are the ones who are able to access high-coworker-wage jobs, and the move could occur due to mechanisms such as promotion. Nevertheless, the event-study evidence suggests that access to higher-coworker-wage workplaces, rather than gradual skill accumulation, plays the dominant role in driving wage gains which provides stronger support for Model 1.

Having seen in Figure 3 that wage jumps tend to accompany moves into higher coworker–wage environments, we next ask how much of this relationship survives once we difference out time–invariant worker attributes. To do so, we model the change in log wages from period -1 to period 0,

$$\Delta y_i \equiv y_i - y_{i,-1},$$

as a function of the corresponding change in coworkers' mean log wages,

$$\Delta \text{owage}_i \equiv \text{owage2}_i - \text{owage1}_i.$$

Table 4 reports pooled first–difference regressions of Δy_i on Δowage_i, first without additional controls and then adding education dummies, a quadratic in experience as of period -1, and a female indicator. Comparing these estimates to the level regressions in Table 3 allows us to gauge what fraction of the large coworker–wage coefficients reflects sorting or fixed individual traits, and what fraction remains even after we remove time–invariant heterogeneity.

Table 4: Wage Changes and Coworker Wage Growth

	Intercept	Δow	Ed9	Ed12	Ed16	exp	exp^2	exp^3	Female	R^2	Adj. R^2
(1) Δy	0.038	0.293	–	–	–	–	–	–	–	0.108	0.108
	(0.002)	(0.006)									
(2) Δy	0.116	0.296	0.004	0.030	0.081	-0.013	0.001	-0.000	-0.025	0.130	0.129
	(0.035)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.000)	(0.000)	(0.004)		

Notes: Each row reports a pooled OLS regression for the change in log wages $\Delta y_i = y_{i0} - y_{i,-1}$ between period -1 and period 0. The regressor Δow is the change in coworkers' mean log wage ($\text{owage2}_i - \text{owage1}_i$). Ed9, Ed12, and Ed16 denote education dummies for 9, 12, and 16 years of schooling; the omitted category is 6 years. *exp*, exp^2, and exp^3 are a cubic polynomial in experience at period -1. *Female* is an indicator for women. Standard errors are in parentheses. R^2 is the coefficient of determination; Adj. R^2 is the adjusted R^2. Sample size is $N = 16{,}969$ in both specifications.

Table 4 shows that controlling for unobserved characteristics of people yields that changes in individual wages are strongly linked to changes in coworker wages across all specifications. In the first pooled model, a one log point rise in coworker wages raises an individual's log wage by about 0.29 points, while women's wage growth is almost 2% lower than that of men. In our following model, we add controls for experience and notice that the effect of coworker's wages remain about the same suggesting that differences in experience barely contribute to explaining coworker wage effect. In the third model, which includes an interaction between gender and coworker wage growth, the coworker wage effect for men is approximately 0.3321 log points while the interaction term having a coefficient of -0.114 means that for women, the coworker wage effect is roughly 0.22 log points.

We confirm this observation on conducting separate regressions by gender where we observe that the coefficient on coworker-wage growth is 0.332, while for women it is 0.218. This indicates that men's wages are much more responsive to movements in the wages of their co workers. On comparing to the OLS levels results form table 3, where the coworker wage coefficients were around 0.66 for men and about 0.60 for women; we interpret that about 50% (0.503) of the coworker wage coefficient for men and about 36% (0.364) of the coworker wage coefficient for women persist even after differencing out the unobserved individual factors. This difference suggests that about half of the relationship seen in the OLS results comes from sorting or unobserved traits. It is only the other half that reflects a real causal effect, which suggests that when people move to jobs with higher paid coworkers, usually their wages increase as well. The comparatively smaller effect for women indicates that women's wages respond less to changes in coworker pay, likely due to unobserved barriers that women face at the workplace. Overall, coworker wages directly affect wage growth, and men benefit more from these workplace effects than women in terms of wages.

Section 5 (Bonus): Shrinkage

Taken together, the evidence from table 3 and table 4 suggests a mixed but informative picture. Sorting on fixed traits clearly amplifies the raw relationship between coworker wages and individual wages, yet there remains a sizeable component on the order of one(half of the original effect) that is consistent with workers' pay responding when they move into higher or lower wage coworker environments. This raises the next question: how stable is this estimated coworker wage effect once we allow for more flexible heterogeneity across experience groups and guard against overfitting in richer specifications?

To further examine how robust the relationship is between coworker wage changes and workers' own wage adjustments, we estimate three shrinkage versions of the first–difference model: OLS, Ridge, and Lasso. All three estimators use the same design matrix, which includes the change in coworker wages Δow_i, a cubic polynomial in experience, and education dummies. Ridge and Lasso apply penalties that shrink the coefficients toward zero, with Lasso additionally performing variable selection.

Table 5: Shrinkage of First–Difference Coefficients: OLS, Ridge, and Lasso

	Intercept	Δow	exp	exp^2	exp^3	Ed9	Ed12	Ed16
OLS	0.099	0.295	-0.011	0.000	0.000	0.004	0.029	0.079
Ridge	0.099	0.293	-0.011	0.000	0.000	0.004	0.028	0.078
Lasso	0.106	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Entries are coefficients from first–difference regressions of the change in log wages $\Delta y_i = y_{i0} - y_{i,-1}$ on the change in coworkers' mean log wages $\Delta\text{ow}_i = \text{owage2}_i - \text{owage1}_i$ and controls. All three estimators use the same design matrix: Δow_i, a cubic polynomial in experience at period -1 (exp, exp^2, exp^3), and education dummies for 9, 12, and 16 years of schooling (Ed9, Ed12, Ed16), with 6 years as the omitted category. Ridge regression and Lasso are estimated with 5 fold cross validation. Coefficients are rounded to three decimal places; very small values are reported as 0.000. Standard errors are not reported for the penalized estimators.

The Ridge estimates in Table 5 lie close to their OLS counterparts. For example, the OLS coefficient on Δow_i is 0.295, and Ridge shrinks it only slightly to 0.293. The experience polynomial terms and education dummy coefficients also move only marginally. This pattern indicates that multicollinearity among the regressors is limited. Namely, Ridge has little "work" to do, and the first–difference estimates appear numerically stable. Economically,

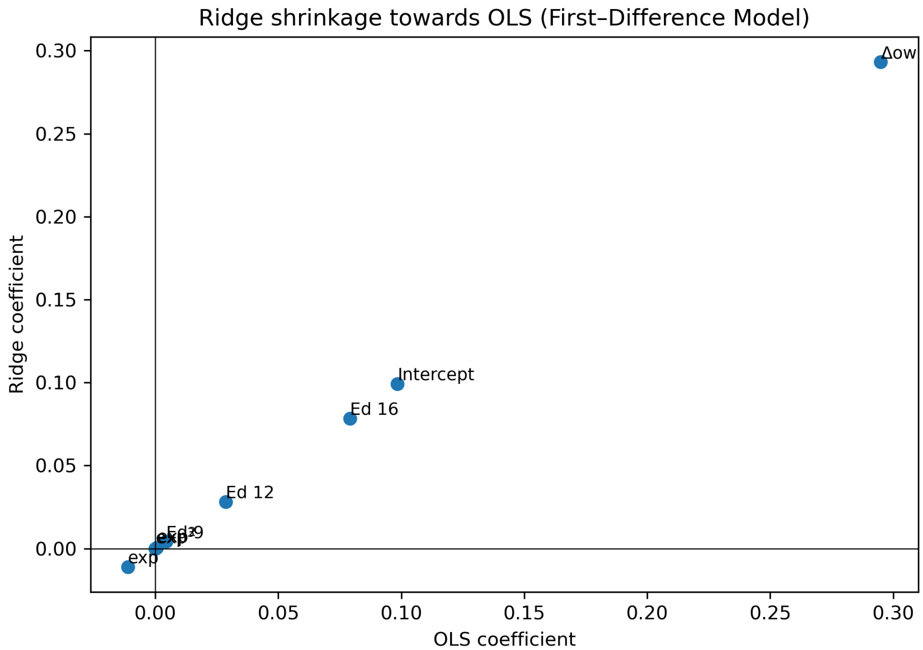


Figure 4: Ridge shrinkage relative to OLS. Each point plots the OLS coefficient (x–axis) against the Ridge coefficient (y–axis). Ridge estimates lie very close to the 45° line, indicating minimal shrinkage.

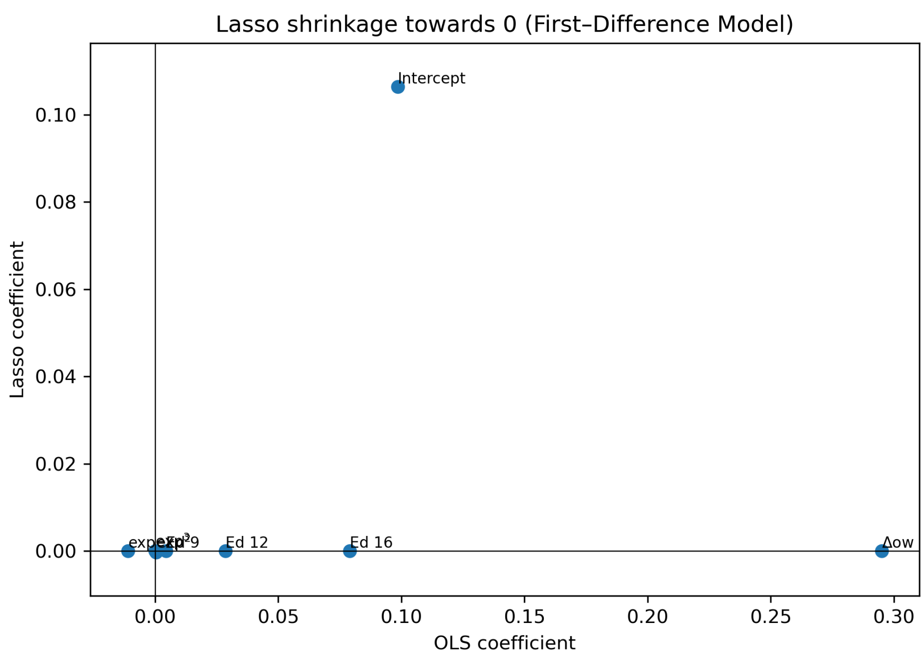


Figure 5: Lasso shrinkage relative to OLS. Almost all coefficients collapse to zero under Lasso penalization, illustrating aggressive variable selection.

this suggests that the estimated causal response of wage changes to coworker wage changes is not being driven by unstable or highly collinear predictors.

The Lasso estimator, by contrast, shrinks nearly all slope coefficients to exactly zero. This model places much stronger penalties on the regressors. In particular, the coefficient on Δow_i is driven fully to zero, as are the experience polynomial terms and all education dummies. Only the intercept remains non-zero. This outcome reflects the Lasso's variable selection property: once complexity is penalized, the model concludes that none of the included regressors explain sufficient variation in Δy_i to justify a nonzero coefficient. Statistically, this shows that the signal in the covariates is weak relative to the noise once differencing removes person specific heterogeneity. Economically, this suggests that the predictive content of coworker wage changes for individual wage changes is fragile and sensitive to model penalization.

Figure 4 plots the OLS coefficients on the x-axis against the Ridge coefficients on the y-axis. The tight clustering of points around the "$y = x$" line confirms that Ridge estimates only minimally adjust the OLS results. In contrast, Figure 5 plots OLS coefficients against Lasso coefficients. Here, almost all points collapse to zero on the y-axis, illustrating aggressive shrinkage toward sparsity. Notably, the Lasso sets the Δow_i coefficient to exactly zero, in stark contrast to both OLS and Ridge.

Together, the shrinkage results paint a nuanced picture. Ridge suggests that the OLS first difference estimates, which already difference out time invariant worker heterogeneity, are numerically very stable, so the modest positive association between coworker wage changes and own wage changes does not appear to be an artifact of multicollinearity. At the same time, Lasso's aggressive shrinkage towards zero is a reminder that this effect is not very large: once we insist on a very sparse model, the change in coworker wages is no longer as important a predictor. Our shrinkage specification uses a flexible polynomial in experience but does not include interactions with experience or gender, so these results should be interpreted as average effects. Combined with the event study and first difference evidence in Section 4: Event Study - Wage Changes around Moves (where we saw larger responses for movers with less experience and for men relative to women), we read the overall pattern as indicating that coworker wages matter for wage growth, but the magnitude of the effect is modest and any heterogeneity by experience or gender is limited and sensitive to how tightly the model is penalized. Overall, the bonus shrinkage exercise reassures us that the baseline first difference estimates are not driven by unstable regressors, while also highlighting the limits of what our data can say about strong, heterogeneous coworker wage effects.

Conclusion

Our analysis began with a descriptive comparison of the men and women in the sample. It revealed a substantial raw gender wage gap: women earn distinctly lower wages, work with lower-paid coworkers, and obtain higher levels of education. These patterns motivated a decomposition of the wage gap into components attributable to observable characteristics and components arising from differential returns.

Standard wage models and the Oaxaca decomposition showed that differences in education and experience explain relatively little of the gender wage gap. Despite women having slightly higher formal schooling on average, men receive systematically higher returns to education and experience. Most of the gap is therefore "unexplained" in the Oaxaca sense. This early finding established a consistent theme that: the gap is driven less by who men and women are, and more by how the labor market rewards them.

Introducing coworker wages into the analysis added an important new dimension. We found that the average wage of one's coworkers is a powerful predictor of individual wages, and that men both sort into higher–wage coworker environments and receive larger returns to these environments. Yet even after accounting for these differences in exposure, coworker wages have a large effect on individual wages, but explain only a small share of the overall wage gap. Most of the gap remains tied to differences in returns, reinforcing the conclusion that workplace environments are not gender–neutral.

The event–study evidence allowed us to explore the mechanisms behind these patterns. Workers who move into higher–wage coworker environments experience immediate and persistent increases in their own wages, while downward moves produce smaller or negative changes. These discontinuities are hard to reconcile with a pure ability–sorting model, since they reflect sharp shifts in coworker wage exposure at the moment of a job transition. Instead, the results support a mixed interpretation: access, networks, or search behavior affect who reaches high–wage workplaces, while genuine spillovers or institutional pay norms help sustain wage differences once workers arrive.

Our first–difference models strengthened this conclusion by differencing out time–invariant worker characteristics. Roughly half of the strong OLS association between coworker wages and own wages survives in the differenced specification. This suggests that sorting on fixed traits explains part, but not all of the relationship. Workers' wages do respond to changes in coworker wage environments, although the size of this response is moderate and heterogeneous.

Finally, our shrinkage exercises confirmed that the OLS and Ridge first–difference estimates are numerically stable, while the Lasso's aggressive shrinkage toward zero highlighted the fragility of the average coworker–wage effect once heavy penalization is imposed. Together, these results suggest that coworker wages matter for wage growth, but their influence is far from overwhelming and depends on model flexibility.

All in all, our findings have led to meaningful progress in understanding the gender wage gap. After this assignment, we are able to conclude that the gap is shaped by a combination of sorting into different workplace environments, differential returns to those environments, and deeper structural differences in how men's and women's characteristics are rewarded. Coworker wages help illuminate these mechanisms, but they do not eliminate the gap. Instead, they clarify that gender differences in wage setting arise not only from observable characteristics but also from workplace dynamics, institutional structures, and other unobserved factors that disproportionately benefit men.

Comprehensively, the evidence implies that the gender wage gap remains persistent even after controlling for education, experience, job transitions, and coworker environments. While our analysis cannot definitively isolate all underlying mechanisms, it demonstrates that both sorting and differential returns play central roles. The results imply that effective policy responses must address both gendered sorting across jobs and unequal returns within jobs, rather than focusing exclusively on education or labor-market attachment. Future work incorporating richer measures of job tasks, bargaining power, and organizational policies would provide an even more complete account of how gender shapes wages in the labor market.

Appendix

A. Additional Event Study Visualizations

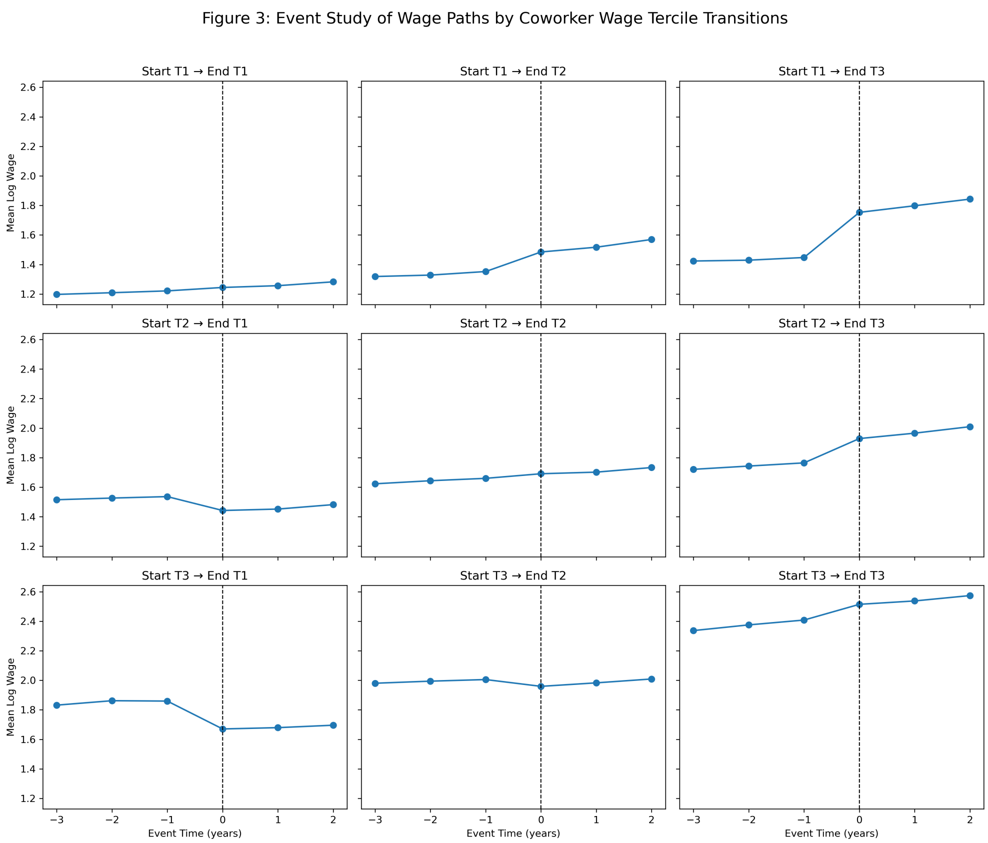


Figure 6: Nine–panel event-study visualization used during exploratory analysis.

References

Blau, F. D. and L. M. Kahn (2017, September). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature 55*(3), 789–865.

Brick, C., D. Schneider, and K. Harknett (2023, November). The Gender Wage Gap, Between-Firm Inequality, and Devaluation: Testing a New Hypothesis in the Service Sector. *Work and occupations 50*(4), 539–577.

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

Wu, A. (2025a). Final Project Analysis Code.

Wu, A. (2025b). Gender Wage Gap Dataset.