

Understanding Wage Growth: the Role of Coworkers*

Long Hong

University of Wisconsin - Madison

Salvatore Lattanzio

University of Cambridge

September 6, 2020

Abstract

This paper studies a critical but understudied driver of wage growth – coworkers. Using Italian employer-employee matched administrative data, we explore coworkers’ effect on wage growth in three directions. First, accounting for the endogenous sorting of workers into peer groups and firms, we estimate the impact of the average coworker’s quality on future wages. We find that a 10 percent rise in coworker’s quality increases one’s wage in the next year by 1.9 percent. The effect decreases gradually over the years and becomes less than half a percent after five years. Heterogeneous analysis suggests that the impact is the highest among new hires and young workers, whose learning from coworkers is essential. Second, we use propensity score matching and an event-study specification to study how a quasi-random entry of a high-performing worker, who could potentially transmit knowledge to other workers, would change the trajectory of her coworker’s wages. Compared to firms that hire an average worker, we find that when a high-quality worker enters, her coworkers’ wages increase by about 2 percent more. Both results suggest that learning could be the most likely channel through which coworker affects one’s future wages. Inspired by the reduced-form evidence, we plan to develop a structural model to quantify the importance of coworkers’ learning in generating future wages.

1 Introduction

The literature has well documented that wages typically increase over the life cycle. However, significant heterogeneity exists in wage growth among workers. In line with the canonical models in Becker (1964) and Ben-Porath (1967), wage growth reflects that workers accumulate knowledge and skills while working. As social interaction is essential in the workplace, it is natural that on-the-job learning is primarily the result of interaction with coworkers. Despite the importance, little is known on the link between coworkers and wage growth from both empirical and theoretical perspectives. If two workers have the same ability to learn, does the worker with better coworkers experience a faster wage growth over time? How much does the learning mechanism contribute to wage growth? How and what direction does knowledge transmission take place in the workplace?

*We are indebted to John Kennan, Rasmus Lentz, Kai Liu, Corina Mommaerts, Jeff Smith, Chris Taber, and seminar participants at the University of Wisconsin-Madison for valuable comments. We thank the Fondazione Rodolfo Debenedetti for its generous help in accessing the data.

What is the labor market value of this learning? Does the heterogeneity on wage growth caused by coworkers widen income inequality? This paper attempts to answer some of these questions.

There is a small but growing literature in understanding the relation between coworkers and wages. Earlier empirical evidence mainly focuses on the effect of coworkers on the contemporaneous wage level in a specific workplace (e.g., Mas and Moretti, 2009) or based on laboratory experiments (e.g., Falk and Ichino, 2006). For example, Mas and Moretti (2009) provide persuasive evidence in a supermarket chain that a cashier’s productivity increases when they work alongside more productive coworkers. Nevertheless, it is unclear to what extent their findings, based on a specific firm or laboratory experiment, apply to the general local labor market. With increasing access to administrative data, researchers also investigate coworkers in one particular local labor market. For example, Cornelissen et al. (2017) use German employer-employee matched administrative data to study the overall impact of coworkers on Munich’s contemporaneous wage level. They find, surprisingly, there is only a small positive effect. However, if knowledge transmission takes time to be reflected in wages, it would be appropriate to consider dynamic effects. That is, the impact of coworkers could materialize on future wage *growth* rather than the current wage level. Despite the potential relevance, only a handful of papers have examined the link between coworkers and wage growth. Two related and complementary working papers, Jarosch et al. (2019) and Herkenhoff et al. (2018), both find a substantial knowledge spillover from coworkers that facilitate wage growth, using data from Germany and the United States, respectively.

This paper aims to explore the effect of coworkers on wage growth in three directions. First, we explore the overall impact of coworker’s quality on one’s future wages. Building on the canonical AKM model (Abowd et al., 1999), our econometric strategy helps circumvent the common reflection problem and account for workers’ endogenous sorting into peer groups and firms. We adopt the novel estimation strategy developed by Hong and Sølvsten (2020) to overcome the estimation challenge induced by the high-dimensional fixed effects. Specifically, we exploit two sources of variation to identify the peer effects: changes in peer quality for workers who switch peer groups, and changes in peer quality for workers who remain with their peer group as other workers join or leave the peer group. Focusing on the later source, we also study how a quasi-random entry (or leave) of a high-performing worker, who could potentially transmit knowledge to other workers, would change the trajectory of her coworker’s average future wages in the same firm. We apply propensity score matching and an event-study specification for this analysis. In particular, we compare firms that hire a high-quality worker to comparable firms that employ a similar-quality worker, i.e., her estimated quality is higher than that of the firm’s average worker quality. Finally, we also plan to develop an equilibrium model with search friction, which builds on Postel-Vinay and Robin (2002) and Bagger et al. (2014). The structural model quantifies how much of the wage growth is through the mechanism of learning from coworkers. It estimates how the heterogeneity in learning from coworkers contributes to the aggregate income inequality. This last part is still a work in progress.

We use the matched employer-employee administrative database in Veneto – the Veneto Worker History dataset – from 1975 to 2001, a subset of the Italian universal INPS data. The data covers the entire working history if a worker has worked in Veneto for at least one day. It also tracks all the corresponding private firms where each worker has worked, including the firms outside Veneto. The data is an excellent fit for the study for the following reasons. First, it contains the entire working population in Veneto, one of Italy’s largest region for an extended period. Second,

we observe every coworker of a worker during his entire working history. Finally, the wage is based on accurate administrative records, which is not top coded.

The preliminary results of the analysis suggest that coworker plays a notable role in generating future wages. The AKM analysis shows that a ten percent increase in average peer quality (measured by worker fixed effects) could increase the next year's wage by around 2 percent, which is similar to the size of the return to college (Nix, 2020). The effects gradually fade out to about half a percent after five years. Moreover, the results are much more extensive for young workers and junior workers. The finding is consistent with a learning story: As a junior worker enters a firm, the better coworkers she has, the more likely she would learn quicker and boost future wages. The event-study analysis shows that a firm hiring a high-performing worker, who could teach the rest of the firm's workers, sees more than two percent higher wage growth than another arguably-identical firm hiring an average worker. On the other hand, a firm employing a worse worker (than the firm's average worker) does not see such a pattern. Both of the analyses above suggest that learning is a critical channel of wage growth. However, the reduced-form evidence only provides somewhat qualitative results. Therefore, we plan to develop a structural model that *quantifies* how much wage growth is through the mechanism of learning from coworkers.

2 Literature Review

Our paper lies at the crossing of three strands of literature. First, it contributes to the literature on the relationship between coworkers and contemporaneous wage levels. The literature typically divides into two approaches. One of them uses field experiments to study the coworker's effect in a specific workplace (e.g., Mas and Moretti, 2009; Papay et al., 2020; Brune et al., 2020). A typical concern of such works is from the external validity that they might not hold in general. With increasing access to the administrative data, the others study the coworker's effect on wage levels in a specific labor market (e.g., Cornelissen et al., 2017; Lengermann, 2002; Battisti, 2017). In general, the previous literature shows that the coworker's quality positively affects the current wage levels. Many suggest that learning and knowledge spillover is one of the critical channels. However, the paper does not consider dynamic effects, which could take some time to materialize. Indeed, if learning from coworkers takes time to be reflected in wages, one shall also pay attention to its effect on future wages. Despite the relevance, there is little evidence in the literature from theoretical and empirical perspectives, and this is where our paper aims to fill in.

Our paper contributes to a small but growing literature on studying the relation between coworker and wage *growth*. In particular, there are a handful of complementary papers (e.g., Jarosch et al., 2019; Herkenhoff et al., 2018; Nix, 2020) in this direction using different methodologies and data sources. All of them find that learning and knowledge spillover play an essential role in generating wage growth. However, the three papers use quite a different approach to measure the worker's quality. In particular, the previous literature uses unobserved permanent worker productivity or the worker fixed effects to measure worker quality. The three papers use only observables. For example, Jarosch et al. (2019) and Herkenhoff et al. (2018) use wages to measure workers' human capital or ability. Such a measure might cause some bias because, for instance, if wage growth is purely due to the firm's effect, then a worker with no human capital improvement can still see

wage growth.¹ On the other hand, our paper follows the standard practice and uses the estimated unobserved permanent worker productivity to measure worker's quality.

Methodologically, our paper fits into a broad literature related to Abowd et al. (1999) and a small one on estimating peer effect using high-dimensional worker fixed effects. For example, Mas and Moretti (2009) uses a two-stage estimate that requires strong assumptions to ensure consistency. Arcidiacono et al. (2012) use an iterative method that ensures consistency under moderate assumptions. However, the accuracy of the estimation and computational time for convergence is highly dependent on the chosen tolerance level. We apply the method by Hong and Sølvssten (2020), who impose minimal assumptions and develop relatively fast algorithms.

3 Theoretical framework

We motivate our empirical analysis by developing a simple two-period principal-agent model. Peer effects affect the worker's future productivity and future wages through knowledge spillover or learning.

3.1 Basic Model

There are two periods in the model. In each period, a firm hires N workers. In the theoretical model, we do not consider the endogenous sorting of workers into firms, which our empirical analysis takes into account. Besides, we assume that the same workers stay in the firm for the two periods. Each worker i chooses their efforts to produce outputs and the firm can only observe the outputs to reward the worker with a wage contract. To simplify the model, we follow the typical literature using a linear wage contract, which the firm would choose.

Worker's problem

Each worker is endowed with ability a , which is invariant over time.

Production function. The production function in each period is as follows.

$$\begin{aligned} q_1 &= a + e_1 + L_1(e_1, \bar{a}_{-i}) + \varepsilon_1, \\ q_2 &= a + e_2 + L_2(e_1, \bar{a}_{-i}) + \delta L_1(e_1, \bar{a}_{-i}) + \varepsilon_2, \end{aligned}$$

where $t \in \{1, 2\}$, q is the output observed to the firm, e is the effort and ε is a random productivity shock with a mean zero. Both e and ε is unobserved to the firm. During each period, a worker could also learn from the coworkers. The learning, denoted by L , depends on their efforts and the peer's quality, measured by the peers' average ability. Learning is not only concurrent - it also affects future productivity with a depreciation rate δ . For simplification, we assume that the learning function follows a straightforward expression below.

$$L(e, \bar{a}_{-i}) = e\lambda\bar{a}_{-i}.$$

¹It might be hard for the complementary papers to estimate the worker's fixed effect since the data is not panel data - they use a small random sample of the working population. As described below, our data use the entire working population and record each worker's full working history.

If there is no learning within the firm, i.e., $\lambda = 0$, it goes to be a simplified two-period principal-agent model, as in Rogerson (1985).

Cost function. The effort will induce some disutility characterized by a standard cost function

$$c(e) = ke^2,$$

as it is increasingly costly for a worker to extract each additional effort.

Utility function. For simplification, we assume (i) the period utility of a worker is linear (risk-neutral); (ii) preference is time-separable. The total utility across two-period is as follows.

$$U = \sum_{t=1,2} [w(q_t) - c_t]$$

where the wage contract $w(\cdot)$ is determined by the firm (see below).

Firm's problem

Assume the firm has a linear contract as shown below

$$\begin{aligned} w_1 &= \alpha + \beta q_1 \\ w_2 &= \alpha + \beta q_2 + \theta q_1 \end{aligned}$$

The firm chooses β to reward the concurrently observed outputs from the worker, and choose θ to reward the first-period effort. In other words, the firm chooses the slope β and θ to maximize profit, given IC and IR holds.

The optimal solution and its implications

Since we have not specified the cardinality of α , it is handy to manipulate α so that the IC holds, and we assume that the interior solutions to worker's maximization problem, i.e., IR holds.

$$\frac{\partial \mathbb{E}[w_2]}{\partial a_{-i}} = \beta^* \cdot \left(\frac{\partial q_2}{\partial e_2^*} \frac{\partial e_2^*}{\partial a_{-i}} + \frac{\partial q_2}{\partial e_1^*} \frac{\partial e_1^*}{\partial a_{-i}} \right) + \theta^* \cdot \frac{\partial q_1}{\partial e_1^*} \frac{\partial e_1^*}{\partial a_{-i}}$$

where apparently all terms are positive. Therefore, $\frac{\partial \mathbb{E}[w_2]}{\partial a_{-i}}$ is positive. Thus, the model suggests that an increase in coworkers' quality could lead to wage growth in the future.²

²The above analysis uses linearity as a simplification assumption, which does not seem to be realistic. We might prove the finding by relaxing the linearity assumptions. Nevertheless, our numeric simulation finds the results still hold if we use non-linear wage contracts and utility functions.

4 Data

We use the social security administration data that contains the entire working population and private firms in the region of Veneto in Northern Italy³ – the Veneto Worker History (VWH) dataset – from 1975 to 2001. We can observe every coworker of each worker over their working life within Veneto. The database contains three types of administrative datasets: (1) a worker-level demographic register, (2) a firm-level record, and (3) an annual firm-worker social security contribution register. A brief description of the database is as follows.

1. The worker register tracks over 3 million workers from 1975 to 2001. It records the entire working history of a worker in the private sector, as long as he/she worked one day in Veneto.⁴ It contains basic demographic information, including birth year and place, gender, nationality.
2. The firm register contains all private firms that employ each worker in the worker register.⁵ It includes a firm's detailed information such as national tax code, address, start and closure dates, industry. We also use the identified national tax code to link the data to AIDA, an internal firm-level balance-sheet data.⁶
3. The last register links the firm and worker registers. A private firm has to report the payment to its workers and the corresponding labor contract to the National Institute of Social Security (INPS) so that the authority could calculate each worker's social security contribution. Therefore, the register contains accurate information on wages (with no top-coding), weeks worked, occupation, etc. The wage has also been inflation-adjusted to the price level of the year 2003.

4.1 Sample selection

In this paper, we use all the workers and firms within the Veneto region only. In other words, we use the working population data within the Veneto labor markets. We only use the period from 1982 to 2001 because the information on working weeks before 1982 is not accurate (Battisti, 2017). Besides, we have a few minimal restrictions, mainly following the standard practice in the literature. First, we keep only a worker's primary job if he works in multiple positions, and we restrict the working ages from 16 to 65. Also, we exclude part-time jobs and apprentices because

³Veneto is the fifth largest region in terms of population and the third most prosperous region in terms of GDP in Italy.

⁴To be more precise, we could observe the working history before and after he/she worked in Veneto as long as it is within Italy. Besides, around 80 percent of the workers have never worked outside Veneto.

⁵There are two important related points. First, the public sector is not included in this database. We think it is important to exclude the public sector for our analysis, primarily because their pay scheme and wage determination structure are different from Italy's private sector. Second, the firm is not at the establishment level. It might be ideal to use establishment-level data for our analysis, but using firm-level data would not make a difference for two reasons. First, most firms, especially in our sample period where the franchise is not typical, are single-establishment firms. Second, the firm size is typically small, with a median size of five workers, and firms with 200 or fewer employees take up around 90 percent of the observations.

⁶AIDA is typically available to download in most Italian business schools. We thank Bocconi University for the generous access.

their wages cannot be meaningful compared to regular full-time employment. Since we are interested in coworkers, we drop single-worker firms. Following the practice of Cornelissen et al. (2017) and Caldwell and Harmon (2019), we also restrict the firm size to be smaller than 5000.⁷ The final dataset contains 1.8 million workers and 110 thousand firms. Each worker is observed for 8 years on average.⁸

Table 1 presents a brief summary statistics of the sample. The workers' average age is 34.5 years old, and the average tenure is around 3.5 years. The average weekly earnings are about 750 euros, which is higher than the national level. As mentioned earlier, the firm size is typically small, with a median of 5 workers.⁹ The share of female workers is 40 percent, and around 70 percent of the jobs are blue-collar jobs. 56 percent of jobs are in the manufacturing sector.

Table 1: Summary statistics

	Median	Mean
Workers		
Age	32.0	34.5
Tenure	3.0	3.5
Weekly earnings	652	744
Firms		
Firm size	5.0	14.6
Labor market		
Share female	40%	
Share blue-collar	69%	
Share manufacture	56%	

4.2 Peer group definition

We define the peer group as all the workers employed in the same firm with the same occupation in a given year. Unfortunately, we do not have a precise definition of occupation in the Veneto database. Therefore, we use the detailed industry code (5-digit Ateco 91) and the professional level (i.e., white-collar, blue-collar, and manager) to infer the occupation. The process gives us 330 different occupations, similar to the number of occupations used in Cornelissen et al. (2017). Admittedly, it is essential to have a robustness check on using different peer definition, which is still working in progress. For example, we could choose all the workers employed in the same firm as a peer group. It is especially plausible for small firms since social interaction is easy even for different occupations within a small firm.

⁷Lastly, due to the identification requirement in the AKM analysis below, we need to restrict the sample to the largest connected set (Abowd et al., 1999), which takes up around 97 percent of the sample.

⁸The average number of coworkers is around 270, but the median number of coworkers is only 37.

⁹The median firm size is calculated after we collapse the sample to the firm level.

5 AKM approach

In this section, we explore the overall effect of coworkers on future wages. In particular, we build our empirical strategy on the canonical AKM model (Abowd et al., 1999), by incorporating the average peer quality and additional fixed effects to better deal with the sorting issue. Besides, we briefly discuss how we adopt the novel method developed by Hong and Sølvesten (2020) for estimation.

5.1 Empirical strategy

In our regression specification, we follow Cornelissen et al. (2017), which builds on Abowd et al. (1999), as expressed in Equation 1:

$$w_{i,t+h} = \alpha_i + \beta \bar{\alpha}_{-i,t} + \mathbf{x}'_{it} \gamma + \psi_{jt} + \eta_{ot} + \theta_{oj} + \varepsilon_{it}, \quad (1)$$

where

$$\bar{\alpha}_{-i,t} = \frac{1}{|M_{-it}|} \sum_{k \in M_{-it}} \alpha_k, \quad \text{where } M_{-it} = \{k : oj(k,t) = oj(i,t), k \neq i\}$$

where w_{it} is the log weekly earnings at time $t + h$, where $h \geq 0$. α_i is the worker fixed effect, which measures the quality or innate ability of a worker. $\alpha_{-i,t}$ is the average coworker's quality at time t .¹⁰ $\mathbf{x}_{i,t}$ is a set of individual time-varying characteristics, including age, age squared, tenure, tenure squared, and a dummy on whether tenure is larger than ten years. ψ_{jt} , η_{ot} , θ_{oj} are firm-year, occupation-year, firm-occupation fixed effects. β is our parameter of interest. It describes how contemporary coworker quality could change future wages.

A critical issue that prevents us from identifying the causal peer effect is sorting. For this reason, we include a rich set of fixed effects beyond those usually included in the canonical AKM model (Abowd et al., 1999).¹¹ The worker fixed effects account for the potential sorting of high-quality workers into high-quality peer groups. Time-varying firm and occupation fixed effects address the possibility that firms (or occupations) may attract better workers and raise wages simultaneously. Firm-by-occupation fixed effects can account for high-quality workers' potential sorting into firms or occupations that pay high wages.

There are two sources of variations for the identification of β . For job switchers, when they move to another firm, the peer quality changes. For job stayers, peer quality changes when other workers join or leave the peer group.

5.2 Estimation

There are two main difficulties for the estimation of β in Equation 1. The first one is that the worker fixed effect needs to be estimated, but at the same time, the average coworker quality is a function of the worker fixed effects. The other difficulty comes from the massive dimension of the fixed effects, making it hard to solve the system. We employ the novel estimation method developed by Hong and Sølvesten (2020), which is discussed below.

¹⁰ $|\cdot|$ defines the modulus of the coworker vector M_{-it} ; hence, it represents the number of coworkers.

¹¹ As a reassuring check, the correlation between worker fixed effects and firm-occupation fixed effects are very tiny

First, we write Equation 1 in a matrix form below.

$$w = X\delta + C\delta\beta + \varepsilon, \quad (2)$$

where $w \in \mathbb{R}^n$ is the wage data, $X \in \mathbb{R}^{n \times k}$ is all the fixed effects and time-varying individual characteristics. $C \in \mathbb{R}^{n \times k}$ is a coworker averaging matrix (see Appendix for a detailed construction of C). $\delta \in \mathbb{R}^k$ is a nuisance parameter that corresponds to the coefficients of the fixed effects. $\beta \in B$ is our parameter of interest, where B is a compact parameter space. Note that $C\delta$ is equivalent to $\bar{\alpha}_{-i,t}$ in Equation 1. In addition, we assume the following.

- exogeneity, $\mathbb{E}[\varepsilon|X, C] = 0$,
- homoskedasticity, $\mathbb{E}[\varepsilon\varepsilon'|X, C] = \sigma^2 I_n$ where $\sigma^2 > 0$ is unknown.
- the designed matrix $X + C\beta$ has full rank k for any $\beta \in B$.

Our parameter of interest β is estimated by solving the objective function Q_n , which is the solution by solving the inner minimization problem, as shown in 3

$$\hat{\beta} = \arg \min_{\beta \in B} Q_n(\beta) = \arg \min_{\beta \in B} \left\{ \min_{\delta \in \mathbb{R}^2} \|Y - X\delta - C\delta\beta\|^2/n \right\} \quad (3)$$

First, Hong and Sølvyten (2020) proves that β is the unique minimizer of the population analogue to Q_n using the assumptions we imposed, so the consistency is ensured. We derive the moment condition for β by taking the first order conditions of β and δ in Equation 3, as shown below.

$$S_n(\beta) = w' M C (R'R)^{-1} R' w / n = 0,$$

where

$$R = X + C\beta \text{ and } M = I_n - R (R'R)^{-1} R'$$

We utilize the sparse matrix operation and conjugate gradients method to speed up the computation. In particular, the sparse matrix helps with the standard matrix operation and memory issues, and the conjugate gradient method could vastly improve the matrix inversion speed.

We have applied the bootstrapping method to calculate the standard error, and the result shows that the standard error is tiny at a magnitude of 0.001. For that reason, we do not report the standard error in Section 5.3.

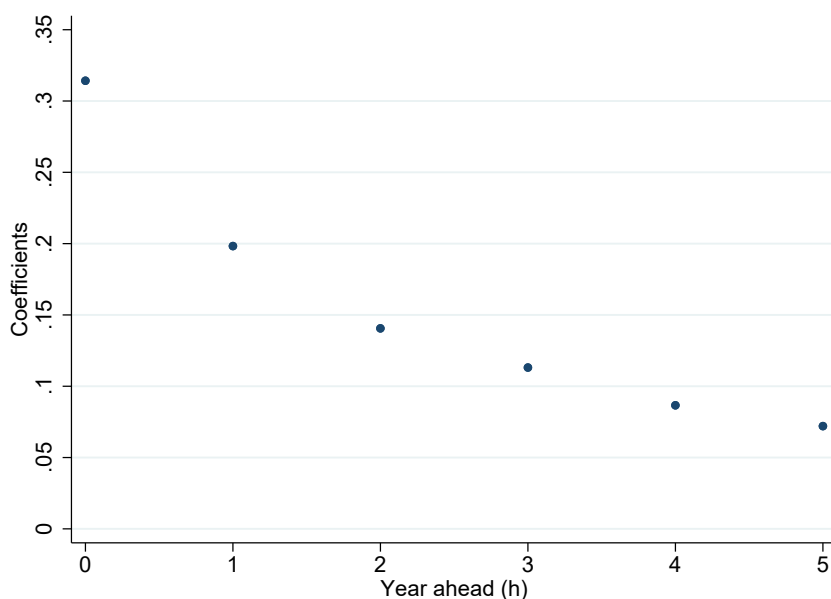
5.3 Results

Figure 1 shows our baseline results. Each dot in the graph represents the estimate β in Equation 1 using the future wages as the dependent variable in each year ahead (h), where $h \geq 0$.¹² The figure

¹²In the current analysis, different future wages are used as outcomes in separate estimations. In the cases when $h > 0$, workers who do not have wages in year $t + h$ are excluded. Therefore, each estimate, to some extent, is not consistently estimated. We have also conducted a robustness check using a sample restricted to workers who have been working continuously for at least five years. The results are very similar, except that the estimates in the first two years are slightly smaller. Besides, the decreasing pattern remains.

shows that the peer effect is large not only for the contemporaneous wage but also for the wages in the following years. A ten percent increase in peer quality could increase the next year’s wage by 2 percent, which is similar to the size of the return to college (Nix, 2020). On the other hand, the effects gradually fade out to around 0.5 percent after five years. Although unreported in the graph, the effect becomes small and flat after five years. It is consistent with other papers that coworkers in the past three years play the most important role in wages (e.g., Caldwell and Harmon, 2019; Nix, 2020). As mentioned above, the standard error is calculated via bootstrapping. Although not reported, all the estimates are statistically significant as the standard errors are very tiny.¹³

Figure 1: Effects of coworker’s quality on future wages (β)



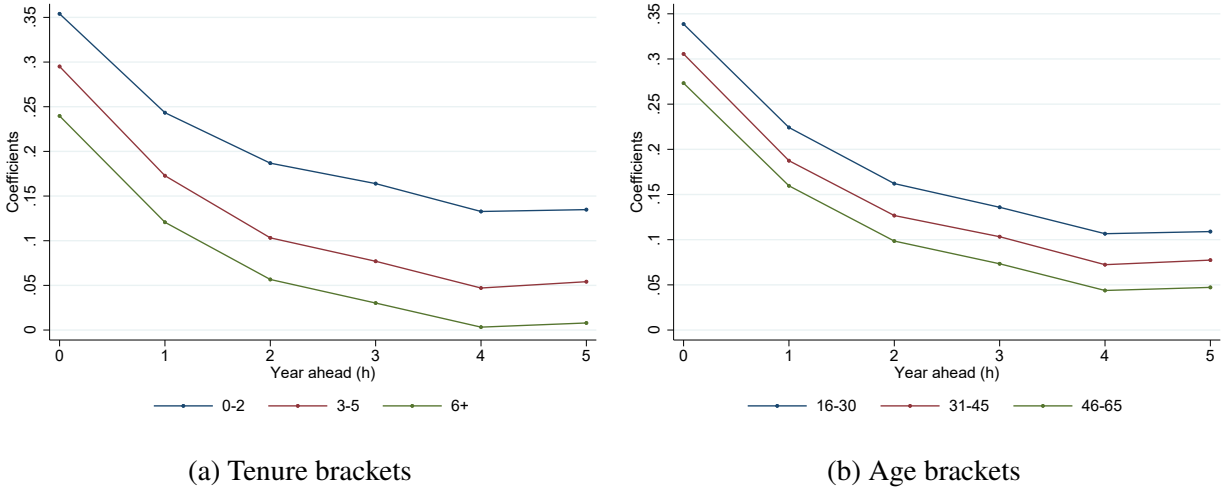
We have also explored the heterogeneous effects across different firm tenures and age groups.¹⁴ Figure 2a shows how the effects differ across different tenure years. The results illustrate a clear pattern that coworker matters the most for the new hires, while the effect decreases as one experiments more years in the same firm. The finding is consistent with a learning process: there is more room for a new hire to learn in a firm. If the new hire has better peers as she enters the firm, she will more likely learn quicker and thus boost future wages.

A similar pattern arises when we explore the heterogeneous impacts across different age groups, as shown in Figure 2b. Like the finding above, the effects are most potent for the youngest workers (below age 30). For older workers, the effect goes down gradually. Again, the same logic applies here. There is greater space for the younger worker to learn as they enter the labor market. A better peer group could help them accumulate human capital faster, thus expedite wage growth.

¹³Due to the tiny standard error, we do not report the standard errors for better visualization of the graphs.

¹⁴We conduct the heterogeneous analysis using the pre-estimated fixed effects from the baseline regression.

Figure 2: Heterogeneous effects across tenure and age brackets



6 Event-study evidence

As mentioned earlier, the sources of identification for β is through the following channels.

- for job switchers, the peer quality changes as they move to another firm
- for job stayers, the peer quality changes when a new worker *enters* the peer group or an old worker *leaves* the peer group.

Since the vast majority of the source is from the last two channels, we now provide an event-study analysis to have a future discussion on them.

In particular, we aim to explore the effect on the stayers' future wages when a high- or low-quality worker enters or leaves. We illustrate our empirical strategy in detail using the case when a new worker enters. The analysis of a worker's leave is methodologically symmetric.

6.1 Empirical Strategy

6.1.1 Event, sample selection, and treatment

Figure 3 illustrates our empirical strategy. We explain in detail below. Primarily, we define the event as a new worker who enters a firm and stays there for three years. We choose three years because we think that it takes time for the knowledge spillover to diffuse among his coworkers.¹⁵

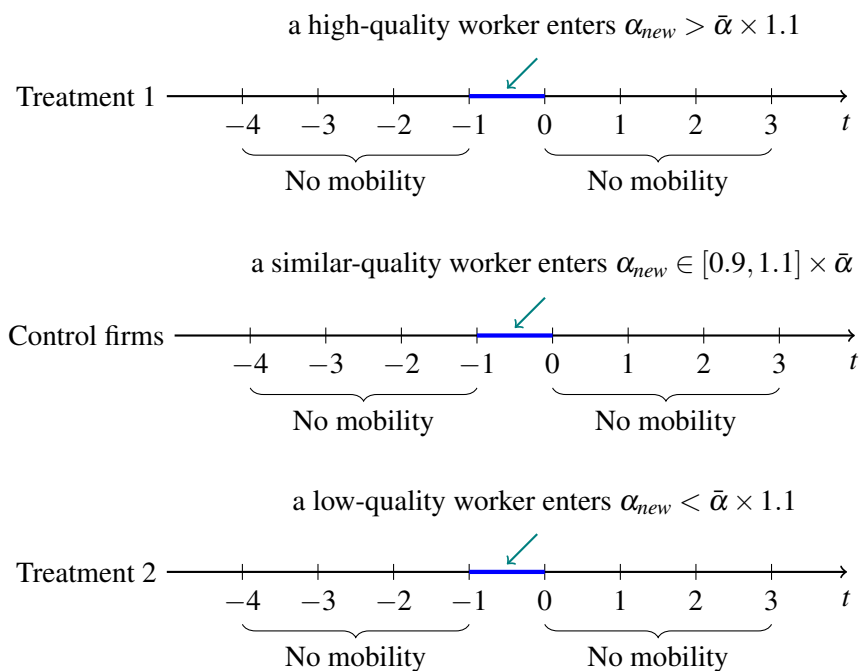
We restrict the sample firms to be observed for at least eight years, in which there is no worker mobility in the four years before and three years after the event. While we risk losing generalization by such a substantial restriction, we believe it is essential for our analysis. First, as mentioned

¹⁵When the event year has multiple workers entering, we exclude the firm from the analysis. Besides, we plan to test some alternative time frame choices, such as two or four years after a worker's entry, as a robustness check.

in Section 5.3, the coworker’s effect typically fades out after three years. After four years, the knowledge spillover within the firm is generally exhausted. When a new worker enters, it is very likely the learning or knowledge spillover is mainly through the new hire channel. Second, we need the pre-event period to examine the pre-event parallel trend assumption from a methodological perspective. Finally, since learning and knowledge spillover take time to be reflected in wages, we need a few years after the event to how the wage trajectory after the new worker enters.

We define the treatment groups as the workers in the firms that have hired a high- or low-quality worker in period $t = -1$. A high- or low-quality worker’s quality (measured by the worker’s fixed effect in Section 5) is 10 percent higher or lower than the firm’s average workers’ quality. We denote the group that hires a high-quality worker as treatment 1 in Figure 3 and the other as treatment 2. We define the control group as the workers in the firms that have hired an average quality worker, whose ability is similar (within 10 percent difference) to the workers in the firm.¹⁶

Figure 3: Treatment and control groups in the event-study analysis



6.1.2 Propensity score matching

A critical issue that prevents us from identifying the effect is that the worker’s entry is not random. For example, the decision to hire a high-quality worker might be endogenous to firm performance, which also affects a worker’s wage growth.¹⁷ While there is no perfect remedy for this, our current

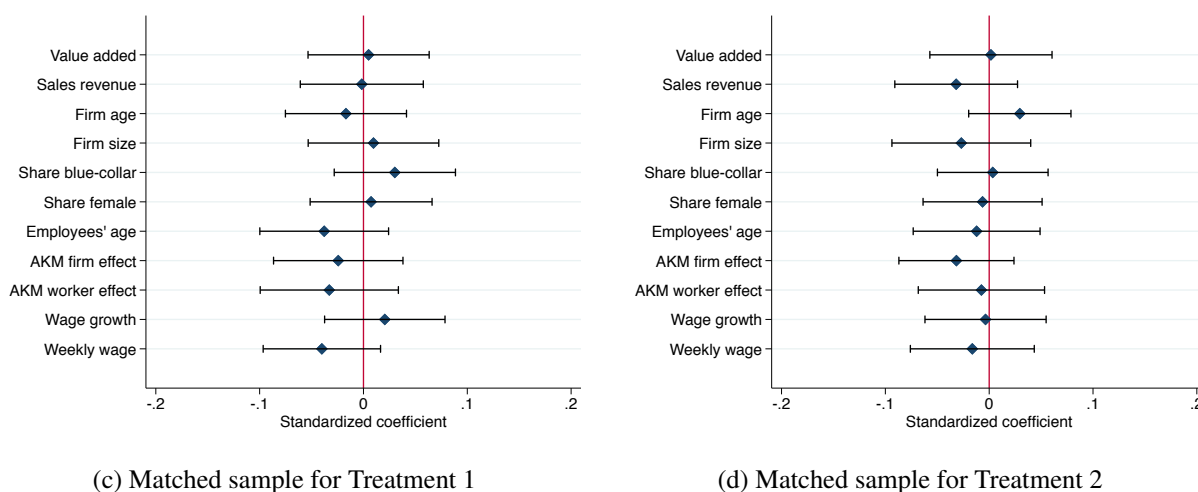
¹⁶In an earlier version of the paper, we use the firms with no worker mobility for the entire eight years as our control group. The finding is very similar to the one using the current control group.

¹⁷Anecdotally, a firm may decide to hire a high-quality worker because he is complementary to some technology the firm decided to invest. Such an investment could raise the productivity of the firm and eventually compensate all employees in wages.

best practice is to construct comparable firms between the treatment and control groups through propensity score matching. The implicit assumption is that similar firms have similar hiring strategies, leading to a quasi-random hiring on average.

We estimate the propensity score using a wide range of firm-level variables and some industrial and geographic variables.¹⁸ We use only the period $t = -3$ in Figure 3 for the matching process, leaving the periods $t = -4$ and $t = -2$ as “out-of-sample” check for our matching.¹⁹ Besides, we use the single nearest neighbor matching with replacement to match the treatment groups 1 and 2 with the control group, separately. In other words, two different matched control groups are respectively comparable to treatment groups 1 and 2. Figure 4 reports the mean differences in the covariates we have used for matching. It suggests the matched control groups are similar to the treatment groups 1 and 2, respectively, from all dimensions.²⁰

Figure 4: Balance test of covariates



In the final sample, 2,230 firms have hired a high-quality worker, and they are matched with 1,415 firms from the control group. 1,959 firms have hired a low-quality worker, and they are matched with 1,397 firms from the control group. The two matched samples consist of 198,720 and 178,972 work-year level observations, respectively.

¹⁸The detailed list of variables we have used is as follows: the AKM firm effects and AKM average worker effects estimated from Section 5, average weekly wages, growth of wages, the average age of employees, the share of female workers, the share of blue-collar workers, firm size, firm age, sales, value-added, 2-digit industry dummies, and province dummies. As a related note, the sales and valued-added variables are merged from the external balance-sheet firm-level database, AIDA. However, there is a good portion of firms that are not covered in AIDA. To utilize the information from balanced-sheet data, as a common practice in the empirical literature, we imputed the missing value as -999 and created a dummy to indicate the missing observations.

¹⁹As a robustness check, we have also used the observations in $t = -4$ only, $t = -2$ only, or all observations from $t = -2$ to $t = -4$ for matching. The results are very similar.

²⁰The same balanced tests of covariates for the unmatched groups are exhibited in Figure A.1. More, Figure A.2 shows that the common support assumption holds since there is sizeable overall lapping on the propensity score between the treatment groups and the control group.

6.1.3 Event-study specification

We use the following event-study specification to study the impact of a high- or low-quality worker’s entry on his coworkers’ future wages.

$$w_{i,t} = \eta_t + \psi_j + \sum_{k \neq -1} \beta_k (Treat_t \times \mathbf{1}\{t = k\}) + \mathbf{x}'_{i,t} \gamma + \varepsilon_{i,t},$$

where $w_{i,t}$ is the log weekly wage of workers, excluding the new hire, in period t ; η_t and ψ_j are year and firm fixed effects, respectively. $\mathbf{x}_{i,t}$ is the time-varying worker characteristics. ε_{jt} is an error term. $Treat_j$ is a dummy variable for treated firms. The coefficients of interest is β_k , which measures the differential impact of hiring a high- or low-quality worker relative to hiring a similar-quality worker on wages in each period k .

6.2 Results

6.2.1 When a high- or low-quality worker enters

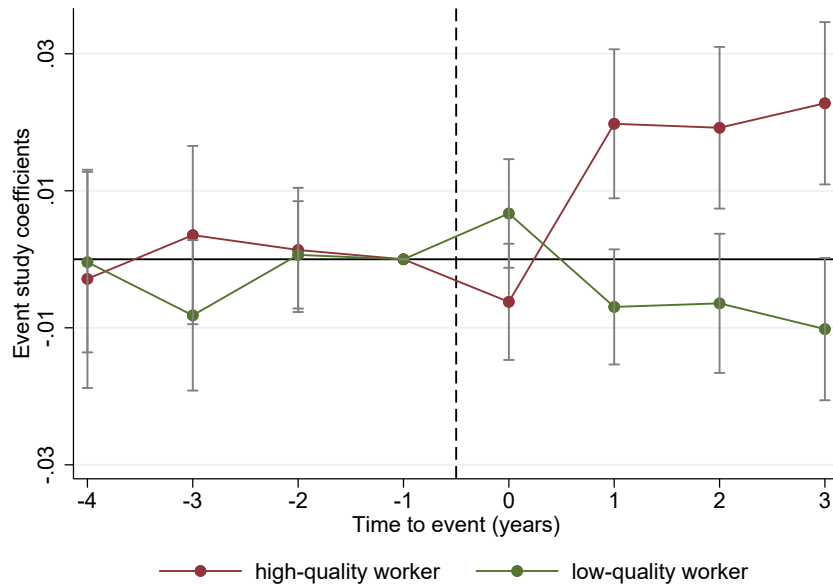
Figure 5 reports the event-study coefficients β_k for each $k \in \{-4, \dots, +3\}$, for both treatment groups. Specifically, the red line and green line are the effects of a high-quality and low-quality worker’s entry, respectively, on their coworker’s future wages. First, the pre-event parallel trend assumption holds as the effect before the event is small and statistically insignificant. As mentioned earlier, we only use period -3 for the propensity score matching. The fact that there is no effect for periods -4 and -2 is reassuring that the matching works well.

Second, the effects are quite different for the two treatments. Compared to the firms with a new hire of similar workers, the workers in firms, which hire a high-quality worker, see a positive and significant effect on future wages. One year after the high-quality worker’s entry, his coworker’s wage is around 2 percent more than that in control firms. The effect persists in the following years. Moreover, a notable observation is that there is no effect in period 0. That is, the high-quality worker’s entry does not impact the coworker’s wage immediately. It takes some time for the knowledge spillover to diffuse and be reflected in wages. On the other hand, when a firm hires a low-quality worker, the effect on his coworker’s future wages is slightly negative, and the impact is statistically insignificant. We believe that the learning story plays a role here. When a high-quality worker enters, he would be able to teach his coworkers, and therefore eventually drive up the wages in the following years. However, when a low-quality worker enters, there is no much to learn from him. Thus it is less relevant for their future wages.²¹

We also explore some heterogeneous effects across different peer groups. Recall that we have defined the peer group in Section 5 as all the workers employed in the same firm with the same occupation. Figure 6 shows the effect of a high-quality worker’s entry on his peer group and non-peer group, where the red line and green line are the effects on the peer group and non-peer group,

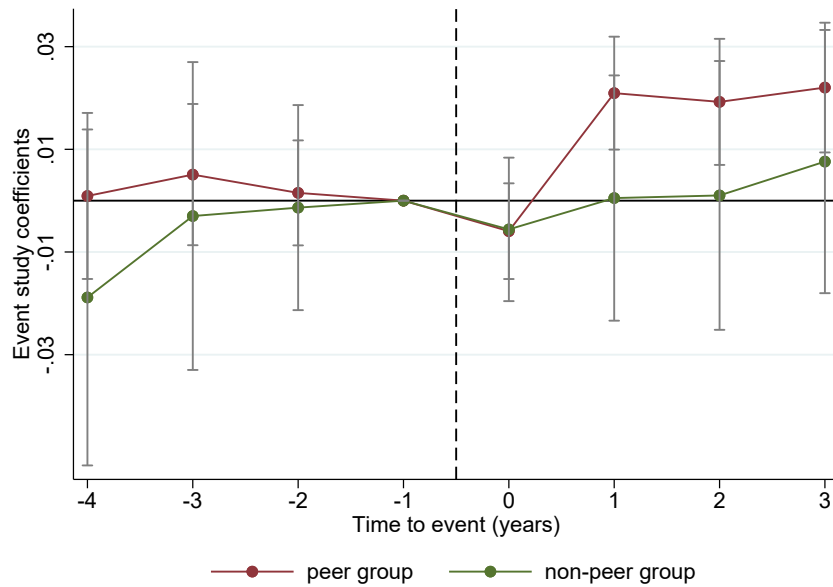
²¹The learning story can be the most likely one, but not necessarily the only one. The entry of a high-quality worker could affect coworkers’ wages through alternative channels. First, a high-ability worker can boost coworkers’ productivity through *peer pressure* (e.g., Mas and Moretti, 2009). For instance, the high-quality worker’s coworkers may feel obliged to put more effort because of the higher competition coming from their peers. Moreover, a high-quality worker is typically positively correlated with a larger connection to other firms (e.g., Caldwell and Harmon, 2019), which will help the firm expand business through the *network effect*. As a result, it increases the future wages of all workers in the firm. Separating all the possible channel is beyond the scope of the current version of the paper.

Figure 5: The effect of a high-/low- quality worker's entry on coworkers' future wages



respectively. For the peer group, the effect is almost identical to the one in Figure 5. On the other hand, there is no effect on the non-peer group. In other words, the knowledge spillover is mainly diffused among the same peer group.²²

Figure 6: Effect of a high-quality worker's entry on the future wages of peers and non-peers



²²Not surprisingly, the finding is reassurance on our definition of the peer group in Section 5

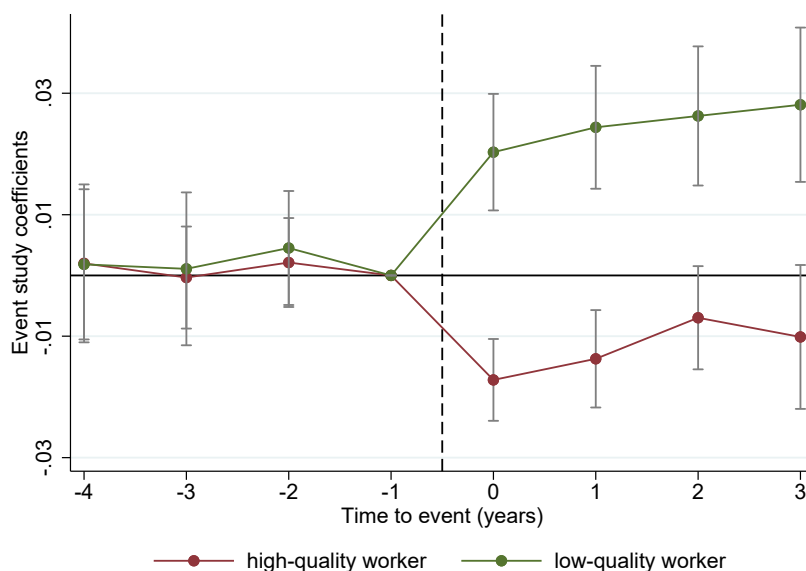
6.2.2 When a high- or low-quality worker leaves

Using the same empirical strategy described in Section 6.1, we also study the effect of a high- or low-quality worker's leave on his coworker's future wages. The only difference is that a worker's entry described above is now a worker's leave. So the implicit assumption is that conditional on comparable firms, the separation between the firm and the worker is as-if random. Figure 7 reports the event-study coefficients β_k for each $k \in \{-4, \dots, +3\}$ for both treatment groups. The red line and green line are the effects of a high-quality and low-quality worker's leave, respectively, on their coworker's future wages.

The interpretation of the effect is somewhat tricky and is quite different from the previous one. When a low-quality worker leaves, it increases the average peer quality in a firm. Therefore, it makes knowledge spillover in the firm more efficient, and eventually increases the wages in the future years.²³ On the other hand, when a high-quality worker leaves, there are potentially two (opposite) effects. First, the leave of a high-quality worker will make the overall peer quality smaller, thus decreasing knowledge spillover efficiency. Second, the high-quality worker's human capital has left to the firm might play a persistent effect over the next few years, which could help boost wage growth. Overall, the result suggests that the first channel might exceed the second, which negatively affects future wages when a high-quality worker leaves.

Moreover, by comparing the effect of high-quality workers in Figure 5 and Figure 7, we could, to some extent, tell a full story on a high-quality worker. When a firm hires a high-quality worker, he could teach the rest of the workers through knowledge spillover and thus increase his coworker's future wages by more than 2 percent. When he leaves the firm, there is an immediate decrease in wage growth due to the loss of his help. However, as the persistent effect of human capital that the high-quality worker has left to the firm, the decrease in wage growth fades out in the following years.

Figure 7: The effect of a high-/low-quality worker's leave on coworkers' future wages



²³In a robustness check, we find most of the effect comes from small firms.

7 Conclusion and future works

This paper explores an understudied component of the wage driver - coworker's quality. Our key finding suggests that the quality of coworkers plays a vital role in generating future wages. In the AKM analysis, we show that a 10 percent increase in coworker's quality raises the next year's wage by 2 percent, which is almost equivalent to the size of the return to college. Moreover, the heterogeneous analysis suggests that the effect is much more significant for the new hire and the young workers, simply because there is substantial room for them to learn from coworkers. In the event-study analysis, we find that if a firm hires a high-quality worker, the rest of the workers' wage trajectory is around 2 percent higher than that in firms hiring a mediocre worker. Both of the evidence suggests that learning is an essential mechanism through which coworker could impact future wages.

In the near future, we would conduct a more heterogeneous analysis in Section 5 and Section 6 to have a more thorough understanding of the mechanisms. Based on the reduced-form evidence, we aim to develop a structural model to quantify the learning channel and understand how much wage growth is due to the coworkers' knowledge spillover.

References

- Abowd, John M., Francis Kramarz, and David N. Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 1999, 67 (2), 251–333.
- Arcidiacono, Peter, Gigi Foster, Natalie Goodpaster, and Josh Kinsler, "Estimating spillovers using panel data, with an application to the classroom," *Quantitative Economics*, 2012, 3 (3), 421–470.
- Bagger, Jesper, François Fontaine, Fabien Postel-Vinay, and Jean-Marc Robin, "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics," *American Economic Review*, June 2014, 104 (6), 1551–96.
- Battisti, Michele, "High wage workers and high wage peers," *Labour Economics*, 2017, 46 (February 2015), 47–63.
- Becker, Gary, *Human Capital*, Chicago: Univ. Chicago Press, 1964.
- Ben-Porath, Yoram, "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 1967, 75 (4), 352–365.
- Brune, Lasse, Eric Chyn, and Jason Kerwin, "Peers and Motivation at Work," *Journal of Human Resources*, 2020, pp. 0919–10416R2.
- Caldwell, Sydnee and Nikolaj Harmon, "Outside Options, Bargaining, and Wages: Evidence from Coworker Networks," *Job Market Paper*, 2019.
- Cornelissen, Thomas, Christian Dustmann, and Uta Schönberg, "Peer effects in the workplace," *American Economic Review*, 2017, 107 (2), 425–456.
- Falk, Armin and Andrea Ichino, "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 2006, 24 (1), 39–57.

- Herkenhoff, Kyle, Jeremy Lise, Guido Menzio, and Gordon M Phillips, “Production and Learning in Teams,” *National Bureau of Economic Research Working Paper Series*, 2018, No. 25179.
- Hong, Long and Mikkel Sølvssten, “Estimation and inference in a peer effects models with heteroscedasticity,” *Work-in-progress*, 2020.
- Jarosch, Gregor, Ezra Oberfield, and Esteban Rossi-Hansberg, “Learning from Coworkers,” *National Bureau of Economic Research Working Paper Series*, 2019, No. 25418.
- Lengermann, Paul A., “Is it who you are where you work or with whom you work? Reassessing the relationship between skill segregation and wage inequality,” *LEHD Technical Paper*, 2002, 10.
- Mas, Alexandre and Enrico Moretti, “Peers at work,” *American Economic Review*, 2009, 99 (1), 112–145.
- Nix, Emily, “Learning Spillovers in the Firm,” *Job Market Paper*, 2020.
- Papay, John P., Eric S. Taylor, John H. Tyler, and Mary E. Laski, “Learning Job Skills from Colleagues at Work: Evidence from a Field Experiment Using Teacher Performance Data,” *American Economic Journal: Economic Policy*, February 2020, 12 (1), 359–88.
- Postel-Vinay, Fabien and Jean-Marc Robin, “Equilibrium Wage Dispersion with Worker and Employer Heterogeneity,” *Econometrica*, 2002, 70 (6), 2295–2350.

Appendix

A Construction of C

Denote i for each row of the observation, where we suppress the notation of i, t to i . First, let's construct a matrix to indicate the location of coworker. For example, for each i ,

$$c_i = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ \vdots \\ 1 \\ \vdots \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ \vdots \\ 0 \\ \vdots \end{pmatrix}$$

where i refers to the coworker's index location. The averaging matrix is constructed as follows.

$$\tilde{c}_i = \frac{c_i}{c_i' \cdot \mathbf{1}} \Rightarrow \tilde{C} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \Rightarrow C = \begin{pmatrix} \tilde{C} \\ 0 \end{pmatrix}$$

where the auxiliary matrix 0 makes the dimension of C the same as X .

A.1 A simple example

As a simple example on how we construct C , suppose we have the following data, where there are only five workers and two peer groups. The first column and second column of the data indicate the indices of worker and peer group, respectively.

$$\text{data} = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 2 \\ 4 & 2 \\ 5 & 2 \end{pmatrix}$$

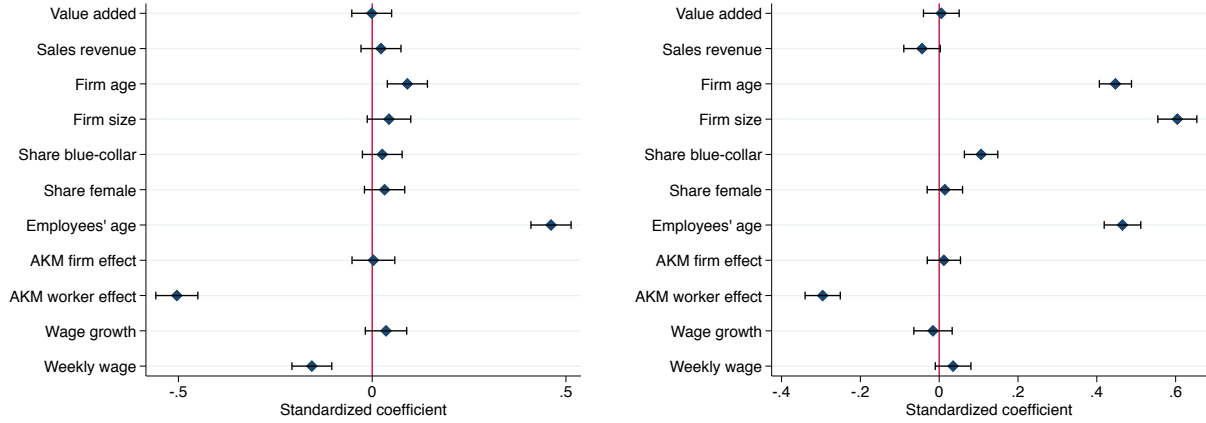
We first construct an averaging matrix \tilde{C} below to detect who is your peer and how much their weight is when calculating the average peer quality. One might read \tilde{C} as follows. The first row of \tilde{C} says: 1 is not a coworker of himself, 2 is his coworker, and 3, 4, 5 are not his coworkers. The third row says, 1 and 2 are not 3's coworkers, 3 is not a coworker of herself, but 4 and 5 are her coworkers. Both of them weight half when calculating the average coworker quality.

$$\tilde{C} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{pmatrix}$$

To make sure C and X have the same dimension, we augment an auxiliary matrix 0 to \tilde{C} as a final component of C . That is, $C = [\tilde{C}, 0]$.

B Propensity score matching

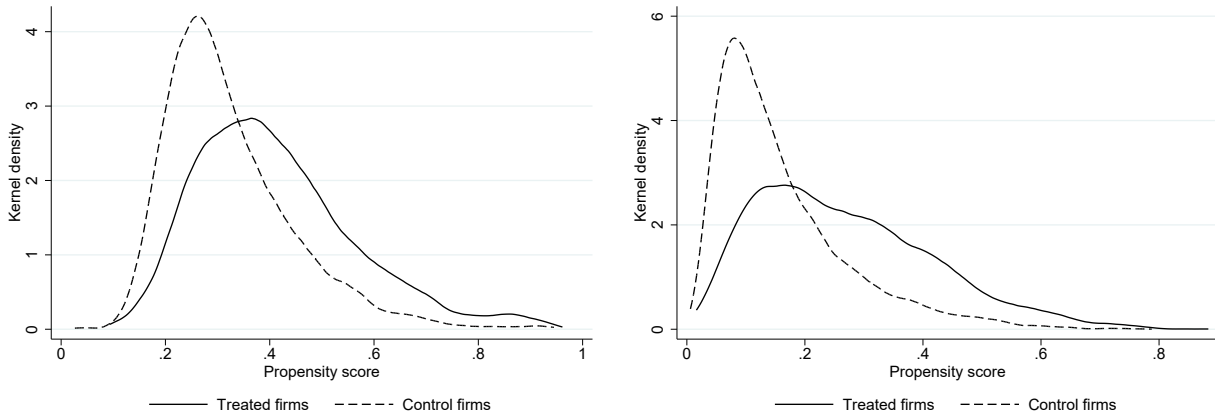
Figure A.1: Balance test of covariates



(a) Unmatched sample for Treatment 1

(b) Unmatched sample for Treatment 2

Figure A.2: Propensity score density



(a) Treatment 1 vs Control group

(b) Treatment 2 vs Control group