Coworker Sorting, Learning, and Inequality

Long Hong*

December 31, 2022

Click here for the latest version

Abstract

Social interaction with coworkers is common in the workplace. This paper explores how coworkers affect inequality through labor market sorting and on-the-job learning. Using matched employer-employee data from Italy, I first document two sets of empirical evidence by estimating an econometric model that incorporates coworkers in a wage regression with a novel estimation method. I find two main mechanisms through which coworkers affect wages: production complementarity and learning from coworkers. I also show that coworkers explain a substantial fraction of wage inequality, similar to that firm heterogeneity explains. To account for wage dynamics induced by these two channels and the subsequent impact on lifetime income inequality, I incorporate coworkers into a labor search model with worker and firm heterogeneity. I find that half of the lifetime income variation is explained by workers' initial ability. Firm heterogeneity explains around 15 percent of the remaining unexplained part, while coworker production complementarity and learning contribute to another 15 percent and 30 percent, respectively.

Keywords: coworkers, wage inequality, lifetime income inequality, peer effect model, labor search model, human capital, labor market sorting, high-dimensional econometrics.

*Department of Economics, University of Wisconsin - Madison (email: long.hong@wisc.edu). I am deeply indebted to my advisors Rasmus Lentz, Corina Mommaerts, and Chris Taber. I also benefit tremendously from discussions with Naoki Aizawa, Michéle Belot, Job Boerma, Natalie Duncombe, Itzik Fadlon, Kevin Hunt, John Kennan, Philipp Kircher, Salvatore Lattanzio, Maurizio Mazzocco, Ross Milton, Chris Moser, Xinyi Peng, Jeff Smith, Mikkel Sølvsten, Shilong Sun, Anson Zhou, and participants at SOLE, Minnesota-Wisconsin joint workshop, North American Summer Meeting of Econometric Society. I acknowledge Fondazione Rodolfo Debenedetti and Bocconi University for accessing the data and the financial support from the Department of Economics at the University of Wisconsin - Madison.

1 Introduction

It is documented that half of the inequality is determined by pre-market ability (e.g., Cunha and Heckman, 2007, 2008). What contributes to the unexplained component remains a question. There is increasing evidence that *where* a worker works plays a critical role (e.g., Abowd, Kramarz and Margolis, 1999; Postel–Vinay and Robin, 2002; Card, Heining and Kline, 2013), but it is still not enough to resolve the puzzle. This paper brings a new perspective and explores the question by studying an essential component of the labor market: coworkers.

Specifically, I explore two primary mechanisms through which coworkers affect wage dynamics and inequality. First, good coworkers bring synergy to work and boost a worker's own productivity and wages, creating a "production complementarity effect". Second, knowledge spillover from coworkers offers a "learning effect", accelerating human capital accumulation and wage growth. Both effects can have a substantial impact on inequality. Due to production complementarity, good workers tend to work with good coworkers, creating labor market sorting and wage dispersion. If two similar workers learn differently from distinct coworkers, it widens the human capital and wage gaps over their working life. Moreover, if a worker both works and learns from good coworkers, it can generate a snowball effect over the working cycle and create substantial lifetime income disparity even among workers who are similar in their early careers.

This paper explores the role of coworkers in inequality in two approaches. First, I empirically assess whether coworkers affect wage dynamics through the two mechanisms – production complementarity and learning – and whether coworkers can explain wage inequality. Second, I develop and estimate a structural model, accounting for wage dynamics induced by these two channels, to quantify the effects of coworkers on lifetime income inequality.

I first document new empirical evidence using matched employer-employee administrative data in a large local labor market in Northern Italy. I find that there are large production complementarities and substantial learning by working with good coworkers. Coworkers also explain a substantial fraction of wage inequality that is similar to what firm heterogeneity can explain, where the latter is traditionally viewed as the driving source of wage inequality in the literature (see a survey in Song, Price, Guvenen, Bloom and Von Wachter, 2019). Specifically, I employ an econometric specification that incorporates the average coworker's quality into a wage equation with worker and firm fixed effects (Abowd et al., 1999, hereafter AKM), where a worker's quality is measured by her permanent ability (worker fixed effects), and estimate using a novel estimation method developed by Hong and Sølvsten (2022). I find that a one-standard-deviation increase in the average coworkers' quality leads to an 8 percent rise in a worker's own wage, suggesting a strong coworker production complementarity. I also find that conditional on the current coworkers, a one-standard-deviation increase in the past coworkers' quality leads to a 2 percent rise in a job mover's current wage. Given that past coworkers are not involved in the current work of job movers, this suggests that the effect is likely coming from the learning effect.

Moreover, to shed light on whether coworkers matter for wage inequality, I decompose the wage variance into worker heterogeneity, firm heterogeneity, and peer effects from coworkers, as well as worker sorting between coworkers and firms. I find that the peer effect from coworkers and firm heterogeneity explain 11 percent and 13 percent of the wage variation, respectively. Sorting between workers and coworkers explains around 30 percent of the wage variance, whereas conditional on coworkers, sorting between workers and firms only explains 3 percent. Given that firms have drawn much attention in the recent literature in explaining wage variance, these results have suggested another critical component: coworkers.

The regression model does not account for the lifetime impact of coworkers. It is plausible, however, that a worker may move to firms with good coworkers. She may also become better by learning from coworkers and is subsequently attracted by better peers. To account for the impact of coworkers on the dynamics of worker quality and lifetime wage trajectory, I develop and estimate a labor search model with coworkers. Specifically, I incorporate coworkers into a model with both worker and firm heterogeneity in a frictional labor market. The model has three main novel features. First, a worker not only accumulates human capital by experience (learning by doing) but also learns differently from distinct coworkers. Second, I allow for production complementarity between workers, coworkers, and firms, and the degrees of complementarity are empirically determined. Finally, the model uses a new wage bargaining framework to allow a worker's wage to change when outside options, human capital, and coworker quality change. I estimate the model using indirect inference (Gourieroux, Monfort and Renault, 1993), and the model can replicate the key features of coworkers in the empirical findings.

Using the estimated model, I explore the lifetime income by calculating the net present value of a worker's 30 years' future earnings in a similar spirit of Flinn (2002) and Bowlus and Robin (2004, 2012). Consistent with Cunha and Heckman (2007, 2008), I find that half of the variation in income inequality is determined by their pre-market ability. I decompose the remaining unexplained part into firms, coworker production complementarity, and learning from coworkers. I find that coworker production complementarity and learning from coworkers contribute to 15 percent and 30 percent of it, and the firm component explains another 15 percent. The result shows coworkers play a substantially more important role in explaining lifetime income inequality.

My paper makes several contributions to the literature. First, it contributes to a growing literature studying workplace peer effects in a large local labor market (e.g., Battisti, 2017; Cornelissen, Dustmann and Schönberg, 2017; Cardoso, Guimaraes, Portugal and Reis, 2018; Nix, 2020; Hong and Lattanzio, 2022).[1] These studies typically incorporate peer effects into an AKM specification, where peer quality is measured using average coworkers' individual fixed effects. The model is typically estimated using an iterative method pioneered by Arcidiacono, Foster, Goodpaster and Kinsler (2012),[2] where the consistency of the estimator relies on a homoscedasticity assumption. I adopt a new estimation method developed by Hong and Sølvsten (2022), which allows heteroskedasticity and ensures consistency of the estimator by a novel bias-correction method. I find substantial bias in using the traditional approach, which is 15 percent smaller than the estimate using the new method.[3]

Second, my paper also contributes to a large body of literature that documents the driving sources of wage and income inequality. The existing studies mainly focus on firm heterogeneity as an important source of wage inequality (e.g., Card et al., 2013; Song

[1]A closely related large literature studies workplace peer effects in different individual firms (e.g., Mas and Moretti, 2009; Brune, Chyn and Kerwin, 2020; Holden, Keane and Lilley, 2021; Thiemann, 2022).

[2]Arcidiacono et al. (2012) was originally used in estimating classroom peer effect, and has recently widely adopted in other contexts, including urban economics (e.g., Baum-Snow, Gendron-Carrier and Pavan, 2020), and international trade (e.g., Dix-Carneiro and Kovak, 2017), and management science (e.g., Thiemann, 2022).

[3]Hong and Sølvsten (2022) shows that the direction of bias is theoretically ambiguous and entirely empirically driven. For example, Hong and Sølvsten (2022) replicates the classroom peer effect application in Arcidiacono et al. (2012) and finds that the estimate using the existing method is *larger* than the new estimate. Specifically, they use the transcript information of all students at the University of Wisconsin-Madison during the semester when the university switched its learning mode to online in 2020. They find that the existing method estimates a positive and significant peer effect, while the new method finds it close to zero and statistically insignificant.

et al., 2019).[4] This paper sheds new light on an essential component of the workplace, coworkers, and shows that it explains a substantial fraction of wage inequality, which is similar in magnitude to the fraction firm heterogeneity explains. Moreover, labor market sorting between workers and firms has been viewed as another important source of wage inequality in the existing literature (e.g., Card et al., 2018; Bagger and Lentz, 2019). This paper documents new empirical evidence on labor market sorting. I find that, conditional on coworkers, sorting between workers and firms becomes minimal. In contrast, sorting in the labor market is primarily driven by coworkers.[5] In addition, recent literature has shown how the firm learning environment affects lifetime income inequality (e.g., Gregory, 2021; Arellano-Bover and Saltiel, 2021). However, the actual content of such a learning environment is still under-explored. As the peer effect is one of the most natural forms of the learning mechanism in a firm, this paper unveils this black box by exploring the learning effect of coworkers in lifetime income inequality.

Finally, the paper contributes to a small but growing literature on equilibrium models that study the effect of coworkers on the labor market (e.g., Herkenhoff, Lise, Menzio and Phillips, 2018; Lopes de Melo, 2018; Jarosch, Oberfield and Rossi-Hansberg, 2021; Corblet, 2022; Freund, 2022). In particular, Herkenhoff et al. (2018) use a search-theoretic model of coworkers in a frictional labor market. They find that learning from coworkers accounts for substantial human capital accumulation, but the market is inefficiently sorted for low-quality workers to learn from high-quality counterparts. Jarosch et al. (2021) use a competitive market model to estimate flexible coworker learning functions. They show that coworker learning is significant, which generates compensating differentials for learning. The two models, however, shed little light on coworkers' effect on wage inequality. My model builds on Herkenhoff et al. (2018) but differs in two main directions. First, it includes both worker and firm heterogeneity, whereas Herkenhoff et al. (2018) assumes that firms are inherently homogeneous, and I model production complementarity

[4]Studies find similar results in countries such as Brazil (Alvarez, Benguria, Engbom and Moser, 2018), Denmark (Bagger, Sørensen and Vejlin, 2013), Germany (Card et al., 2013), Portugal (Card, Cardoso, Heining and Kline, 2018), Sweden (Håkanson, Lindqvist and Vlachos, 2021), the UK (Mueller, Ouimet and Simintzi, 2017), and the US (Song et al., 2019).

[5]Lopes de Melo (2018) also surveyed similar (but biased) results using correlations among workers, coworkers, and firms by plugging in the estimates directly from the AKM model. As discussed in Kline, Saggio and Sølvsten (2020), the plug-in computations yield substantial bias due to the limited mobility issues (Andrews, Gill, Schank and Upward, 2008). I differ from Lopes de Melo (2018) in two directions. First, I estimate workers, coworkers, and firms coherently in a wage regression model. Also, I adopt a similar technique in Kline et al. (2020) and correct the bias in all calculations.

not only between workers and coworkers but also between workers and firms. Including both worker and firm heterogeneity not only establishes a close link between the model and the key features of the matched employer-employee data but also allows me to compare my model to canonical labor search frameworks with two-sided heterogeneity (e.g., Postel–Vinay and Robin, 2002).[6] Second, I adopt a new multilateral wage bargaining setting (similar to Lentz and Mortensen, 2012; Elsby and Gottfries, 2022), which generates wage predictions that are consistent with my empirical findings.

The paper proceeds with a description of the data used in the next section. Sections 3 document two sets of novel empirical evidence on coworkers. In Sections 4 and 5 develop and estimate a search-theoretical model, and Section 6 conducts counterfactual exercises to quantify the importance of coworkers in the labor market. Section 7 concludes.

2 Data

This paper uses the matched employer-employee administrative data – Veneto Worker History (VWH) dataset – which contains the entire working population and private firms in Veneto's region in Northern Italy from 1975 to 2001.[7] The database contains three sets of administrative records: a worker-level demographic registry, a firm-level record, and an annual social security contribution register that links the worker and firm records.

The worker registry tracks over three million workers. It records a worker's entire working history in the private sector, including all job spells outside Veneto, as long as he/she worked at least one day in Veneto. It contains rich demographic information, including gender, age, and residency. The firm-level record contains all private firms that employ any worker in the worker register, which contains detailed information such as national tax code, address, and industry, where the national tax code allows researchers

[6] My model also relates to the assignment literature that studies labor market sorting with heterogenous firms and multiple heterogeneous workers, following Becker (1973). Kremer (1993) studies the assignment of multiple workers to each firm and finds there is positive assortative matching whenever the production function is supermodular. Eeckhout and Kircher (2018) derives conditions under which sorting is assortative when firms can hire multiple worker types. In equilibrium, a firm matches with a single worker type. Boerma, Tsyvinski and Zimin (2021) show that when production is instead submodular, equilibrium features a rich, mixed sorting pattern and substantial within-firm worker heterogeneity. To explain the sorting patterns between workers and firms, this paper integrates a supermodular production technology and a submodular learning technology into a frictional labor market model.

[7] The database stopped collecting this information after 2001. As a related comparison in the U.S., Veneto has a similar population size as Wisconsin. Besides, Veneto is one of the most advanced economies in Italy, which shares similar labor market features with many other European countries.

to link external balance-sheet databases.[8] Finally, the social security contribution 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 can calculate each worker's social security contribution. Therefore, the register contains accurate information for each job spell on total earnings (without top-coding), weeks worked, occupation (white-collar, blue-collar, manager, apprentice), type of contract (fixed-term or open-ended), and type of working schedule (full-time or part-time). Earnings have been inflation-adjusted to the price level of the year 2003.

The database is particularly suitable for the study. First, I observe every coworker of each worker over their working life in Veneto. Second, it provides accurate wage records with rich information from both worker and firm sides. One drawback of this database, like many similar matched employer-employee databases, is that it is not possible to identify the exact coworkers with whom each worker frequently interacts. While finer data on coworker interactions can be helpful (e.g., Mas and Moretti (2009) uses data from a large supermarket chain), it is rare to obtain such data on a large labor market scale. In subsection 2.2, I discuss the definition of peer groups in detail.

2.1 Sample selection

I select the period from 1995 to 2001 as the sample of analysis. The labor market in Veneto was in a steady environment with nearly full employment during that period (see Tattara and Valentini, 2010; Serafinelli, 2019).[9] Following the standard practice in the literature, I keep only a worker's primary job if he or she works in multiple positions. Specifically, if a worker has two or more employment contracts in a year, I keep the job with the highest annual earnings and number of weeks worked. I break a few ties by randomly choosing the primary job. I restrict the sample be workers aged 18 to 65 and exclude part-time jobs and apprentices because their wages are set according to a very different rule from regular full-time employment. I also require that each peer group must have at

[8] The firm is not at the establishment level. It would be ideal to use establishment-level data for the analysis, but it likely does not make a difference in my case because the vast majority of firms are single-establishment firms.

[9] Since the econometric model (described below) uses the worker fixed effects as a measure of worker's ability, I use a long enough panel to be able to effectively measure worker flows (essential for identification) but do not want to push the ability fixed-effects assumption too much by keeping the panel relatively short. A similar approach has also been used in Card et al. (2013) and Lentz, Piyapromdee and Robin (2018), for example.

least two workers.[10] Moreover, I restrict the sample workers to private non-agricultural firms within the Veneto region because I cannot observe a complete picture of a worker's coworkers once she leaves Veneto. Lastly, due to the identification requirement in the AKM-type model below, I restrict the sample to the largest connected set (Abowd et al., 1999), which makes up around 98 percent of the resulting observations.

Descriptive statistics Table 1 presents descriptive statistics of the sample used in the analysis. The final sample contains around five million person-year observations with more than one million workers and around seventy thousand firms. A full-time worker earns a mean weekly wage of 833 euros. The average age and tenure are 35 year-olds and 5 years, respectively. Overall, 40 percent of workers change jobs at least once throughout the whole period of analysis. The female labor force participation is relatively low, with 35 percent of workers being female, and the majority of workers are employed in blue-collar occupations (68%). Firms are small, reflecting the structure of the Italian labor market, with a mean firm size of 21 employees and a median of 9. Similarly, the peer group size – that is, workers in the same occupation and firm – has a mean of 10 and a median of 3.

Job mobility The job mobility rate is relatively high in Veneto. Specifically, on average, around 8% of workers move from another firm annually, which is similar to the mobility pattern in the US. On average, around 5 workers move in and out of a firm annually. The overall firm turnover rate is around 20%, meaning that an average firm replaces one-fifth of its workers in a given year. One concern is that this number might be purely driven by small firms. As shown in Figure 1, which plots the annual turnover rate by showing the yearly fraction of newly entered and newly separated workers across different firm sizes, the turnover rate is relatively similar across different firm sizes. Even very large firms, with 1000+ employees, have a turnover rate of around 15%.

2.2 Peer group definition

For my main specifications, I define the peer group as all the workers employed in the same firm with the same occupation in a given year, where the occupation is defined by

[10] Workers in one-worker firms earn about 30 percent lower wages than the labor market average, and they account for only 1.8% of the total observations.

general professional levels (blue-collar, white-collar, and managers). There are a few potential concerns with the definition. First, having managers as one separate peer group might be too conservative because the manager might also have spillovers to the supervising workers. In a robustness check, I assign the managers to either blue-collar or white-collar by exploring which position the manager is promoted from.[11] The results are very similar. This is likely due to the fact that managers only account for 2 percent of my sample as shown in Table 1, which is unlikely to alter the main results.

Second, the peer definition might be somewhat too broad given that there might be multiple occupations within the same professional types (e.g., consultants and accountants are both white-collar workers). However, there is no clear reason why consultants and accountants cannot work in a team. In fact, in many firms, teamwork commonly involves interdisciplinary skills. Given the peer group size is typically small (a median of 3), such teamwork is very likely to happen.[12]

Finally, if the peer group does not perfectly reflect the true interaction in the firm, it is likely to introduce a specific source of measurement error, as discussed in Cornelissen et al. (2017) and Nix (2020), which would attenuate the estimates. If this is the case, my approach provides a lower bound of the true peer effect.

3 Empirical evidence

This section estimates the effect of coworkers on wages using an econometric specification that incorporates coworkers into a wage regression with worker and firm fixed effects (Abowd et al., 1999). It provides empirical evidence for the two main mechanisms – production complementarity and learning – through which coworkers affect wage dynamics and also shows that coworkers explain a substantial fraction of wage inequality.

[11] To be specific, if the manager is promoted from a white/blue-collar worker, I group her to white/blue-collar. For a few observations, a worker might always be a manager. I check whether there are white-collar workers in the firm. If all the coworkers are blue-collar, I group her to blue-collar.

[12] Cardoso et al. (2018) finds that the peer effects are quantitatively similar with different occupational definitions within a firm using the matched employer-employee administrative database in Portugal, which has a very similar institutional setting as Veneto.

Table 1: Summary statistics

	(1) Mean	(2) Standard Dev.	(3) Median
Worker level			
Weekly wage	832.91	2731.75	701
Age	34.79	9.85	33
Tenure	4.60	5.02	3
Mover	0.36	0.48	
Woman	0.35	0.48	
Blue-collar	0.68	0.47	
White-collar	0.30	0.46	
Manager	0.02	0.14	
Firm level			
Firm size	21	92	9
Peer group size	10	51	3
Annual mobility	5	36	2
Person-year observations		4,828,066	
Number of workers		1,203,965	
Number of firms		68,883	

Notes. The table reports the average, standard deviations, and medians of each variable in columns (1) to (3). The worker-level statistics are computed based on all person-year observations, and I calculate the firm-level statistics by collapsing the data such that the unit of observation is at the firm-year level.

3.1 Empirical Strategy

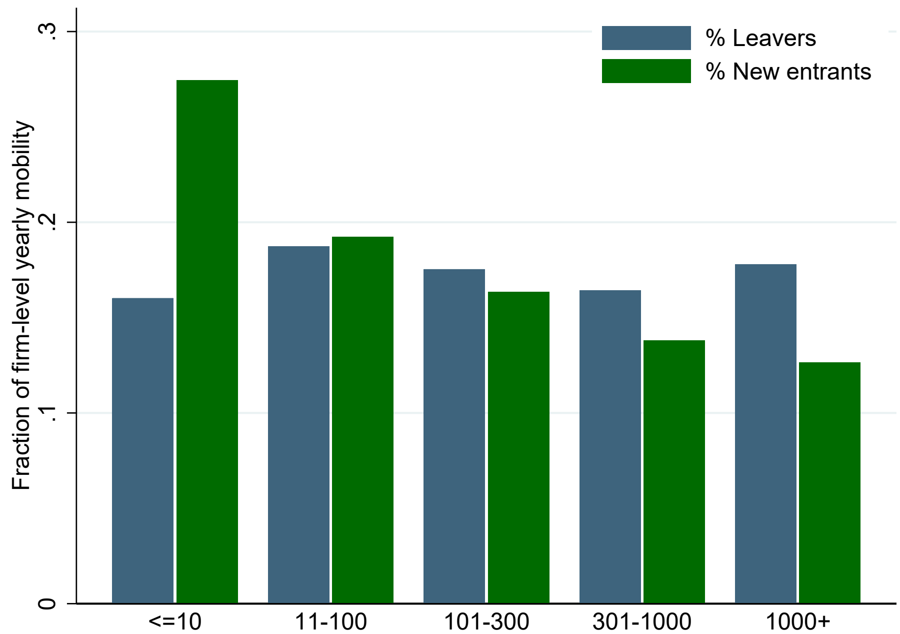
Formally, I estimate a peer effect model by incorporating a coworker component into the canonical AKM econometric specification. Specifically, I estimate a wage equation as expressed below.

$$y_{it} = \alpha_i + \beta \cdot \bar{\alpha}_{-i,t} + \psi_{j(it)} + \mathbf{x}'_{it}\,\xi + \varepsilon_{it} \tag{1}$$

where i, j, and t are worker, firm, and year indices, respectively. y_{it} is the log weekly wage; α_i is the worker fixed effect, which captures a worker's unobserved permanent ability/human capital or a measure of worker *quality*; $\psi_{j(it)}$ is the firm fixed effect, which measures the firm permanent productivity or wage premium. The covariates $\mathbf{x}_{it}$ are the time-varying observables, including age squared,[13] tenure, tenure squared, log firm size,

[13] Age is not included as it is equal to calendar year (t) minus birth year (absorbed by worker fixed effect) and is not identified in this model. For the age-squared term, I normalize the age to age - 40 since the

Figure 1: Firm annual turnover by firm size



Notes. The figure reports, at the firm level, the average fractions of the annual newly entered and newly separated.

occupation fixed effects, and year fixed effects. ε_{it} is the error term. All of the above terms are standard in the AKM framework. The disparity comes from the incorporation of coworker's average quality, $\bar{\alpha}_{-i,t}$, measured as below.

$$\bar{\alpha}_{-i,t} = \frac{1}{|M_{-i,t}|} \sum_{\ell \in M_{-i,t}} \alpha_{\ell}, \text{ and } M_{-i,t} = \{\ell : oj(\ell, t) = oj(i, t), \ \ell \neq i\}. \tag{2}$$

where $M_{-i,t}$ is the set of worker i's peer group but *excluding* worker i in the same occupation (o) and firm (j). In other words, $\bar{\alpha}_{-i,t}$ is the average coworker's unobserved permanent ability at time t .

The parameter β measures the peer effect of average coworker quality on her own wages. There are two sources of variation for the identification of β. For job switchers, peer quality changes when they move to another firm. For job stayers, peer quality changes when other workers join or leave the peer group.[14] As a comparison with the

average wage tends to be flat after age 40. As pointed out in Card et al. (2018), such normalization can be critical in estimating worker fixed effects.

[14]The mobility may come from a within-firm promotion (e.g., from blue-collar to white-collar), which is an ideal variation for analysis. However, such mobilities only account for less than 3 percent of the mobility

identification requirement to the AKM model, which only requires job switchers linking all the firms (in the largest connected set), the identification for the peer effect model uses variations from *both* job stayers and job movers. Therefore, the entire regression model, including worker and firm fixed effects, is identified using the two sources of variations.

Moreover, similar to Abowd et al. (1999), the conditional expectation of the error term is assumed to be zero. As discussed in Card et al. (2013), a *sufficient* condition for this assumption, in the case of only worker and firm fixed effects (i.e., AKM model), is that the conditional job mobility from firm to firm is as-if exogenous. They specifically use an event study approach to explore this condition. Such a strategy may not be easily extended to my setting because the analogous condition is that job mobility has to consider both coworker quality, which is time-varying, and firm types.

One natural concern is that there is no clear reason for the *average* coworker quality to be a "correct" measure for coworkers. It could be the median coworker or the best coworker who has the highest influence on worker i. The reason for using the average coworker's quality as my preferred measure is primarily two-fold. First, from the econometric perspective, the average assumption is critical for the consistency of β.[15] Second, given that firm sizes are typically small in Italy, the average may not be a wild measure. It also lies in a large empirical literature that fosters the linear-in-means model, which allows me to compare my results to other work.

3.2 Estimation

Estimation of Equation 1 is challenging for two primary reasons. First, due to the large number of workers and firms, the fixed effects have a large dimension. Second, the equation is non-linear in parameters, i.e., $\beta \cdot \bar{\alpha}_{-i,t}$, where $\bar{\alpha}_{-i,t}$ is a function of the worker fixed effect estimates.

Previous studies rely on the iterative method pioneered by Arcidiacono et al. (2012).[16] The main idea of their method involves two steps. First, they start with a guess (or the estimates from the prior iteration) of α_i to construct $\bar{\alpha}_{-i,t}$ and estimate equation (1) by OLS. Second, conditional on the estimates, they update α_i, and thus $\bar{\alpha}_{-i,t}$. They repeat

in the data, which is unlikely to alter the main results.

[15] Finding an alternative estimator of β if coworkers are not average is beyond the scope of this paper but is promising for future research.

[16] Holden et al. (2021) has a parallel development on the estimation method using a similar idea and has a faster computational speed.

the two steps by solving a fixed-point problem until all estimates coverage. Their method shows that β can be consistently estimated under the assumption that the error term is homoskedastic. As discussed in Appendix A, relaxing this assumption can violate the consistency of the estimator β.[17]

This paper adopts a new method developed by Hong and Sølvsten (2022), which allows for heteroskedasticity and uses a novel bias-correction method for the consistency of the estimator. Correcting such a bias is also important from a practical point of view because, as shown in Hong and Sølvsten (2022), the direction of bias is theoretically ambiguous and entirely empirically driven. A brief description of the estimation method is provided in Appendix A. It is also worth noting that the fixed effects are not consistently estimated, but the method ensures that β is consistently estimated even if the fixed effects are not. Moreover, inference of β can be challenging. A common approach in the literature is to use the wild bootstrapping method (e.g., Arcidiacono et al., 2012; Cornelissen et al., 2017). I adopt a novel analytical solution derived in Hong and Sølvsten (2022) to compute the standard error of β.

Finally, the interpretation of peer effect β is canonical: a one-standard-deviation increase in the average peer quality $\sigma_{\bar{\alpha}_{-i,t}}$ corresponds to a $\beta\, \sigma_{\bar{\alpha}_{-i,t}}$ percent increase in wages, on average.[18] To this end, I adopt a similar approach in Kline et al. (2020) to estimate the standard deviation of the peer quality to correct potential bias in the variance of the fixed effect estimates.

3.3 Mechanism 1: production complementarity

Table 2 shows the baseline results of the peer effect in Column (1), where β is estimated to be around 0.41, which is statistically significant at the one percent level, and the bias-corrected standard deviation of $\bar{\alpha}_{-i,t}$ is 0.19. This means that a one-standard-deviation increase in peer quality increases the wage level by 7.8 percent. The effect is also economically meaningful when one compares it with other drivers of wages. For example, it is similar in magnitude to the return to schooling, where Lucifora, Comi and Brunello (2000) finds that the return to one year of schooling in the same period is about 7 percent

[17] In line with the discussion in White and Domowitz (1984), with heteroskedasticity, β is, in general, inconsistent in a non-linear model. In equation (1), there is non-linearity in parameters, i.e., $\beta \cdot \bar{\alpha}_{-i,t}$.

[18] As the average peer quality is close to a normal distribution as shown in Appendix Figure 1, one can also consider that as a roughly 34 percentile increase from the mean or the median.

in Italy.

The large contemporaneous peer effect suggests that there is strong *production complementarity* by working with coworkers.[19] The story is also consistent with recent studies in Herkenhoff et al. (2018) and Jäger (2016), which find large production complementarity exists among workers from both using a structural model approach and a quasi-natural experiment method, respectively.

Table 2: Contemporenous peer effects

	Baseline (1)	Robust 1: Common shocks (2)	Robust 2: Placebo peers (3)
Peer effect $\widehat{\beta}$	0.413 (0.035)	0.367 (0.028)	0.016 (0.005)
Std Dev. of $\bar{\alpha}_{-i,t}$	0.191	0.198	0.210
1-SD effect	7.8%	7.3%	0.33%
Worker FE	x	x	x
Firm FE	x		x
Occupation FE	x		x
Year FE	x		x
Firm $\times$ Occup $\times$ Year FE		x	

Notes: The table reports the contemporaneous peer effects on wages under different specifications. The estimation and inference follow Hong and Sølvsten (2022), and the standard deviation of peer quality is estimated using a similar approach of Kline et al. (2020).

Comparison to the existing method As mentioned in Section 3.2, the exiting method may be biased when the homoscedasticity assumption is relaxed. I use the same estimator as proposed by Arcidiacono et al. (2012), and the peer effect β is estimated to be around 0.351, which is around 15 percent smaller than the one estimated using Hong and Sølvsten (2022) as shown in Table 2 Column 1.

Common shocks The peer model in equation (1) can be extended to include the potential presence of time-varying peer group-specific wage shocks that are correlated with

[19] Some papers interpret the peer effect as peer pressure (e.g., Mas and Moretti, 2009). From a practical point of view, one can also consider peer pressure as production complementarity because it increases a worker's productivity when good coworkers are around. However, such interactions do not generate knowledge spillover or learning from coworkers.

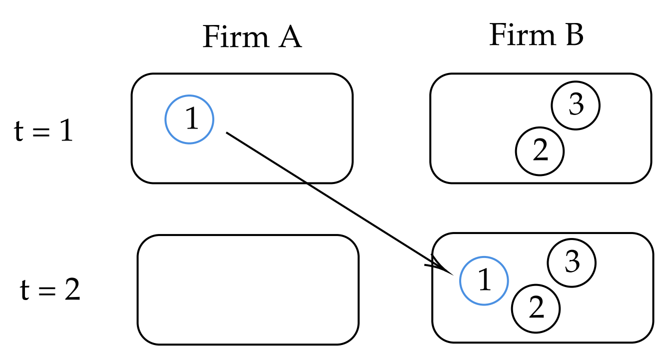
shocks to peer group quality. For example, a firm might adopt a new technology or a new machine specific to one occupation only, which simultaneously raises wages and worker quality in that occupation relative to other occupations in the firm.

To deal with this issue, I estimate the following equation, where ξ_{jot} is the firm-occupation-year fixed effects, which absorbs the time-varying peer group-specific common shocks.

$$y_{it} = \alpha_i + \beta \cdot \bar{\alpha}_{-i,t} + \xi_{jot} + x'_{it}\xi + \varepsilon_{it}$$

With job switchers, one can still separately identify worker and firm-occupation-year fixed effects (Engbom, Moser and Sauermann, 2022). Moreover, one can identify the peer effect β as the average peer quality is calculated by excluding worker i. The identifying variations for the peer effect β, however, come from job stayers: in particular, the peer composition change and the peer size changes. Figure 2 shows a simplified example, which provides the key intuition for identifying variation. Consider two firms, A and B, with only white-collar workers. In the first period, Firm B has Worker 2 and Worker 3. In the second period, Worker 1 moves from Firm A to Firm B. For worker 1, her new peer group is entirely captured by firm-year fixed effect (being workers 2 and 3). However, for *stayer* Worker 2 (an analogous effect also happens to Worker 3), her peer composition changes from Worker 3 to Worker 1 and Worker 3 with average peer quality changing from α_3 to $\alpha_1/2 + \alpha_3/2$. As a result, there are two changing forces: (i) Worker 1 contributes to the new peer quality, and (ii) Worker 3's contribution halves since the peer quality is measured with average. In other words, both the peer composition change and the size of peer groups change contribute to the identification of the peer effect.

Figure 2: A simplified example of identifying mobility



Notes. Depiction of three individuals (denoted 1, 2, and 3) and Worker 1's mobility among two firms (denoted A and B). In the first period, individuals 2 and 3 are peers, while individuals 1, 2, and 3 are peers in the second period. The worker composition except Worker 1 in Firm A is omitted for simplification.

Table 2 Column 2 shows that the peer effect β is estimated to be around 0.37, which is very close to the baseline results in Column 1. Given the similar results, I keep it as a robustness exercise and keep using equation (1) as my main specification so that I can use identifying variation from both job movers and job stayers. It also allows the peer effect model to be closely linked to the widely studied AKM specification in the literature.

Placebo peer groups An implicit assumption imposed in equation (1) is that there is little spillover across different peer groups. I explore this assumption in two directions. First, the manager's group is likely to have an influence on the other two groups (white-collar and blue-collar workers). As discussed above, as a robustness check, I assign the managers to either blue-collar or white-collar by exploring which position the manager is promoted from. The resulting peer effect estimate is almost identical. Since managers only account for 2 percent of the total observation, it may be infeasible to detect it using the current empirical framework.

The other is to explore whether the blue-collar workers are affected by the white-collar workers and vice versa. To shed light on this question, I define the peer group as all the workers in other occupations and re-estimate the equation (1). Table 2 Column 3 reports that the placebo peer effect from the non-peer is economically small, suggesting that the across-peer spillover is minimal.

3.4 Mechanism 2: human capital accumulation

A natural key difference between working with machines and coworkers is that knowledge spillovers are likely to take place while working with coworkers. As teamwork is common in the workplace, learning from coworkers can potentially contribute to non-trivial on-the-job human capital accumulation. However, few studies have documented this aspect of peer effects. To shed light on this question, I utilize a sample of job movers and explore a dynamic peer effect model described below.

$$y_{it} = \alpha_i + \beta_0 \cdot \bar{\alpha}_{-i,t} + \beta_1 \cdot \bar{\alpha}_{-i,t-1} + \psi_{j(i,t)} + x_{it}'\eta + \varepsilon_{it} \tag{3}$$

where $\bar{\alpha}_{-i,t-1}$ is the average coworkers' quality and the firm fixed effects in the period before moves.[20] β_0 and β_1 are the concurrent and lagged peer effects, respectively.

The parameter of interest is β_1, which measures human capital spillover from the prior coworkers. The idea is straightforward: since past coworkers are not involved in the current production as the worker has moved to another firm, past coworkers are likely affecting current wages through the human capital channel by learning from them.

I estimate equation (3) with the full sample, conditional on observing workers for at least two periods. I then use the estimated $\bar{\alpha}_{-i,t}$ and $\bar{\alpha}_{-i,t-1}$ to re-run the same regression using the mover sample. Table 3 Column 1 shows that the current peer effect, $\widehat{\beta}_0$, is similar to that in the baseline result in Table 2, and the lagged peer effect, $\widehat{\beta}_1$, is estimated to be around 0.095. It means that a one-standard-deviation increase in the past coworker's quality increases a job mover's current wage by 2 percent. As the average return to one year of working experience in Veneto is around one percent (calculated in Appendix Table 1), the effect of past coworkers is still sizable.

One concern is that a poaching firm gives a higher wage offer as workers' coworkers are better. Intuitively, a firm needs to provide a higher financial incentive to compensate a worker leaving a good coworker environment. To mitigate such an issue, I use a subsample of displaced movers who became unemployed due to mass layoffs and firm closures.[21] In this case, all the workers are hired from the unemployment spell, which makes the human capital channel more straightforward. As shown in Table 3 Column 2, the effect from the displaced sample is somewhat smaller but is still sizable.

3.5 Wage inequality

The findings from the previous subsections suggest that coworkers matter for both production complementarity and learning. These mechanisms can have a dramatic impact on wage inequality. First, with production complementarity, good workers are more likely to be sorted with good coworkers, creating wage polarization. Second, differential learning from coworkers can widen the wage growth gap. As a result, coworkers can contribute

[20] With slight abuse of notation, I refer to the previous period as the prior job spell. Nevertheless, the majority of the last spells refer to the previous year.

[21] I follow the definitions in existing literature (summarized in Arellano-Bover and Saltiel, 2021). Specifically, I define mass layoffs as a firm's total employment dropping below half of its prior three-year moving average without subsequently recovering and define firm closure as a firm is closed that does not subsequently reappear in the data. Moreover, I restrict firms with at least 20 employees at their pre-layoff average.

Table 3: Dynamic peer effects on wages

	Baseline: Movers	Robust: Displaced workers
Current peer effect β_0	0.399	0.376
	(0.019)	(0.069)
Lagged peer effect β_1	0.095	0.086
	(0.022)	(0.027)
Std dev. of current peer, $\bar{\alpha}_{-i,t-1}$	0.190	0.189
Std dev. of past peer, $\bar{\alpha}_{-i,t-1}$	0.211	0.186
1-SD effect		
Current peer	7.6%	7.1%
Past peer	2.0%	1.6%

Notes: The table reports the coefficients of the current and past coworkers in equation (3) using the mover sample in Column 1 and the displaced worker sample in Column 2.

to non-trivial inequality. However, little is documented in the literature. This section uses a wage variance decomposition exercise to shed new light on this question.

3.5.1 Wage variance decomposition

Throughout the paper, I use log wage variance as a measure of wage inequality, similar to many existing studies (notably, Card et al., 2013). Following Card et al. (2018) and Sorkin (2018), I use the "ensemble" decomposition method for the variance of the log wages, where I decompose the wage variance into each regressors in equation (1).[22]

$$\text{var}(y_{it}) = \text{cov}(y_{it}, \alpha_i) + \text{cov}(y_{it}, \psi_j) + \text{cov}(y_{it}, \beta \cdot \bar{\alpha}_{-i,t}) + \text{cov}(y_{it}, x_{it}'\eta) + \text{cov}(y_{it}, \varepsilon_{it}) \quad (4)$$

The formula provides direct evidence about the contribution and the importance of each component in equation (1) to the variance of log wages (y_{it}). For example, the share of the variance of wages explained by the coworker effect is captured by the third term in equation (4) divided by the wage variance, i.e., $\text{cov}(y_{it}, \beta\bar{\alpha}_{-i,t}) / \text{var}(y_{it})$.

One can also plug in y_{it} using equation (1) into each term on the right-hand side of equation (4) and fully expand the expression (which will not be shown here as it is very

[22] I first plug in equation (1) to the following equation: $\text{var}(y_{it}) \equiv \text{cov}(y_{it}, y_{it}) = \text{cov}(y_{it}, \alpha_i + \psi_j + \beta \cdot \bar{\alpha}_{-i,t} + x_{it}'\eta + \varepsilon_{it})$, and expand the right hand side of the equation as covariance is a linear operator.

messy). From the full expression, one can collect the covariance terms between α_i and $\bar{\alpha}_{-i,t}$ and the covariance terms between α_i and ψ_j, which measure labor market sorting between workers and coworkers as well as workers and firms, respectively, as expressed below, omitting the remaining terms.[23]

$$\operatorname{var}(y_{it}) = ... + 2\operatorname{cov}(\alpha_i, \bar{\alpha}_{-i,t}) + 2\operatorname{cov}(\alpha_i, \psi_j) + ... \tag{5}$$

From there, the fractions of the wage variance explained by the sorting between workers and coworkers or firms are simply $2\operatorname{cov}(\alpha_i, \bar{\alpha}_{-i,t}) / \operatorname{var}(y_{it})$ and $2\operatorname{cov}(\alpha_i, \psi_j) / \operatorname{var}(y_{it})$, respectively.

Studies have shown that the wage variance decomposition method can generate biased values if one directly plugs in the estimates from equation 1 (e.g., Andrews et al., 2008; Bonhomme, Holzheu, Lamadon, Manresa, Mogstad and Setzler, 2020). To address this issue, the paper adopts the technique developed by Kline et al. (2020), which is used to correct such bias in the AKM model. I conduct a similar bias correction exercise by incorporating the method into the peer model.[24]

3.5.2 AKM model decomposition

Before I show the wage variance decomposition results, it is useful to compare the results to well-documented wage variance decomposition using the canonical AKM model as expressed below:

$$y_{it} = \alpha_i + \psi_j + x_{it}'\eta + \varepsilon_{it}. \tag{6}$$

where α_i and ψ_j are worker and firm fixed effects, and x_{it} contains the same set of covariates as in equation (1).

Table 4 Column 1 shows the wage variance decomposition results using equation (6). The worker heterogeneity (α_i) and firm heterogeneity (ψ_j) explain around 53% and 18%

[23]One can expand the equation and collect the covariance terms as shown below. For simplification, I ignore the covariates (x_{it}) and the error term ε_{it} here.

$$\begin{aligned}\operatorname{var}(y_{it}) &= \operatorname{cov}(\alpha_i + \bar{\alpha}_{-i,t} + \psi_j, \alpha_i) + \operatorname{cov}(\alpha_i + \bar{\alpha}_{-i,t} + \psi_j, \alpha_i) + \operatorname{cov}(\alpha_i + \bar{\alpha}_{-i,t} + \psi_j, \alpha_i) \\ &= \operatorname{var}(\alpha_i) + \operatorname{var}(\beta\bar{\alpha}_{-i,t}) + \operatorname{var}(\psi_j) + 2\operatorname{cov}(\alpha_i, \beta\bar{\alpha}_{-i,t}) + 2\operatorname{cov}(\alpha_i, \psi_j) + 2\operatorname{cov}(\beta\bar{\alpha}_{-i,t}, \psi_j)\end{aligned}$$

If I include the covariates, the expression can be much messier. Therefore, I use the "ensemble" decomposition in equation (4) as my preferred decomposition method instead of using a fully expanded expression.

[24]Appendix Table 3 presents a version of the results when the bias correction is not performed. Comparing them to the bias-corrected results in Table 4 below, the biases are very sizable.

of the wage variation, respectively. These estimates are similar to what the literature has documented in Europe (e.g., Card et al., 2018) and in the United States (e.g., Sorkin, 2018). There is a considerable amount of positive sorting between worker and firm, which explains roughly 11 percent of the wage variance, which corresponds to a moderate correlation of 0.211.

3.5.3 Peer model decomposition

Table 4 Column 2 shows the wage variance decomposition of the peer effect model in equation (1). The fractions explained by worker heterogeneity and firm heterogeneity are around 47% and 13%, respectively, which is somewhat smaller than the ones reported using the AKM model. On the other hand, the fraction explained by the coworker component $\beta\bar{\alpha}_{-i,t}$ explains roughly 11 percent of the wage variance, which is similar to what the firm heterogeneity explains.

Turning to the sorting, the results are surprisingly different compared to the one reported in the AKM model. In particular, the sorting between workers and firms only explains around 3% of the wage variance, corresponding to a small correlation of less than 0.1 between workers and firms, which is substantially smaller than the sorting between workers and firms explained by the AKM model. On the other hand, the sorting between workers and coworkers explains around 28% of the wage variance. The resulting correlation between workers and coworkers is as high as 0.596.[25] The finding deviates somewhat from the existing studies of sorting in the labor market, which typically is between workers and firms, and suggests that, when considering coworkers, the labor market sorting is primarily dominated by coworker sorting instead of sorting between workers and firms.

4 Search-theoretic model

The previous section shows that coworkers matter for wage inequality, and identifies two main mechanisms – production complementarity and learning – through which coworkers affect wage dynamics. However, it is challenging to disentangle how these two mechanisms contribute to wage inequality. As the empirical evidence cannot account for the

[25] Lopes de Melo (2018) estimated worker fixed effects from an AKM specification and calculates $\text{cov}(\alpha_i, \bar{\alpha}_{-i,t})$ by directly plugging the estimated worker fixed effects, and find a similar correlation of 0.55 in Brazil data.

Table 4: Wage variance decomposition

Variance explained by	AKM Model (1)	Peer Model (2)
Worker effects, α_i	53.3%	47.1%
Firm effects, ψ_j	18.2%	13.3%
Peer effects, $\beta \cdot \bar{\alpha}_{-i,t}$		11.2%
Sorting: $2\text{cov}(\alpha_i, \psi_j)$	10.8%	3.39%
Sorting: $2\text{cov}(\alpha_i, \beta\bar{\alpha}_{-i,t})$		27.6%
Correlation: (worker, firm)	0.211	0.080
Correlation: (worker, coworker)		0.596

Notes: The table reports the wage variance decomposition of different components in the AKM model and the peer model. I adopt a similar method as Kline et al. (2020) to correct the bias stemming from the limited mobility in the data for all the estimates.

endogenous worker mobility induced by these two channels, without a structural model, it is difficult to understand counterfactual environments where production complementarity and learning are not considered.

To better quantify how coworkers affect wage dynamics and inequality, I develop a structural model that allows for endogenous worker mobility induced by production complementarity and learning. Specifically, I incorporate coworkers into a labor search model with worker and firm heterogeneity in a frictional market. It has three main novel features. First, a worker not only accumulates human capital using learning by doing but also learns differently from distinct coworkers. Second, a worker has production complementarity both with her coworkers and firms, where the degrees of complementarity are empirically determined. Finally, the model uses a new wage bargaining framework to allow a worker's wages to change when outside options, human capital, and coworker quality change. The model is in a steady-state economy, and time is discrete.

4.1 Workers

The model consists of a continuum of workers with measure 1. A worker is born with a permanent ability that is acquired before entering the labor market. She accumulates human capital while working, which is similar to experience, and her human capital evolves differently with different types of coworkers (and no coworkers).

Formally, a worker is characterized by the tuple (a, k, x), i.e., ability, human capital level, and employment status. A worker's pre-market permanent ability level a is drawn from an exogenous distribution $\Phi(.)$, which is parameterized to be a log-normal distribution $LogN(0, \chi_w^2)$. While working, a worker can gain different levels of human capital h_k, where k is discrete and goes from 1 to K, and its value evolves in a polynomial functional form:

$$\log(h_k) = b_0 + b_1 k + b_2 k^2$$

where I normalize the first human capital level to be one, $h_1 = 1$, so that $b_0 = -b_1 - b_2$.

A worker's productivity is $\pi_{ka} = h_k \cdot a$. Throughout the rest of the paper, I denote worker type ka as a worker with human capital h_k and ability a.

4.2 Firms and teams

The model also consists of a continuum of firms with measure μ and a continuum of teams with measure 1. A firm is characterized by its permanent productivity and is consisted of a collection of teams. Within a firm, the teams are ex-ante the same – consisting of two potential vacancies. A team produces based on the team members' productivity. A two-worker team may produce more than the combination of two single-worker teams due to complementary production technology. A firm has a two-layer production function. Across teams, a firm produces using a constant return to scale technology across the teams, i.e., teams are producing independently within a firm. For each team, the firm uses its own productivity and team production to produce according to another complementary production technology. I describe the team and firm separately below for better illustration.

The setting of a team follows Herkenhoff et al. (2018). A team or a job within a firm is ex-ante homogeneous and is characterized by its vacancy status $z = (.,.)$, where each team can match up to two workers.[26] A team uses workers' productivity to produce according to the following simplified CES production function, where ρ governs the degree

[26] This is a simplification assumption to reduce the computational burden. However, extending the team size does not change the key tradeoffs. Moreover, since a firm has a collection of teams, I can still match the firm size distribution in the data.

of the production complementarity between the two workers.

$$t(z) = \left(\pi_{ka}^{\rho} + \pi_{\ell t}^{\rho}\right)^{\frac{1}{\rho}} = \begin{cases} 0 & \text{if } z = (0\ ,0\ ) \\ \left(\pi_{ka}^{\rho} + 0\right)^{\frac{1}{\rho}} = \pi_{ka} & \text{if } z = (ka, 0\ ) \\ \left(\pi_{ka}^{\rho} + \pi_{\ell t}^{\rho}\right)^{\frac{1}{\rho}} & \text{if } z = (ka, \ell t) \end{cases} \tag{7}$$

Specifically, if there is no worker in the team, it produces nothing. If there is one worker ka, then it produces using the worker's productivity π_{ka}. If there are two workers ka and ℓt (the coworker has a human capital level ℓ and ability t), then the team produces according to a CES function specified in equation (7). If ρ is estimated to be smaller than one, then the team production function is supermodular, which can lead to an (ex-post) sorting between workers and coworkers in the labor market.

A firm has permanent productivity p drawn from an exogenous (population) distribution $\Psi(.)$, which is parameterized to be a log-normal distribution, $LogN(0, \chi_f^2)$. The firm production contains two parts. First, *across teams*, the production follows a constant return to scale production – that is, each team produces independently from the others. Conceptually, one might consider different teams as very distinct jobs which operate independently.

$$F(p) = \sum_{z} f(p, z) \tag{8}$$

where $f(p,z)$ is a team-specific firm production, which leads to the second part of the firm production function. A firm also uses the team production and its own productivity to produce according to the following simplified CES production function, where η governs the degree of the production complementarity between the firm and the team and f_0 is a normalization scale parameter that helps to match the overall production level in the data.

$$f(p,z) = \begin{cases} 0 & \text{if } z = (0\ ,0\ ) \\ f_0\ \left(p^{\eta} + t(z)^{\eta}\right)^{\frac{1}{\eta}} & \text{otherwise} \end{cases} \tag{9}$$

Specifically, if there is no worker in a team, the firm also produces nothing. If the team produces, then the firm uses its productivity p and the team's production $t(z)$ to produce according to equation (9). If η is estimated to be smaller than one, then the firm production is supermodular, which can generate an (ex-post) sorting between workers and firms.

The firm size is assumed to be exogenous and is drawn from the empirical firm size distribution, and I delay the detailed description in Section 4.4.1 after I have properly defined the stationary distribution of the workers and teams.

4.3 Timing

The timing of the model is discrete, and each period is divided into a few stages.

Learning At the beginning of each period, a worker's human capital involves with probability $g_h(k' \mid k, x) \in [0,1]$, which depends on the current human capital and employment status. Note that I assume that a worker can only change human capital by at most one level within a period because the model will be simulated at a monthly frequency. The learning function takes the following form:

$$g_h(k' \mid k, x) = \begin{cases} \beta_s & \text{if } x = 0, \quad \text{where } k' \in [k, \min\{k+1, K\}] \\ \beta_s + \beta_c \cdot \pi_{\ell t} & \text{if } x = \ell t, \quad \text{where } k' \in [k, \min\{k+1, K\}] \end{cases} \tag{10}$$

The key feature of the learning function is that the learning rate depends on the employment status. If a worker is employed with no coworker, the probability of increasing human capital is β_s, which basically measures the learning by doing. On the other hand, if a worker is employed with a coworker, apart from learning by doing, the worker could also learn from the coworker ℓt, which is captured by $\beta_c \cdot \pi_{\ell t}$. I use the simple linear form of learning to be consistent with the peer effect model in the empirical part.[27]

I additionally assume that worker does not depreciate human capital, which might be particularly relevant during unemployment. The assumption echoes the institutional setting where the Italian government provides extensive unemployment training programs for unemployed workers, presumably to prevent human capital depreciation.[28]

[27] Nevertheless, the learning function is relatively restrictive, and exploring a more flexible functional form is an important arena for future work.

[28] It is straightforward to extend the learning function by adding a probability of human capital depreciation. However, as the paper is more concerned about the relative importance between learning-by-doing (β_0) and learning from coworkers (β_c), adding human capital depreciation does not provide an additional tradeoff in this aspect.

Entry-exit At the entry-exit stage, a worker leaves the market with probability $\sigma \in [0,1]$, and she will be replaced by a new worker. The new worker enters the labor market unemployed with human capital h_1 and draws her ability a from an exogenous distribution $\Phi(.)$.

Search-and-match The labor market is frictional, and the search is random. There is an exogenous separation rate λ_s, and a worker randomly meets a firm with probability λ_u if unemployed and with probability λ_e if employed. Upon meeting, the firm and worker will match if their joint value is larger than their combined outside values. Because a job has a vacancy constraint, if the team of two workers meets a good worker, it can replace an inferior worker with a good worker. Once the match is formed, the unit starts to operate according to the production functions in equations (7) and (9).

Similar to Bagger, Fontaine, Postel-Vinay and Robin (2014), I assume all the events in the enter-exit and search-and-match stages are *mutually exclusive*.

4.4 Value functions and the equilibrium

There are four types of (joint) value functions in the model. I denote $V^p_{0,0}$, $V^p_{ka,0}$, $V^p_{ka,\ell t}$, and U_{ka} as the value of an empty firm with productivity p, the joint value between a firm with productivity p and a worker ka, the joint value between a firm with productivity p, a worker ka, and a worker ℓt, and the unemployment value of a worker ka, respectively.

I denote $\hat{V}$ as the values in the next period when the human capital evolves. Due to higher human capital, a worker might find it more valuable to stay unemployed, thus dissolving the match. As a result, $\hat{V}$ has the following expressions:

$$\hat{V}^p_{ka,0} = \max\{V^p_{ka,0}, V^p_{0,0} + U_{ka}\} \tag{11}$$

$$\hat{V}^p_{ka,\ell t} = \max\{V^p_{ka,\ell t}, V^p_{\ell t,0} + U_{ka}, V^p_{ka,0} + U_{\ell t}\} \tag{12}$$

Now, I describe each joint value function in detail separately below.

4.4.1 Stationary distributions

Before proceeding to the value functions, it is helpful to introduce some notation for the stationary distributions of workers and teams. Specifically, I denote u, e, and n, as the measures of unemployed workers, employed workers, and teams, at the production stage of the period. Formally, I denote u_{ka} as the measure of unemployed workers of type ka, $e^p_{ka,0}$ as the measure of workers of type ka who are employed without a coworker at firm type p, and $e^p_{ka,\ell t}$ as the measure of workers of type ka, employed with a coworker of type ℓt at firm type p.

The distribution of workers implies a distribution of teams. The measure of teams with only one worker under firm type p is $n^p_1 = \sum_{ka} e^p_{ka,0}$, and the measure of teams with two workers under firm type p is $n^p_2 = \sum_{ka,\ell t} e^p_{ka,\ell t}/2$. The remaining teams are empty with a measure of $n^p_0 = \pi^p - n^p_1 - n^p_2$, where π^p is exogenously determined by the firm size distribution in the data, where I infer firm types by ranking the empirical average wages of firms. It follows that $\sum_p \pi^p = 1$, given the teams have a unit measure. Because both team and worker have a unit measure, firms naturally have empty teams (for future workers to join), which serves as a convenient approach to not explicitly modeling the boundary of firm sizes.

4.4.2 Team without workers

The value $V^p_{0,0}$ of a team of type p without employees is such that

$$V^p_{0,0} = 0 + \delta \Big\{ \sum_{ka,x,q} \underbrace{m(ka,x,q)}_{\substack{\text{prob the firm} \\ \textit{meets} \text{ a worker}}} (1-\gamma) \max \{ \underbrace{V^p_{ka,0} - V^p_{0,0}}_{\substack{\text{firm's valuation} \\ \text{of worker } ka}} - \underbrace{v(ka,x,q)}_{\substack{\text{worker's} \\ \text{outside option}}}, 0\} + V^p_{0,0} \Big\} \quad (13)$$

In the current period, the team produces nothing. In the next period, the team contacts a worker of type (ka, x) in a firm type q, with probability $m(ka, x, q)$, where

$$m(ka,x,q) = \begin{cases} \lambda_u \, u_{ka} & \text{if } x = u \\ \lambda_e \, e_{ka,0} \; n^q_1 & \text{if } x = 0 \\ \lambda_e \, e_{ka,\ell t} \; n^q_2 & \text{if } x = \ell t \end{cases}$$

Conditional on meeting, the team extract a fraction $1-\gamma$ of the surplus $V^p_{ka,0} - V^p_{0,0} - v(ka, x, q)$, where $V^p_{ka,0}$ is the joint value of the team and the worker, $V^p_{0,0}$ is the outside option of the team, and $v(ka, x, q)$ is the outside option of the meeting worker, which is defined as below.

$$v(ka,x,q) = \begin{cases} U_{ka} & \text{if } x = u \\ \widehat{V}^q_{ka,0} - \widehat{V}^q_{0,0} & \text{if } x = 0 \\ \widehat{V}^q_{ka,\ell t} - \widehat{V}^q_{\ell t,0} & \text{if } x = \ell t \end{cases}$$

4.4.3 Team with one worker

The joint value $V^p_{ka,0}$ of a team of type p and a worker of type ka is such that

$$\begin{aligned} V^p_{ka,0} = \underbrace{f(p,ka,0)}_{\substack{\text{production} \\ \text{with } ka}} + \delta\, \mathbb{E}_{ka'} \Bigg\{ & \underbrace{\sigma[V^p_{0,0} - \widehat{V}^p_{ka',0}]}_{\text{worker leaves market}} + \underbrace{\lambda_s[V^p_{0,0} + U_{ka'} - \widehat{V}^p_{ka',0}]}_{\text{match separates}} \\ & + \sum_{\ell t,x,q} m(\ell t,x,q)(1-\gamma)\max\left\{\widehat{V}^p_{ka',\ell t} - \widehat{V}^p_{ka',0} - v(\ell t,x,q), 0\right\} \\ & + \sum_{z,q} \underbrace{\lambda_e p(z,q)}_{\substack{\text{prob worker} \\ \text{is poached by} \\ \text{a firm (z,q)}}} \gamma \max\Big\{ \underbrace{v(ka',z,q)}_{\substack{\text{outside option} \\ \text{induced by} \\ \text{poaching firm}}} - \underbrace{(\widehat{V}^p_{ka',0} - V^p_{0,0})}_{\substack{\text{firm's valuation} \\ \text{of worker } ka'}}, 0\Big\} + \widehat{V}^p_{ka',0} \Bigg\} \end{aligned} \tag{14}$$

In the current period, the joint unit produces $f(p,ka,0)$. In the next period, the following happens. At the beginning, with probability $g(k'|k,x)$, the human capital evolves from k to k'. I denote $ka' = (k',a)$ for simplification. At the entry-exit stage, the worker exits the labor market with probability σ, which results in an empty firm continuation value of $V^p_{0,0}$. A worker can also move into unemployment with an exogenous probability λ_s, which leads to a continuation of $V^p_{0,0} + U_{ka'}$. Also, the firm can contact a worker $(\ell t,x,q)$ with probability $m(\ell t,x,q)$, as specified above, and firm extracts a fraction $1-\gamma$ of the surplus between the firm and the contacted worker, which is $V^p_{ka',\ell t} - V^p_{ka',0} - v(\ell t,x,q)$. Lastly, the worker ka can also be poached by a firm of productivity q and vacancy status z with probability $\lambda_e p(z,q)$, where

$$p(z,q) = \begin{cases} n^q_0 & \text{if } z = (0,0) \\ n^q_1\, e^q_{ka,0} & \text{if } z = (ka,0) \\ n^q_2\, e^q_{ka,\ell t}/2 & \text{if } z = (ka,\ell t) \end{cases}$$

Conditional on meeting a poaching team, the coalition between the team and the current worker ka extracts a fraction γ of the surplus $v(ka', z, q) - (\widehat{V}^p_{ka',0} - V^p_{0,0})$. I denote $v(ka', z, q)$ as the marginal value of worker ka to firm (q, z), which is equivalent to the difference between the joint value of the match with worker ka and the joint value of the match without the worker, as shown below, and $\widehat{V}^p_{ka_+,0} - V^p_{0,0}$ is the marginal value of the worker to the current coalition.[29]

$$
v\left(ka, z, q\right) = \begin{cases} \widehat{V}^q_{ka,0} - V^q_{0,0} & \text{if } z = (0,0) \\ \widehat{V}^q_{ka,\ell t} - \widehat{V}^q_{\ell t,0} & \text{if } z = (\ell t, 0) \\ \max_{i \in \{1,2\}} \left\{ \widehat{V}^q_{ka,\ell t_i} + U_{\ell t_i} - \widehat{V}^q_{\ell t_1, \ell t_2} \right\} & \text{if } z = (\ell t_1, \ell t_2) \end{cases}
$$

4.4.4 Team with two workers

The joint value $V^p_{ka,\ell t}$ of a team type p, a worker ka and a worker ℓt is such that

$$
\begin{aligned}
V^p_{ka,\ell t} = \underbrace{f(p, ka, \ell t)}_{\substack{\text{production with} \\ ka \text{ and } \ell t}} + \delta \mathbb{E}_{ka',\ell t'} \Bigg\{ & \sigma(\widehat{V}^p_{ka',0} - \widehat{V}^p_{ka',\ell t'}) + \lambda_s(\widehat{V}^p_{ka',0} + U_{ka'} - \widehat{V}^p_{ka',\ell t'}) \\
& + \underbrace{\sigma(\widehat{V}^p_{\ell t',0} - \widehat{V}^p_{ka',\ell t'}) + \lambda_s(\widehat{V}^p_{\ell t',0} + U_{\ell t'} - \widehat{V}^p_{ka',\ell t'})}_{\text{worker } \ell t \text{ leaves market or match separates}} \\
& + \sum_{s\tau, x, q} m(s\tau, x, q)(1-\gamma) \cdot \\
& \qquad \max\Big\{0, \underbrace{\max\{\widehat{V}^p_{ka', s\tau} + U_{\ell t'}, \widehat{V}^p_{\ell t', s\tau} + U_{ka'}\}}_{\text{which one to replace if } s\tau \text{ is hired}} - \widehat{V}^p_{ka',\ell t'} - v(s\tau, x, q)\Big\} \\
& + \sum_{z,q} \lambda_e p(z,q) \gamma \max\{v(ka', z, q) - (\widehat{V}^p_{ka',\ell t'} - \widehat{V}^p_{\ell t',0}), 0\} \\
& + \sum_{z,q} \underbrace{\lambda_e p(z,q) \gamma \max\{v(\ell t', z, q) - (\widehat{V}^p_{ka',\ell t'} - \widehat{V}^p_{ka',0}), 0\}}_{\text{worker } \ell t \text{ is poached by a firm } (z,q)} + \widehat{V}^p_{ka',\ell t'} \Bigg\}
\end{aligned} \tag{15}
$$

For the sake of brevity, I only discuss the differences between expressions (14) and (15). First, the worker ℓt can exit the labor market or move into unemployment, leaving the worker ka the only worker in the team. Second, when the team meets a worker of type $s\tau$,

[29] The last equation shows that the poaching team needs to break up with one of its current employees to make room for the new hire, which will be useful for $V^p_{ka,\ell t}$ below.

it captures a fraction $1-\gamma$ of surplus. However, upon hiring, the team also has to decide on the worker to replace $s\tau$. Finally, the other worker ℓt in the team can also be poached by another team.

4.4.5 Unemployed worker

The value of unemployment U_{ka} to a worker ka is given by

$$U_{ka} = f(0, ka, 0) + \delta \mathbb{E}_{ka'} \left\{ \sigma[0 - U_{ka'}] + \sum_{z} \lambda_u p(z, q) \gamma \max\{v(ka', z, q) - U_{ka'}, 0\} + U_{ka'} \right\} \tag{16}$$

In the current period, the worker enjoys home production of $f(0, ka, 0)$. In the next period, the worker leaves the market with probability σ, or contacts a team with vacancy status z and productivity q with probability $\lambda_u p(z, q)$, in which case, the worker extracts a fraction γ of surplus $v\left(ka', z, q\right) - U_{ka'}$.

4.4.6 Equilibrium

A *Stationary Equilibrium* is a list of value functions $\{U, V\}$ and a distribution of workers across employment states $\{u, e\}$ such that (i) the value functions satisfy conditions (11)-(16) given $\{u, e\}$, and (ii) the distribution $\{u, e\}$ is stationary given the transitions probabilities implied by the value functions. The law of motion of the stationary distributions of $\{u, e\}$ is described in detail in Appendix C. Note that the wage *per se* does not affect the equilibrium as it is an internal transfer between team and worker. Nevertheless, it is important for estimation. The next section discusses how wages are determined in detail.

4.5 Wage setting

The wage is determined through bargaining on the surplus of a match (Cahuc, Postel-Vinay and Robin, 2006; Dey and Flinn, 2005). Unlike the standard bilateral bargaining, the unit of production in the model can consist of up to three agents: the firm and the two workers in the same team. I employ a three-way bargaining protocol building on the previous multilateral bargaining literature (Lentz and Mortensen, 2012; Brügemann, Gautier and Menzio, 2019; Elsby and Gottfries, 2022).

Formally, worker ka working in a firm p with a coworker ℓt have a (net present) value, $W^p_{ka,\ell t}(w)$, if she receives a wage w:

$$W^p_{ka,\ell t}(w, A_{ka}) = A_{ka} + \gamma \left(V^p_{ka,\ell t} - V^p_{0,\ell t} - A_{ka}\right) \tag{17}$$

where the total surplus is the joint value of the entire production unit (i.e., firm p, worker ka, and worker ℓt) minus the joint value of the production unit without worker ka minus the outside option of worker ka. I denote the first part of the surplus as the *marginal value* of the worker ka to the production unit, i.e., $MV^p_{ka,\ell t} = V^p_{ka,\ell t} - V^p_{0,\ell t}$, which is also the maximum value the production unit can offer to worker ka. I assume that the unit *remembers* the past outside option A_{ka} until it is updated with better outside offers during the job tenure (similar to Cahuc et al., 2006).

Upon matching, the starting wage is set according to equation (17). Within a job, wages can be changed under three main circumstances. First, when a worker is poached by a good offer, which exceeds the current value of the worker but is not good enough to separate the worker, the wage increases due to the improved outside option values (i.e., A_{ka} improves). Second, when a worker's human capital increases, the wage increases as the joint value of the production unit increases (i.e., $V^p_{ka,\ell t}$ rises).[30] Finally, wages can change when the coworker changes – leaves, joins, or is replaced – or the human capital of the coworker improves (i.e., the marginal value of worker ka, $MV^p_{ka,\ell t} = V^p_{ka,\ell t} - V^p_{0,\ell t}$, changes).[31] In particular, if coworker quality improves, both the value of the production unit, $V^p_{ka,\ell t}$ and the value without worker ka, $V^p_{0,\ell t}$ increases. However, if the team production complementarity is estimated to be supermodular, the increase in the former term likely exceeds that in the latter one, leading to a wage rise.

The wage setting is particularly relevant in my setting. First, it becomes the standard two-way bargaining if the involved parties are only a firm and a worker. Second, the value-splitting protocol ensures that both workers are equally treated as the (net present) value of a worker's wage is simply a function of his marginal value and the outside option. Finally, the wage setting gives predictions that are consistent with the empirical

[30] It might decrease in one rare case. The human capital increase can also increase the unemployment value. If such an increase is higher than the increase in the joint value of production, wages can decrease. In the estimated model, I do not observe such cases.

[31] One special case is that a worker's outside option value might be larger than her marginal value when her coworker leaves. In other words, the outside option value exceeds the maximum the firm can offer. In that case, the unit allows the worker to extract all the match surplus by lowering her down to match her marginal value.

findings. For one, it predicts that wages increase as human capital increases. For the other, it predicts that an improvement in coworkers' quality increases a worker's own wages.

4.6 Wage equations

Before delving into the value functions, It may be useful to introduce a few additional notation. As mentioned in Section 4.5, a worker's (net present) value of wage w can be updated within a job. The updating happens when a worker's outside option improves, the human capital increases, or her coworker changes.

For brevity, I illustrate the value updating using a case when a worker ka is employed with a coworker ℓt in firm p, which corresponds to a value $W^p_{ka,\ell t}(A)$.[32] A worker's continuation value for the next period is denoted as $\widehat{W}^p_{ka',\ell t'}(A')$ as workers' human capital can change at the beginning of the period. At the same time, it improves the unemployment outside option, which may be larger than the outside option A that the production unit remembers from the last period, especially when the worker was previously hired from unemployment. Moreover, the outside option cannot be higher than the maximum the production unit can offer, which is theoretically possible but not observed in the estimated model. As a result, the updated outside option A' is as

$$A' = \min\{\max\{A, U_{ka'}\}, \widehat{V}^p_{ka',\ell t'} - \widehat{V}^p_{0,\ell t'}\}.$$

Lastly, the worker can, in theory, find her value smaller than the unemployment value $U_{ka'}$. In this case, the worker will move to unemployment. Similar to equation (11), I denote

$$\widehat{W}^p_{ka',\ell t'}(A') = \max\{W^p_{ka',\ell t'}(A'), U_{ka'}\}.$$

I now lay out the detailed expression of the wage equations below.

[32] With abuse of notation, I denote worker's value as $W^p_{ka,\ell t}(A)$ without referring to the wages. Because a worker is risk-neutral, wages are simply the difference between the current worker's net present value and the discounted worker's future net present value (see the wage equations below).

4.6.1 Wage without coworkers

The (net present) value of a worker ka with wage w in firm p with no teammate with an outside option value A is expressed below.

$$
\begin{aligned}
W^p_{ka,0}(w,A) = \underbrace{w}_{\substack{\text{flow}\\\text{wage}}} + \delta\, \mathbb{E}_{k'} \Bigg\{ & \underbrace{\widehat{W}^p_{ka',0}(A')}_{\substack{\text{value updates}\\\text{when h.c. }\uparrow}} + \sigma[0 - \widehat{W}^p_{ka',0}(A')] + \lambda_s[U_{ka'} - \widehat{W}^p_{ka',0}(A')] \\
& + \sum_{\ell t,x,q} \underbrace{m(\ell t,x,q)\ h^p_{ka',0}(\ell t,x,q)}_{\substack{\text{meet a worker } \ell t\\\text{and hire her with prob } h}} [\underbrace{\widehat{W}^p_{ka',\ell t}(A')}_{\substack{\text{value updates}\\\text{if } \ell t \text{ is hired}}} - \widehat{W}^p_{ka',0}(A')] \qquad (18) \\
& + \sum_{z,q} \lambda_e p(z,q) \Big[\max \Big\{ \widehat{W}^p_{ka',0}(A'), \min\{(1-\gamma)\nu(ka',z,q) + \gamma(\widehat{V}^p_{ka',0} - V^p_{0,0}), \\
& \qquad \gamma\nu(ka',z,q) + (1-\gamma)(\widehat{V}^p_{ka',0} - V^p_{0,0})\} \Big\} - \widehat{W}^p_{ka',0}(A') \Big] \Bigg\}
\end{aligned}
$$

As the worker is risk neutral, wage w is the flow value that is equivalent to the difference between the current worker's value and the expected future values. Therefore, in the current period, the worker receives a wage of w. In the next period, human capital can improve at the beginning, so the continuation value becomes $\widehat{W}^p_{ka',0}(A')$. A worker might leave the market with zero continuation value with probability σ or move to unemployment with a continuation value of $U_{ka'}$ with probability λ_s. With probability $m(\ell t,x,q)$, the team meets a worker of type ℓt and employment status (x,q). Conditional on meeting, the team hires the worker with probability $h^p_z(\ell t,x,q)$, where, in the current case, $z = (ka',0)$. That is, the firm p and worker ℓt mutually agree to match if the match surplus in firm p is larger than that in firm q (if employed) or the unemployment value.

$$
h^p_z(\ell t,x,q) = \begin{cases} 1 & \text{if } \nu(\ell t,z,p) > v(\ell t,x,q) \\ 0 & \text{otherwise} \end{cases}
$$

If the worker is hired, the continuation value is changed $\widehat{W}^p_{ka',\ell t}(A')$. With probability $\lambda_e p(z,q)$, the worker is poached by a firm type (z,q). The following cases happen depending on the value of the poaching offer. I denote as $v_q = \nu(ka',z,q)$ and $v_p = \widehat{V}^p_{ka',0} - V^p_{0,0}$ as the poaching value and the maximum the incumbent firm can offer (i.e., the marginal value of worker ka' to the production unit), respectively. First, if $v_q > v_p$, then the worker moves to the poaching firm with value $\gamma v_q + (1-\gamma)v_p$. In that case, her next period's outside option value is also updated to v_p. Second, if $v_q \leq v_p$, but it is large enough to bind the outside option A', i.e., $v_q > A'$, then the worker's value increases to $\gamma v_p + (1-\gamma)v_q$. As a result, the outside option value updates to v_q. Finally, if $v_p \leq A'$, the outside option does not bind, and nothing changes.

4.6.2 Wage with coworkers

The value of a worker ka receiving wage w in firm p with a coworker ℓt and an outside option value A is expressed below.

$$
\begin{aligned}
W^p_{ka,\ell t}(w, A) = w + \delta \mathbb{E}_{ka',\ell t'} \Bigg\{ & \underbrace{\widehat{W}^p_{ka',\ell t'}(A')}_{\substack{\text{value updates when} \\ \text{one or both h.c. } \uparrow}} + \sigma[0 - \widehat{W}^p_{ka',\ell t'}(A')] + \delta[U_{ka'} - \widehat{W}^p_{ka',\ell t'}(A')] \\
& + [\sigma + \delta + \sum_{z,q} \lambda_e p(z,q) h^q_z(\ell t', ka', p)][\underbrace{\widehat{W}^p_{ka',0}(A')}_{\substack{\text{value updates} \\ \text{if } \ell t' \text{ leaves}}} - \widehat{W}^p_{ka',\ell t'}(A')] \\
& + \sum_{s,x,q} m(s,x,q) h^p_{ka',\ell t'}(s,x,q) \Big[r^p_{ka',\ell t'}(s, ka')\, U_{ka'} + \underbrace{r^p_{ka',\ell t'}(s,\ell t')\, \widehat{W}^p_{ka',s}(A')}_{\substack{\text{value updates when} \\ \text{coworker replaces}}} - \widehat{W}^p_{ka',\ell t'}(A') \Big] \\
& + \sum_{z,q} \lambda_e p(z,q) \Big[\max \Big\{ \widehat{W}^p_{ka',\ell t'}(A'), \min \Big\{ (1-\gamma)\nu(ka',z,q) + \gamma(\widehat{V}^p_{ka',\ell t'} - \widehat{V}^p_{0,\ell t'}), \\
& \qquad \gamma \nu(ka',z,q) + (1-\gamma)(\widehat{V}^p_{ka',\ell t'} - \widehat{V}^p_{0,\ell t'}) \Big\} \Big\} - \widehat{W}^p_{ka',\ell t'}(A') \Big] \Bigg\}
\end{aligned} \tag{19}
$$

For brevity, I only describe the terms that are substantially different from equation (18). In the next period, both workers might increase their human capital with some probability, so the continuation value becomes $\widehat{W}^p_{ka',\ell t'}(A')$. With probability $[\sigma + \delta + \sum_{z,q} \lambda_e p(z,q) h^q_z(\ell', k', p)]$, the coworker ℓ either leaves the market, or exogenously moves to unemployment, or moves by another firm, which leads to a continuation value of worker ka as $\widehat{W}^p_{ka',0}(A')$. With probability $m(s,x,q) \cdot h^p_{ka',\ell t'}(s,x,q)$, the team meets and hires a worker (s,x,q). Upon hiring, the team has to decide to replace one current worker. Worker ka is replaced by worker s with probability $r^p_z(s, ka)$, where $z = (ka, \ell t)$,

$$r^p_z(s, ka) = \begin{cases} 1 & \text{if } \widehat{V}_{s,\ell t} + U_{ka} > \widehat{V}_{ka,s} + U_\ell \\ 1/2 & \text{if } \widehat{V}_{s,\ell t} + U_{ka} = \widehat{V}_{ka,s} + U_\ell \\ 0 & \text{if } \widehat{V}_{s,\ell t} + U_{ka} < \widehat{V}_{ka,s} + U_\ell \end{cases}$$

If the worker is replaced, then the worker moves to unemployment with a continuation value of $U_{ka'}$. If the coworker is replaced, the worker's value is updated to $\widehat{W}^p_{ka',s}$.

4.6.3 Discussion

The wage setting and model prediction is broadly consistent with the existing literature as well as the empirical findings in Section 3. First, when a worker's human capital increases, the firm pays a higher wage due to her increased contribution to the joint production unit. This feature of the wage setting is also similar to settings where a worker bargains on the rental price of her human capital as used in Taber and Vejlin (2020): given a manually agreed rental price, a worker's wage increases when human capital increases. Second, when a worker's coworker quality increases, depending on the degree of complementarity between worker and coworker, the wage is theoretically possible to increase or decrease. If the production complementarity is modular, the wage setting would predict that a worker will accept a wage cut to join a team with a better coworker in expecting to learn more from a better coworker, which mirrors the wage setting in a competitive model used Jarosch et al. (2021). If the production is supermodular, the model would likely predict a wage rise when the coworker quality increases as gain from the production complementarity can exceed the potential wage cut stemming from learning. This prediction, however, deviates somewhat from the wage setting used in Herkenhoff et al. (2018), where they assume that the wage would either stay unchanged or potentially decrease. Nevertheless, my model prediction is consistent with the empirical findings where I find an increased coworker quality can lead to a non-trivial wage gain as shown in Table 2.

5 Estimation

Throughout the model, I have the following parameters to estimate, as collected in Θ.

$$\Theta = \{\underbrace{\delta, \sigma, \lambda_s, \lambda_u, \lambda_e}_{\text{labor market}}, \underbrace{\Phi(.), b_1, b_2, \beta_s, \beta_c}_{\text{worker side}}, \underbrace{\Psi(.), \rho, \eta, f_0, \gamma}_{\text{firm side}}\}$$

Some of the parameters are estimated *outside* the model following the standard literature. I choose the length of a period in the model to be one month. In particular, I set the preference monthly discount rate δ to be 0.9957, corresponding to an annual discount rate of 0.95. The probability of retiring from the market is set by allowing workers to spend 35×12 months working on average, i.e., $\sigma = 1/(35 \times 12) = 0.00238$.

The remaining parameters $\theta = \{\lambda_s, \lambda_u, \lambda_e, \chi_w, b_1, b_2, \beta_0, \beta_c, \chi_f, \rho, \eta, f_0, \gamma\}$ are estimated *internally* in the model using indirect inference (Gourieroux et al., 1993). The corresponding estimator is shown below.

$$\hat{\theta} = \arg\min_{\theta} \left[\mathbf{m}(\theta_0) - \mathbf{m}(\theta)\right]' D \left[\mathbf{m}(\theta_0) - \mathbf{m}(\theta)\right]$$

where $\mathbf{m}(\theta_0)$ and $\mathbf{m}(\theta)$ are vectors of data moments and auxiliary statistics – described in detail in the following subsection – computed with real data and the simulated data using model parameters θ, respectively. D is a diagonal weighting matrix whose diagonal is the inverse of the squared value of the data moments.[33]

Standard errors are calculated using the standard sandwich formula, where the gradients are calculated numerically, and the variance-covariance matrix is obtained by bootstrapping. Specifically, I randomly draw samples of workers with their entire working histories with replacement and re-generate the auxiliary moments 5,000 times using bootstrapped samples.

[33] The approach deviates from the most common approach where the weight is the inverse of the variance of the data moments. In practice, this approach ensures that all the moments are equally weighted. It is straightforward to re-estimate the model using the standard weighting matrix, but given the good fit of the model, the weighting matrix should not matter too much. A similar approach has also been used in other studies that use large-scale population data, e.g., Denmark in Taber and Vejlin (2020) and the United States in Wallskog (2022).

5.1 Identification

In this section, I describe the key moments for the identification of the parameters. In particular, I use the same number of auxiliary statistics as the number of parameters to be estimated. I also add one additional moment, which is my key objective of interest – the wage variance.

Table 5: Identification Map

Auxiliary moment	Structural parameter
UE rate	Unemployed job arrival rate, λ_u
EE rate	Employed job arrival rate, λ_e
EU rate	Exogenous separation rate, λ_s
Wage correlation, $\text{cov}(y_{it}, \bar{y}_{-i,t})$	Team production function, ρ
Between firm wage variance	Firm production function, η
Lagged peer effect for job movers	Learning from coworkers, β_c
Average wage growth	Learning by doing β_0
Coefficient on tenure^2	Wage bargaining power γ
Coefficient on experience	Human capital polynomial 1, b_1
Coefficient on experience^2	Human capital polynomial 2, b_2
Worker FEs variance	Firm type distribution shape parameter, χ_f
Firm FEs variance	Ability type distribution shape parameter, χ_w
Average wage levels	Production scalar, f_0

Job mobility To account for monthly job mobility, I extend the annual dataset described in Section 2 to a monthly frequency by utilizing the start and ending date of each job contract. From there, I can construct the monthly unemployment-to-employment (UE), employer-to-employer (EE), and employment-to-unemployment (EU) transition rates, which help identify the offer arrival rate of the unemployed worker, λ_u, the offer arrival rate of the employed worker, λ_e, and the exogenous separation rate, λ_s, respectively.

Mincerian regression I estimate the following Mincerian regression using the non-left censored workers to capture regress log wages on the experience and tenure, which are measured by the number of months divided by twelve.

$$y_{ist} = \xi_{is} + \underbrace{\xi_{e1} exp_{ist} + \xi_{e2} exp_{ist}^2}_{\text{experience}} + \underbrace{\xi_{t2} ten_{ist}^2}_{\text{tenure}} + \epsilon_{ist}, \tag{20}$$

where ξ_{is} is the worker-spell fixed effect. I match ξ_{e1} and ξ_{e2} for the identification of human capital process b_1 and b_2, and the tenure coefficient ξ_{t2} for the identification of bargaining parameter γ. The mapping between experience and the human capital process is straightforward. On the other hand, the bargaining parameter, γ, will determine the share of the total surplus extracted at the beginning of a new match. Therefore, a higher γ will result in a lower left-over surplus to be extracted within a match or a lower within-job wage growth. Ideally, one could capture it by having the coefficient on tenure. However, as pointed out by Altonji and Shakotko (1987) and Topel (1991), experience and tenure are perfectly correlated within a spell. One cannot identify the coefficient on tenure directly since I have included the worker-spell fixed effects. Instead, use the coefficient on tenure squared, which measures the rate of within-job wage stops increasing, to pick up the importance of bargaining.

AKM regression I target the distribution of the worker and firm heterogeneity, i.e., χ_w and χ_f by targeting the distribution of the estimated worker and firm fixed effects from the AKM regression (Abowd et al., 1999) as shown below.

$$y_{it} = \xi_i + \xi_{j(it)} + \xi_t + \epsilon_{it},$$

where ξ_i and $\xi_{j(it)}$ are worker and firm fixed effects, respectively, and ξ_t is the year fixed effects. The variance of ξ_i and ξ_j capture the pre-market skills or ability, χ_w, and firm heterogeneity, χ_f, in the model, respectively. One concern with using the AKM moments is that it might induce a heavy computation due to high dimensions of ξ_i and ξ_j. I utilize the sparse matrix operation to mitigate the massive memory consumption, as the corresponding dummy matrices of the fixed effects contain a substantial fraction of zeros, and then apply the conjugate gradients method for fast computation of the estimation.

Between-firm wage variation I target the firm production complementarities, η, to the between-firm variation. The idea is that if the firm production complementarity is large, then the between-firm variation is amplified.

Lagged peer effects Using a similar idea in equation (3), I explore the effects of past coworkers' wages on the current wages, as shown in the equation below. If I consider wage as a (noisy) measure of productivity, my past coworker's productivity can impact my current productivity through human capital accumulation. So, ξ_{c1} will have information on the coworker learning parameter.

$$y_{it} = \xi_0 + \xi_{c0}\, \bar{y}_{-i,t} + \xi_{c1}\, \bar{y}_{-i,t-1} + \varepsilon_{it} \tag{21}$$

To make sure this channel is as clear as possible, the sample only contains the movers who went through unemployment so that, through the lens of the model, the past coworkers could affect the current productivity only through human capital.[34]

Average coworker's wages Since the team production complementarity, ρ, governs the ex-post sorting between workers and coworkers, the wage correlation between them provides sufficient information for identifying ρ. In particular, I utilize the fact that the simulated data replicates the real data structure and use the average coworkers within a firm to compute the statistic.

Wage level and wage growth Finally, I identify the production scalar f_0 by using the average log weekly wage, as it only affects the production levels in the economy. I target the learning-by-doing parameter β_s by the average annual log wage growth from age 21 to age 50 in the data.

Table 5 provides an overview of the identification idea. It lists the auxiliary statistic that also which structural parameter each one primarily helps identify. While all structural parameters are determined by all auxiliary parameters, taking everything else as given, the structural parameter in the model is driven by changes in the auxiliary parameter next to it.

5.2 Estimates and model fit

Table 6 shows the comparison between the moments generated from the simulated data and the data. As one can see, the fit is very good. This is partially due to the fact I use a similar number of auxiliary moments to match the number of model parameters, although such a fit is not guaranteed due to the complexity of the underlying model

[34] In the data, I identify mobility through unemployment by exploring the panel at the monthly frequency. An employment-unemployed-employment mover is defined as a worker who moves to another firm through an unemployment spell that lasts at least six months.

and the nonlinearity of the wage equation. The fit of the coefficient on tenure squared is slightly off, but this is probably due to the fact the model generates slightly more volatile wage changes within a job than what is observed in the data.

Table 6: Model fit: the moment values estimated from the simulated data and real data

	Simulated	Data	(s.e.)
Monthly UE rate	0.103	0.100	(0.001)
Monthly EU rate $\times 100$	0.766	0.761	(0.003)
Monthly EE rate $\times 10$	0.207	0.204	(0.000)
Log weekly wages, mean	6.562	6.584	(0.001)
Log weekly wages, variance	0.186	0.182	(0.002)
Mincer regression, experience coeff, $\xi_{s1} \times 100$	2.851	2.987	(0.038)
Mincer regression, experience^2 coeff, $\xi_{s2} \times 1000$	-0.412	-0.395	(0.021)
Mincer regression, tenure^2 coeff, $\xi_{t2} \times 1000$	-0.013	-0.088	(0.032)
AKM regression, variance of worker fixed effects	0.112	0.103	(0.002)
AKM regression, variance of firm fixed effects	0.039	0.037	(0.001)
Annual log wage growth (in log points) $\times 100$	1.192	1.260	(0.015)
Between-firm wage variance	0.079	0.081	(0.001)
Lagged peer effect ξ_{c1}	0.172	0.173	(0.006)
Wage correlation between workers and coworkers	0.651	0.660	(0.002)

Notes: The standard errors of the data moments, in parentheses, are computed via bootstrapping the sample 5,000 times. Specifically, I recalculate the moments for each bootstrapped sample, where I randomly draw, with replacement, samples of workers with their entire working histories.

Estimates The corresponding parameter estimates are presented in Table 7. In general, the interpretation of the estimates is not interesting unless it is explored in the context of counterfactual exercises. I mainly discuss two sets of parameters here – production complementarity and learning – as they are important in the counterfactual exercises below.

First, the team production parameter ρ is estimated to be around 0.470, which is smaller than 1, meaning that team production is supermodular. The firm production parameter η is estimated to be 0.751, which is also smaller than 1 and suggests that the firm production is supermodular. Both parameters generate positive sorting in the labor market between worker and coworker as well as between worker and firm, respectively. The estimated model shows that the correlation between a worker's productivity, π_{ka}, and her coworker's productivity, $\pi_{\ell t}$, is 0.451, while the correlation between worker productivity and firm productivity, p, is around 0.093. These results are consistent with the empiri-

Table 7: Parameter Estimates

Parameter	Description	Estimate	(s.e.)
ρ	Team production function	0.470	(0.001)
η	Firm production function	0.751	(0.002)
$\beta_c \times 100$	Learning from coworkers	0.460	(0.003)
$\beta_s \times 100$	Learning by doing	0.391	(0.009)
γ	Wage bargaining power	0.712	(0.015)
b_1	Human capital polynomial 1	0.273	(0.001)
b_2	Human capital polynomial 2	-0.013	(0.002)
λ_u	Unemployed job arrival rate	0.181	(0.002)
λ_e	Employed job arrival rate	0.052	(0.001)
λ_s	Exogenous separation rate	0.028	(0.001)
χ_f	Firm type distribution shape parameter	0.393	(0.068)
χ_w	Ability type distribution shape parameter	0.412	(0.101)
f_0	Production scalar	110.71	(0.286)

Notes: Standard errors are calculated using the standard sandwich formula, where the gradients are calculated numerically, and the variance-covariance matrix is obtained by bootstrapping the sample 5,000 times.

cal evidence that suggests that the labor market sorting is primarily driven by coworker sorting instead of the sorting between workers and firms.

The parameter of learning from coworkers, β_c, is estimated to be 0.0046, while the estimated parameter of learning by doing is 0.0039. Given the average productivity in the estimated model is about 1.57, the probability of increasing human capital due to knowledge spillover from an average coworker can be as high as 0.0072. In other words, working with an average coworker can almost double the probability of human capital accumulation, which is in the range of the literature (e.g., Herkenhoff et al., 2018; Jarosch et al., 2021). The result is also consistent with the empirical findings in subsection 3.4, where I find the wage return to past coworkers' quality is almost twice the return to experience.

Validation To shed light on the validation of the model estimates, I have estimated the peer regression model in equation 1 using the simulated data and conducted the wage variance decomposition exercise as depicted in Table 4, which is duplicated in Table 8 Column 1. As shown in Column 2 of Table 8, the wage decomposition exercise using simulated data is very similar to empirical ones, despite that these moments are not directly used in the estimation of the model.

Table 8: Non-targeted moments: cross-sectional wage variance decomposition

	Data (1)	Simulated data (2)
Variance decomp		
Worker effect	47.1%	51.2%
Firm effect	13.3%	15.0%
Coworker effect	11.2%	12.3%
Corr(workers, firms)	0.080	0.035
Corr(workers, coworkers)	0.596	0.530

Notes: Column 1 replicates what is reported in Table 4; and Column 2 reports the wage variance decomposition of the peer regression model in equation 1. Again, I adopt a similar method as Kline et al. (2020) to correct the bias stemming from the limited mobility in the data for all the estimates.

6 How does coworkers impact inequality?

In this section, I present structural decomposition exercises to understand how coworkers impact inequality, in particular, lifetime income inequality, and shed light on the relevant importance of the two mechanisms of coworkers – production complementarity and learning – and that of the firms. Moreover, I conduct a policy counterfactual to understand how exposure to high-quality coworkers in the early career can impact the a worker's lifetime income and human capital accumulation.

6.1 Lifetime income inequality

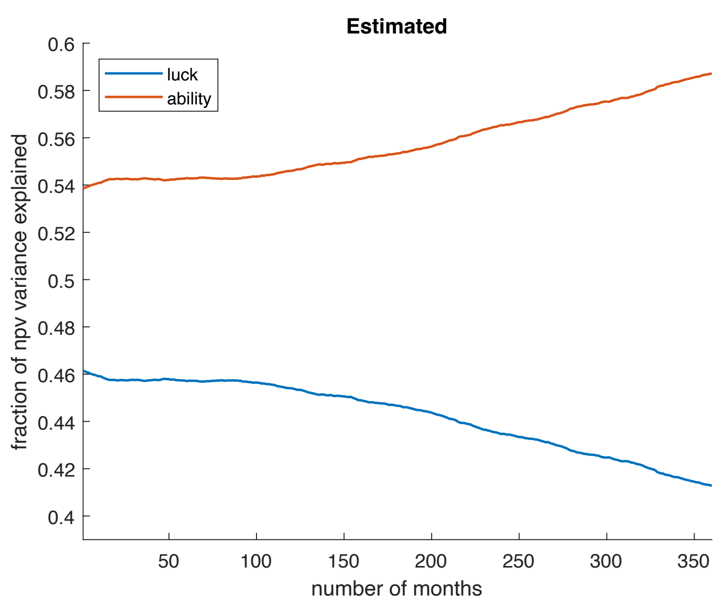
As discussed in Flinn (2002), Bowlus and Robin (2004), and Bowlus and Robin (2012), cross-sectional wage inequality misses the important life-cycle dynamics of wages, and one should consider lifetime income inequality as the underlying welfare measure of a worker. This is particularly relevant in a setting with coworkers. In particular, if a good worker not only works with good coworkers but also learns from them, then over the working life, there can be a snowball effect that amplifies the income dispersion among different workers. Moreover, Cunha and Heckman (2007, 2008) shows that half of the inequality in the present value of lifetime earnings is determined by ability, and the rest is determined by uncertainty or luck. However, what contributes to the luck component is less explored. In this section, I explore how coworkers contribute to the luck component.

To do this, I simulate a cohort of workers who enter the labor market at the same time, who are only different in terms of their ability a, and simulate their entire working life until they leave the market. I calculate PV_τ the *net present value* of their 30 years' future earnings as a measure of lifetime income after they work for τ months, as shown below.[35] For example, PV_0 is discounted lifetime income calculated at the time when they enter the labor market, and PV_{120} is the net present value of the future 30 years' earnings 10 years after they enter the labor market.

$$PV_\tau = \sum_{m=1+\tau}^{M} \frac{w_m}{(1+\delta)^{m-1-\tau}}$$

where M is the total months in 30 years, i.e., 360 months, $\delta = 0.9957$ is a monthly discount rate used in the model, corresponding to an annual discount rate of 0.95, and w_m is the monthly earnings and is equal to zero during unemployment. I denote $z_\tau = log(PV_\tau)$ as the log value of the net present value at month τ.

Figure 3: Life-cycle log wage variance decomposition



To decompose the (log) net present value into the ability and luck component, I utilize

[35] All the calculations are conditional on that a worker is still in the market 30 years from month τ. As discussed in Section 5, the exogenous retirement rate is set as $\sigma = 1/(35 \times 12) = 0.00238$ so that workers, on average, leave the market after 35 years.

the law of iterated expected in the variance of z in terms of ability a. Specifically,

$$\mathrm{Var}[z_\tau] = \underbrace{\mathrm{Var}_a\left[\mathbb{E}\left(z_\tau|a\right)\right]}_{\text{ability component}} + \underbrace{\mathbb{E}_a\left[\mathrm{Var}\left(z_\tau|a\right)\right]}_{\text{luck component}}$$

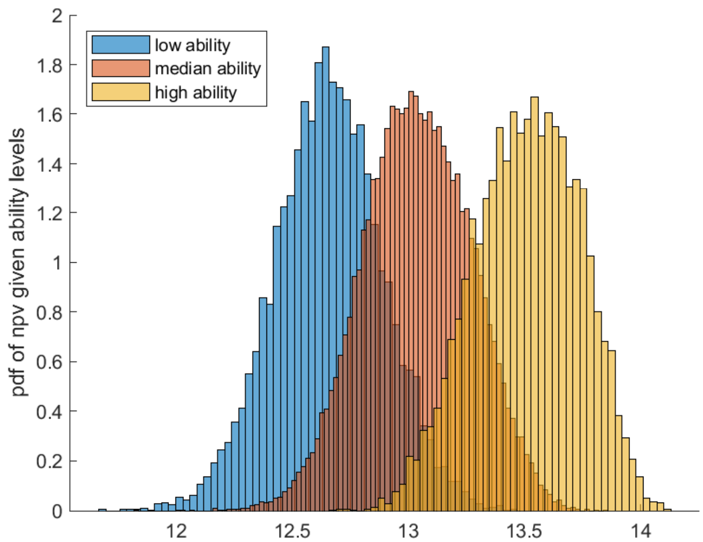
where the first between-ability term captures the variance of log income explained by workers' innate ability, and the within-ability term captures the luck/uncertainty. As shown in Figure 3, the ability component contributes to 53 percent of the overall income inequality at the start of a career, and the luck component is a sizable 47 percent, which is similar to what Cunha and Heckman (2007, 2008) find in the United States. However, when a worker gains experience in the market, the (initial) ability becomes more important, and the luck component decreases.

To shed light on how variable a worker's lifetime income can be among workers who enter the labor market with the same ability, i.e., within-ability variance, $\mathbb{E}_a\left[\mathrm{Var}\left(z_\tau|a\right)\right]$, I plot the distribution of z_0 given ability types. Figure 4 plots the distribution of lifetime income given ability types (high, medium, and low types). Despite that, the high-ability workers earn the most, on average, there is a large variation in the lifetime income within the same ability. In other words, given identical abilities entering the labor market, the income trajectory can be very different purely due to luck. In particular, the luckiest worker (90th percentile) earns twice the lifetime income of the least lucky worker (10th percentile).

This leads to the essential question of this paper: what are the driving sources of the luck part? I address this question by utilizing the features of the models and decomposing the luck component into (i) firm component (firm heterogeneity and firm production complementarity), (ii) coworker production complementarity, (iii) learning from coworkers, (iv) learning by doing, and (v) the remaining part, which mostly captures the changes in bargaining position over the life cycle (i.e., outside option value changes and being on-and-off employment). In principle, there are in total 24 orders ($4! = 24$) one can conduct the decomposition. Following Lentz, Piyapromdee and Robin (2021), I conduct all 24 decompositions (depicted in the Appendix) and take the average of results.

As shown in Table 9. The firm component explains around 17.3 percent of the luck, which is similar to the fraction coworker production complementarity can explain. On the other hand, learning from coworkers plays the most significant role by explaining around 30 percent of the luck component. The result highlights the importance of learning from

Figure 4: within-ability variation for the luck component



coworkers from a life-cycle perspective. If a worker is paired with a good coworker early in the career (by luck), it boosts her human capital accumulation and improves productivity, which will subsequently help match with better coworkers and further accelerate the human capital accumulation, creating a snowball effect and leading to a higher earning trajectory.

Table 9: Decomposition of the luck component

Fraction of the luck component explained by	Fraction (1)
Firm component	17.3%
Coworker production complementarity	16.3%
Coworker learning	29.6%
Learning by doing	16.7%
Remaining part	20.1%

Notes: The results are the average of 24 different orderings of decomposition. The full picture of the decomposition can be found in the Appendix.

6.2 Early exposure to high-quality coworkers

The previous exercise shows that experiencing high-quality coworkers over the working life is essential in shaping lifetime income trajectories. However, how about the initial luck? In particular, How does early exposure to high-quality coworkers impact lifetime income and human capital accumulation? The literature has demonstrated mixed evidence on how firms in a young worker's early career can have a lifetime impact (e.g., Topel and Ward, 1992; Arellano-Bover, 2020), but little is known about how coworkers can play a role.

To explore this question, I artificially match workers with high-quality coworkers only for the first year (i.e., 12 months) of their careers. After the first year, the firm can keep, replace, or separate the worker with respect to the frictional labor market. As shown in 10, overall, the initial exposure to high-quality coworkers increases their lifetime income by around 7 percent and the human capital accumulation by 6 percent, as compared to entering the labor market unemployed. In other words, a worker is willing to pay up to 7 percent of their lifetime income just to be matched with good coworkers in their early career.

The effect is relatively heterogeneous across workers with different initial abilities. In particular, a worker with high ability values the initial match with high-quality coworkers much more. This is because the high-ability worker is a good match to the firm and thus is less likely to separate her even when the binding initial year ends. On the other hand, the low-ability workers are likely to separate from the high-quality coworkers as the firm may find other workers to replace them when the binding first year ends.

Table 10: The effect of early exposure to high-quality coworkers

	Lifetime income	HC after 30 yr
Overall effect	6.8 %	6.0 %
Heterogeneous effect		
High-ability workers	7.8 %	6.9 %
Median-ability workers	6.6 %	6.2 %
Low-ability workers	4.5 %	3.3 %

7 Conclusion

While social interactions are common in the workplace, the role of coworkers is often overlooked in the literature. This paper explores how coworkers affect inequality through labor market sorting and on-the-job learning.

Using matched employer-employee administrative data in a large local labor market in Italy, I document two sets of empirical evidence on coworkers by estimating an econometric specification that incorporates coworkers into an AKM model. First, I find that a one-standard-deviation increase in the contemporaneous coworker quality leads to an 8 percent wage rise, which suggests substantial production complementarity by working with good coworkers. Also, using a subsample of movers, I show that a one-standard-deviation increase in the past coworker quality increases the current wage by 2 percent. As past coworkers cannot affect the current production of a job mover, the effect is likely coming from the human capital channel by learning from the past coworkers.

Second, I conduct a wage variance decomposition exercise and explore how coworkers can contribute to wage inequality in the market. I find that peer effect alone can account for more than 10 percent of the wage variance, which is on par with the fraction that can be explained by firm heterogeneity. Moreover, sorting between worker and coworker explains around 30 percent of the wage variance, whereas conditional on coworkers, sorting between worker and firm can only explain around 3 percent of the wage variance, indicating that coworkers are the primary driver of the labor market sorting.

To quantify the effect of production complementarity and learning on wage inequality by accounting for the endogenous mobility induced by the two channels, I develop and estimate a search-theoretical model in a frictional labor market with both heterogeneous workers and firms. The model contains key features that are motivated by empirical evidence. It first has both learning-by-doing and learning from coworkers. It also allows production complementarity between worker and firm, as well as production complementarity between worker and coworker. Finally, it provides a flexible wage setting that can generate the model prediction, which is consistent with the empirical findings.

Using the estimated model, I find that half of the variation in income inequality is determined by their pre-market ability. I decompose the remaining unexplained part into firms, coworker production complementarity, and learning from coworkers. I find that coworker production complementarity and learning from coworkers contribute to 15 percent and 30 percent of it, and the firm component explains another 15 percent. The result shows coworkers play a substantially more important role in explaining lifetime income inequality. Moreover, using a policy counterfactual, I show that early exposure to high-quality coworkers has a substantial lifetime impact.

As the knowledge economy is taking off and teamwork has become more and more common in the workplace, coworkers are likely to become more important in the future. My findings also provide some insights for policymakers, especially with the rise of remote work in recent years. High-skilled workers may be less likely to be hurt by, if not benefit from, such a new work mode, as they may still maintain similar production complementarity and learning via effective virtual tools. Low-skilled workers, on the other hand, may experience a hard time getting synergy and learning using such a mode. For example, a blue-collar worker may find it harder to operate machines via a virtual meeting. Such a differential effect might, in the long run, lead to a more polarized labor market. Whether one could use some policy instruments to solve this issue is still understudied and will be interesting to explore in future research.

The study also opens more questions for future research. For one, I have explored a relatively homogeneous effect of production complementarity and learning. How do these effects vary across different observables, such as gender, race, age, location, and occupation? Can differential peer effects between males and females partially explain the gender pay gap? Also, the study sheds little light on the firm perspective. If hiring a superior worker is important to attract other workers, how can a firm optimize its resource between building up better physical capital and paying higher wages for hiring star workers? In terms of learning, can a firm re-optimize its worker training scheme by internalizing learning from coworkers? I leave these questions to be answered in my future research.

References

Abowd, John M., Francis Kramarz, and David N. Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 1999, *67* (2), 251–333.

Altonji, Joseph G and Robert A Shakotko, "Do wages rise with job seniority?," *The Review of Economic Studies*, 1987, *54* (3), 437–459.

Alvarez, Jorge, Felipe Benguria, Niklas Engbom, and Christian Moser, "Firms and the decline in earnings inequality in brazil," *American Economic Journal: Macroeconomics*, 2018, *10* (1), 149–89.

Andrews, Martyn J, Len Gill, Thorsten Schank, and Richard Upward, "High wage workers and low wage firms: negative assortative matching or limited mobility bias?," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2008, *171* (3), 673–697.

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.

Arellano-Bover, Jaime, "Career consequences of firm heterogeneity for young workers: First job and firm size," 2020.

— and Fernando Saltiel, "Differences in on-the-job learning across firms," 2021.

Bagger, Jesper and Rasmus Lentz, "An empirical model of wage dispersion with sorting," *The Review of Economic Studies*, 2019, *86* (1), 153–190.

— , 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.

— , Kenneth L Sørensen, and Rune Vejlin, "Wage sorting trends," *Economics Letters*, 2013, *118* (1), 63–67.

Battisti, Michele, "High wage workers and high wage peers," *Labour Economics*, 2017, *46* (February 2015), 47–63.

Baum-Snow, Nathaniel, Nicolas Gendron-Carrier, and Ronni Pavan, "Local productivity spillovers," 2020.

Becker, Gary S, "A theory of marriage: Part I," *Journal of Political economy*, 1973, *81* (4), 813–846.

Boerma, Job, Aleh Tsyvinski, and Alexander P Zimin, "Sorting with team formation," Technical Report, National Bureau of Economic Research 2021.

Bonhomme, Stéphane, Kerstin Holzheu, Thibaut Lamadon, Elena Manresa, Magne Mogstad, and Bradley Setzler, "How Much Should we Trust Estimates of Firm Effects and Worker Sorting?," 2020.

Bowlus, Audra J and Jean-Marc Robin, "Twenty years of rising inequality in US lifetime labour income values," *The Review of Economic Studies*, 2004, *71* (3), 709–742.

_ and _ , "An international comparison of lifetime inequality: How continental Europe resembles North America," *Journal of the European Economic Association*, 2012, *10* (6), 1236–1262.

Brügemann, Björn, Pieter Gautier, and Guido Menzio, "Intra firm bargaining and Shapley values," *The Review of Economic Studies*, 2019, *86* (2), 564–592.

Brune, Lasse, Eric Chyn, and Jason Kerwin, "Peers and Motivation at Work," *Journal of Human Resources*, 2020, pp. 0919–10416R2.

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin, "Wage Bargaining with On-the-Job Search: Theory and Evidence," *Econometrica*, 2006, *74* (2), 323–364.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline, "Firms and labor market inequality: Evidence and some theory," *Journal of Labor Economics*, 2018, *36* (S1), S13–S70.

_ , Jörg Heining, and Patrick Kline, "Workplace Heterogeneity and the Rise of West German Wage Inequality*," *The Quarterly Journal of Economics*, 05 2013, *128* (3), 967–1015.

Cardoso, Ana Rute, Paulo Guimaraes, Pedro Portugal, and Hugo Reis, "The returns to schooling unveiled," 2018.

Corblet, Pauline, "Education Expansion, Sorting, and the Decreasing Education Wage Premium," *Job Market Paper*, 2022.

Cornelissen, Thomas, Christian Dustmann, and Uta Schönberg, "Peer effects in the workplace," *American Economic Review*, 2017, *107* (2), 425–456.

Cunha, Flavio and James Heckman, "A new framework for the analysis of inequality," *Macroeconomic Dynamics*, 2008, *12* (S2), 315–354.

_ and James J Heckman, "The evolution of inequality, heterogeneity and uncertainty in labor earnings in the US economy," 2007.

Dey, Matthew S and Christopher J Flinn, "An equilibrium model of health insurance provision and wage determination," *Econometrica*, 2005, *73* (2), 571–627.

Dix-Carneiro, Rafael and Brian K Kovak, "Trade liberalization and regional dynamics," *American Economic Review*, 2017, *107* (10), 2908–46.

Eeckhout, Jan and Philipp Kircher, "Assortative matching with large firms," *Econometrica*, 2018, *86* (1), 85–132.

Elsby, Michael WL and Axel Gottfries, "Firm dynamics, on-the-job search, and labor market fluctuations," *The Review of Economic Studies*, 2022, *89* (3), 1370–1419.

Engbom, Niklas, Christian Moser, and Jan Sauermann, "Firm pay dynamics," *Journal of Econometrics*, 2022.

Flinn, Christopher J, "Labour market structure and inequality: A comparison of Italy and the US," *The Review of Economic Studies*, 2002, *69* (3), 611–645.

Freund, Lukas, "Superstar Teams: The Micro Origins and Macro Implications of Coworker Complementarities," *Cambridge Working Paper in Economics CWPE2276*, 2022.

Gourieroux, Christian, Alain Monfort, and Eric Renault, "Indirect inference," *Journal of applied econometrics*, 1993, *8* (S1), S85–S118.

Gregory, Victoria, "Firms as Learning Environments: Implications for Earnings Dynamics and Job Search," *Job Market Paper*, 2021.

Håkanson, Christina, Erik Lindqvist, and Jonas Vlachos, "Firms and Skills The Evolution of Worker Sorting," *Journal of Human Resources*, 2021, *56* (2), 512–538.

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.*

Holden, Richard, Michael Keane, and Matthew Lilley, "Peer effects on the united states supreme court," *Quantitative Economics*, 2021, *12* (3), 981–1019.

Hong, Long and Mikkel Sølvsten, "Estimation and Inference in a Peer Effects Model with Heteroscedasticity," *Manscript*, 2022.

_ and Salvatore Lattanzio, "The Peer Effect on Future Wages in the Workplace," *Available at SSRN*, 2022.

Jäger, Simon, "How substitutable are workers? Evidence from Worker Deaths," *Available at SSRN*, 2016.

Jarosch, Gregor, Ezra Oberfield, and Esteban Rossi-Hansberg, "Learning from coworkers," *Econometrica*, 2021, *89* (2), 647–676.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten, "Leave-out estimation of variance components," *Econometrica*, 2020, *88* (5), 1859–1898.

Kremer, Michael, "The O-ring theory of economic development," *The quarterly journal of economics*, 1993, *108* (3), 551–575.

Lentz, Rasmus and Dale T Mortensen, "Labor market friction, firm heterogeneity, and aggregate employment and productivity," *University of Wisconsin Discussion Paper*, 2012.

_ , Suphanit Piyapromdee, and Jean-Marc Robin, "On worker and firm heterogeneity in wages and employment mobility: Evidence from danish register data," *PIER Discussion Papers*, 2018, *91.*

_ , _ , and _ , "The Anatomy of Sorting–Evidence from Danish Data," *Unpublished manuscript*, 2021.

Lopes de Melo, Rafael, "Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence," *Journal of Political Economy*, 2018, *126* (1), 313 – 346.

Lucifora, Claudio, Simona Comi, and Giorgio Brunello, "The returns to education in Italy: A new look at the evidence," Technical Report, IZA Discussion Papers 2000.

Mas, Alexandre and Enrico Moretti, "Peers at work," *American Economic Review*, 2009, *99* (1), 112–145.

Mueller, Holger M., Paige P. Ouimet, and Elena Simintzi, "Wage Inequality and Firm Growth," *American Economic Review*, May 2017, *107* (5), 379–83.

Nix, Emily, "Learning Spillovers in the Firm," 2020.

Postel–Vinay, Fabien and Jean–Marc Robin, "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 2002, *70* (6), 2295–2350.

Serafinelli, Michel, ""Good" firms, worker flows, and local productivity," *Journal of Labor Economics*, 2019, *37* (3), 747–792.

Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter, "Firming up inequality," *The Quarterly journal of economics*, 2019, *134* (1), 1–50.

Sorkin, Isaac, "Ranking firms using revealed preference," *The quarterly journal of economics*, 2018, *133* (3), 1331–1393.

Taber, Christopher and Rune Vejlin, "Estimation of a Roy/Search/Compensating Differential Model of the Labor Market," *Econometrica*, 2020, *88* (3), 1031–1069.

Tattara, Giuseppe and Marco Valentini, "Turnover and excess worker reallocation. The Veneto labour market between 1982 and 1996," *Labour*, 2010, *24* (4), 474–500.

Thiemann, Petra, "The persistent effects of short-term peer groups on performance: Evidence from a natural experiment in higher education," *Management Science*, 2022, *68* (2), 1131–1148.

Topel, Robert, "Specific capital, mobility, and wages: Wages rise with job seniority," *Journal of political Economy*, 1991, *99* (1), 145–176.

Topel, Robert H and Michael P Ward, "Job mobility and the careers of young men," *The Quarterly Journal of Economics*, 1992, *107* (2), 439–479.

Wallskog, Melanie, "Entrepreneurial Spillovers Across Coworkers," 2022.

White, Halbert and Ian Domowitz, "Nonlinear regression with dependent observations," *Econometrica*, 1984, pp. 143–161.

Appendix A Estimation Strategy: A Brief Summary

The estimation is based on a recent method developed by Hong and Sølvsten (2022). The main idea of the paper is as follows.

A simplified framework To facilitate the discussion, it can be helpful to rewrite Equation 1 in a simplified form:

$$y_\ell = x_\ell'\delta + \beta_0 \cdot a_\ell'\delta + \varepsilon_\ell, \qquad \ell = 1, \ldots, n. \tag{A.1}$$

where x_ℓ and a_ℓ are observed K-dimensional vectors, δ is a vector of nuisance parameters. The vector a_ℓ is a function of the peer group that observation ℓ belongs to and is, intuitively, similar to an averaging vector which constructs $\bar{\alpha}_{-i,t}$ in Equation 1.[36] y_ℓ and ε_ℓ remain the same meaning, but I have suppressed the subscripts for simplification. $\beta_0 \in B$ is the true parameter of β, where B is a compact parameter space. Consistent with the peer effect literature, it is natural to consider $B \subset (-1, 1)$. This parameter space restricts the impact of the average peer quality to be smaller in magnitude than a worker's own effect α_i.

A "cross-fit" objective function The least squares estimator applied to (A.1) yields the following estimator of β_0:

$$\hat{\beta}^{\text{LS}} = \arg\min_{\beta \in B} \min_{\delta \in \mathbb{R}^k} \sum_{\ell=1}^{n} \left(y_\ell - x_\ell'\delta - a_\ell'\delta \cdot \beta \right)^2. \tag{A.2}$$

To give a representation of $\hat{\beta}^{\text{LS}}$ that is more amenable to analysis and intuition, one could eliminate the nuisance vector δ using the Frisch–Waugh–Lovell theorem. To do so, we define the entries of the matrix that residualizes against the regressor $x_\ell + a_\ell\beta$ as $M_{\ell k}(\beta) = \mathbf{1}\{\ell = k\} - (x_\ell + a_\ell\beta)'S(\beta)^{-1}(x_k + a_k\beta)'$. We can then represent $\hat{\beta}^{\text{LS}}$ as the solution to a minimization problem that does not involve δ:

$$\hat{\beta}^{\text{LS}} = \arg\min_{\beta \in B} \hat{Q}_n(\beta) \qquad \text{where } \hat{Q}_n(\beta) = \sum_{\ell=1}^{n} \sum_{k=1}^{n} M_{\ell k}(\beta) y_\ell y_k. \tag{A.3}$$

[36] Additionally, a_ℓ is appended with a vector of zeroes in place of the control variables so that x_ℓ and a_ℓ are both K-dimensional vectors.

The representation of the least squares estimator as a minimizer of the objective function $\hat{Q}_n$ implies that an almost necessary condition for consistency of $\hat{\beta}^{\text{LS}}$ is that the population analog $Q_n(\beta) = \mathbb{E}[\hat{Q}_n(\beta) \mid X, A]$ has a unique minimum at β_0, where $A = (a_1, \ldots, a_n)'$ and $X = (x_1, \ldots, x_n)'$. Assuming no serial correlation in the error term, and let $\sigma_\ell = \mathbb{E}[\varepsilon_\ell^2 \mid X, A]$ be the variance in the ℓ-th error term. Defining $\tilde{a}_\ell(\beta) = \sum_{k=1}^{n} M_{\ell k}(\beta) a_k'$, one could show that

$$Q_n(\beta) = (\beta - \beta_0)^2 \sum_{\ell=1}^{n} (\tilde{a}_\ell(\beta)'\delta)^2 + \sum_{\ell=1}^{n} M_{\ell\ell}(\beta)\sigma_\ell^2. \tag{A.4}$$

Assuming X and A are full rank matrices, $\sum_{\ell=1}^{n} (\tilde{a}_\ell(\beta)'\delta)^2 > 0$ for all $\beta \in B$ so that the first part of Q_n is uniquely minimized at β_0. However, the second term is, in general, *not* minimized at the truth. The presence of the second part will therefore lead to inconsistency of the least-squares estimator except in special cases.[37]

In order to solve this problem, Hong and Sølvsten (2022) proposes a "cross-fit" objective function as follows, which allows β_0 to be the unique minimizer of the objective function (A.5).

$$\hat{Q}_n^{\text{CF}}(\beta) = \hat{Q}_n(\beta) - \textstyle\sum_{\ell=1}^{n} M_{\ell\ell}(\beta)\hat{\sigma}_\ell^2(\beta). \tag{A.5}$$

where $\hat{\sigma}_\ell^2(\beta)$ is the cross-fit variance estimator proposed in Kline et al. (2020) as shown below.

$$\hat{\sigma}_\ell^2(\beta) = \frac{y_\ell \hat{\varepsilon}_\ell(\beta)}{M_{\ell\ell}(\beta)} \tag{A.6}$$

where $\hat{\varepsilon}_\ell(\beta)$ is the regression residual at β. The first order condition of equation (A.5) leads to the proposed estimator in Hong and Sølvsten (2022).

$$\hat{\beta}^{\text{CF}} = \underset{\beta \in B}{\arg\text{zero}}\ \hat{m}_n^{\text{CF}}(\beta) \qquad \text{where } \hat{m}_n^{\text{CF}}(\beta) = \nabla_\beta \hat{Q}_n(\beta) - \sum_{\ell=1}^{n} \nabla_\beta M_{\ell\ell}(\beta)\hat{\sigma}_\ell^2(\beta). \tag{A.7}$$

[37] Under homoscedastic error terms, we have that $\sigma_\ell^2 = \sigma^2$. This property implies that the second part of Q_n simplifies as follows.

$$\sum_{\ell=1}^{n} M_{\ell\ell}(\beta)\sigma_\ell^2 = \sigma^2 \sum_{\ell=1}^{n} M_{\ell\ell}(\beta) = \sigma^2(n-K).$$

Thus, Q_n is uniquely minimized at β_0. In the special case of the peer effects model (A.1) without additional control variables w_{it}, this observation was also made by Arcidiacono et al. (2012).

where $\hat{Q}_n(\beta)$ is the first order condition for the standard least-squared estimation, and the second term is the bias-correlation component due to heteroskedasticity.

Appendix B Additional Tables/Figures

Appendix Table 1: Return to experience in Veneto from 1995 to 2001

	log(wages)
Experience $\times 100$	3.660
Experience^2 $\times 100$	-0.087
$\mathbb{E}[\frac{\partial y_{it}}{\partial e_{it}}] = \hat{\xi}_1 + 2\,\hat{\xi}_2 \mathbb{E}[e_{it}] = 1.38\%$	

Notes. I use workers who are *not* left-censored, which allows me to calculate their actual working experience. I then estimate a simple Mincer regression below,

$$y_{it} = \alpha_i + \xi_1 e_{it} + \xi_2 e_{it}^2 + \varepsilon_{it}$$

where e_{it} is the experience, and the average experience in the sample is around 13.1 years.

Appendix Table 2: Peer effects over time

	1983-1989 (1)	1986-1992 (2)	1989-1995 (3)	1992-1998 (4)	1995-2001 (5)
Peer effect $\widehat{\beta}$	0.247	0.193	0.169	0.278	0.413
	(0.021)	(0.019)	(0.015)	(0.030)	(0.035)
Std of $\bar{\alpha}_{-i,t}$	0.178	0.175	0.186	0.200	0.191
1-SD effect	4.4%	3.4%	3.1%	5.5%	7.8%

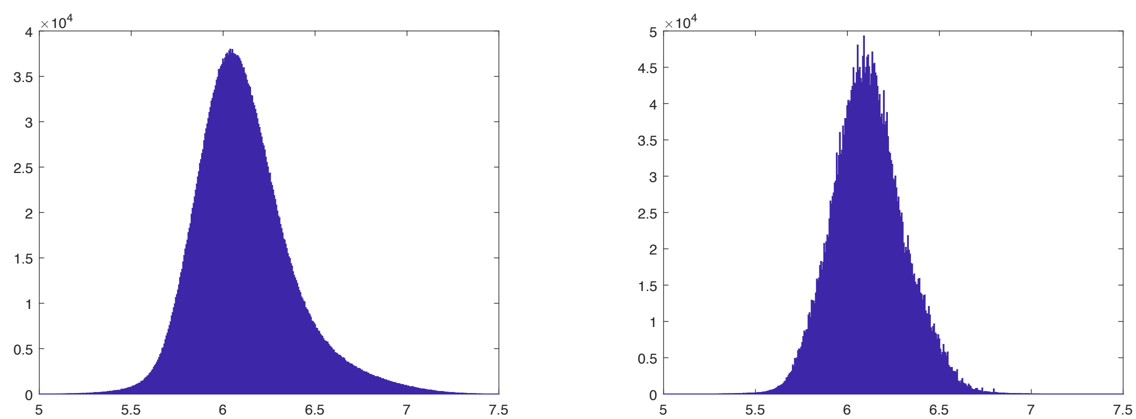
Notes: the table reports the peer effects using a sample with a seven-year interval, and the exercise is repeated every three years.

Appendix Table 3: Wage variance decomposition without bias correction

Variance explained by	AKM Model (1)	Peer Model (2)
Worker effects, α_i	58.1%	50.4%
Firm effects, ψ_j	20.2%	10.0%
Peer effects, $\beta \cdot \bar{\alpha}_{-i,t}$		13.7%
Sorting: 2cov(α_i, ψ_j)	-1.17%	-23.8%
Sorting: 2cov$(\alpha_i, \beta\bar{\alpha}_{-i,t})$		31.7%
Correlation: (worker, firm)	-0.010	-0.186
Correlation: (worker, coworker)		0.570

Notes: The table reports the same statistics in Table 4. However, all the estimates are calculated using the pre-estimated parameters *without* bias-corrected using a similar method as Kline et al (2020).

Appendix Figure 1: Histograms of worker quality and peer quality



Notes. The histograms plot the estimated worker fixed effects (left) and the average coworker's quality (right) from equation (1).

Appendix C Laws of Motion

This section describes, in detail, the law of motion for stationary distribution. Despite the lengthy algebra below, the key idea is that the inflow of a certain type of worker is the same as the outflow of the same type of worker at each stage within a period.

The measures at the search-match stage Let u_m, $e_m(p, ka, 0)$, $e_m(p, ka, \ell t)$ denote the distribution of unemployed worker ka, worker ka without coworkers at firm type p, and worker ka with a coworker ℓt in firm type p, respectively, at the search-match stage.

The measure of unemployed workers, $u_m(ka)$, is expressed below. The first term, being initially unemployed, is the measure of unemployed workers ka who remain unemployed, which can be due to the fact that they either did not contact a firm or contacted a firm but were not hired. Workers who are initially employed without coworkers can also move to unemployment due to exogenous separation from a match. If a worker is initially employed with a coworker ℓt, he can move to unemployment due to either exogenous separation or the team hires a worker $s\tau$ and replaces ka.

$$
\begin{aligned}
u_m(ka) =& u(ka) \left\{ (1-\lambda_u) + \lambda_u \left[\sum_y p(z,q) \left(1 - h_z^q(ka, u)\right) \right] \right\} \\
&+ e(p, ka, 0)\delta + \sum_{\ell t} e(p, ka, \ell t) \left\{ \delta + \left[\sum_{s\tau, x} m(s\tau, x, q) h^p_{ka,\ell t}(s\tau, x) r^p_{ka,\ell t}(s\tau, ka) \right] \right\}
\end{aligned}
$$

The measure of worker ka with no coworker at firm type p, $e_m(p, ka, 0)$, is described below. If the worker is initially unemployed, the worker is hired by an empty team at the firm type p. If the worker is initially employed with no coworkers, then the worker can remain in the same situation if the following three cases happen. First, A poaching firm meets her, but the worker did not move. Second, the firm meets a worker $(\ell t, x, q)$, but the firm does not hire her. Finally, nothing happened: the worker does not encounter exogenous job destruction, meet a poaching firm, or meet a worker. If the worker is initially employed with a coworker ℓt at a firm p, then the worker becomes a solo worker when the worker is hired by an empty team at another firm type p, or the coworker is hired by a poaching firm, or the coworker is exogenously separated. If the worker is employed at a firm type $q \neq p$ regardless of the coworker types, then the worker can be poached by an empty team in firm type p.

$$
\begin{aligned}
e_m(p,ka,0) =& u(ka)\left[\lambda_u p((0,0),p) h^p_{0,0}(ka,u,0)\right] + e(p,ka,0)\left\{\left[\sum_{z,q}\lambda_e p(z,q)\left(1-h^q_z(ka,0,p)\right)\right]\right.\\
&+\left[\sum_{\ell t,x,q} m(\ell t,x,q)\left(1-h^p_{ka,0}(\ell t,x,q)\right)\right] + \left.\left[1-\delta-\lambda_e-\sum_{\ell t,x,q} m(\ell t,x,q)\right]\right\}\\
&+\sum_{\ell t} e(p,ka,\ell t)\left\{\lambda_e p(0,0) h_{0,0}(k,a,\ell t) + \left[\sum_{z,q}\lambda_e p(z,q) h^q_{z,q}(\ell t,ka)\right] + \delta\right\}\\
&+\sum_{q\neq p} e(q,ka)\ \lambda_e p((0,0),p) h^p_{0,0}(ka,0,q)\\
&+\sum_{q\neq p,\ell t} e(q,ka,\ell t)\ \lambda_e p((0,0),p) h^p_{0,0}(ka,\ell t,q)
\end{aligned}
$$

The measure of worker ka employed with a coworker of type ℓt at firm type p, $e_m(p,ka,\ell t)$, is described below. If the worker is initially unemployed, she can be hired by a team $(\ell t, 0)$ or replace the worker i at a team $(\ell t, i)$ in a firm type p. If the worker is initially employed without a coworker at firm type p, then the firm can hire a worker ℓt. And the worker is poached and hired by a team $(\ell t, 0)$ or replaces the worker i at a team $(\ell t, i)$ in a firm type p. If the worker is initially employed with a coworker ℓt, then the worker can keep in the same situation if the following cases happen. First, either ka or ℓt is poached, but neither is hired. Second, the team meets a worker but does not hire her. Finally, nothing happens: neither of the workers is exogenously separated; they are not poached by a firm; the firm does not meet any worker. If the worker is initially employed with a coworker with type s, then the worker is poached and hired by a team $(\ell t, 0)$ or replaces the worker i at a team $(\ell t, i)$ in a firm type p. The team can also meet a worker type ℓt and decides to hire and replace worker s. Finally, if the worker is employed at a firm type $q \neq p$ regardless of the coworker types, then the worker can be poached by an empty team in firm type p.

$$
\begin{aligned}
e_m(p, ka, \ell t) = {} & u(ka)\lambda_u \Bigg\{ p((\ell t, 0), p) h^p_{\ell t,0}(ka, u, 0) \\
& + \sum_i \left[p((\ell t, i, p)) h^p_{\ell t,i}(ka, u, 0) r^p_{\ell t,i}(ka, i) + p(i, \ell t, p) h^p_{i,\ell t}(ka, u, 0) r^p_{i,\ell t}(ka, i) \right] \Bigg\} \\
& + e(p, ka, 0) \Bigg\{ \lambda_e \left[\sum_i p(\ell t, i, p) h^p_{\ell t,i}(ka, 0) r^p_{\ell t,i}(ka, i) + p(i, \ell t) h_{i,\ell t}(ka, 0) r_{i,\ell t}(ka, i) \right] \\
& + \lambda_e p(\ell t, 0, p) h^p_{\ell t,0}(ka, 0, p) + \left[\sum_{x,q} m(\ell t, x, q) h^p_{ka,0}(\ell t, x, q) \right] \Bigg\} \\
& + e(p, ka, \ell t) \Bigg\{ \lambda_e \left[\sum_{z,q} p(z, q) \left(1 - h^q_z(ka, \ell t, p)\right) + \sum_{z,q} p(z, q) \left(1 - h^q_z(\ell t, ka, p)\right) \right] \\
& + \sum_{(i,x,q)} q(i, x, q) \left(1 - h^p_{ka,\ell t}(i, x, q)\right) + \left[1 - 2\delta - 2\lambda_e - \sum_{(i,x,q)} m(i, x, q) \right] \Bigg\} \\
& + \sum_s e(p, ka, s) \Bigg\{ \lambda_e p((\ell t, 0), p) h^p_{\ell t,0}(ka, s, p) + \lambda_e \sum_i \Big[p(\ell t, i, p) h^p_{\ell t,i}(ka, s, p) r^p_{\ell t,i}(ka, i) \\
& + p(i, \ell t, p) h^p_{i,\ell t}(ka, s, p) r^p_{i,\ell t}(ka, i) \Big] + \sum_{x,q} m(\ell t, x, q) h^p_{ka,s}(\ell t, x, q) r^p_{ka,s}(\ell t, s) \Bigg\} \\
& + \sum_{q \neq p} e(q, ka)\ \lambda_e \Bigg\{ p((\ell t, 0), p) h^p_{\ell t,0}(ka, 0, q) \\
& + \sum_i \left[p((\ell t, i), p) h^p_{\ell t,i}(ka, 0, q) r^p_{\ell t,i}(ka, i) + p((i, \ell t), p) h^p_{i,\ell t}(ka, 0, q) r^p_{i,\ell t}(ka, i) \right] \Bigg\} \\
& + \sum_{q \neq p, s} e(q, ka, s)\ \lambda_e \Bigg\{ p((\ell t, 0), p) h^p_{\ell t,0}(ka, s, q) \\
& + \sum_i \left[p((\ell t, i), p) h^p_{\ell t,i}(ka, s, q) r^p_{\ell t,i}(ka, i) + p((i, \ell t), p) h^p_{i,\ell t}(ka, s, q) r^p_{i,\ell t}(ka, i) \right] \Bigg\}
\end{aligned}
$$

The measures immediately before the production stage Let u_p and e_p denote the distribution of workers across employment states immediately before the production stage. Before production, each team will have to check whether the participation constraint is violated. If so, the match can be dissolved. Denote $d^p_z(ka)$ is the probability that the firm p and worker ka mutually agree to dissolve the match.

The measure of unemployed workers, $u_p(ka)$, is shown below. If a worker is employed, the worker can move to unemployment if the match is dissolved.

$$u_p(ka) = u_p(ka) + \sum_p e_p(p, ka, 0)\ d^p_{ka,0}(ka) + \sum_{\ell t, p} e_p(p, ka, \ell t) d^p_{ka,\ell t}(ka),$$

The measure of workers with no coworker, $e_p(ka, 0)$, is described below. If the worker is initially employed with no coworker, then the match is not dissolved. If the worker is employed with a coworker, then the coworker is dissolved.

$$e_p(p, ka) = e_p(p, ka) \left(1 - d^p_{ka,0}(ka)\right) + \sum_{\ell t} e_p(ka, \ell t) d_{ka,\ell t}(\ell t)$$

The measure of workers employed with a coworker of type ℓt, $e_p(p, ka, \ell t)$ is expressed below, and neither is dissolved from the match.

$$e_p(p, ka, \ell t) = e_p(p, ka, \ell t) \left(1 - d_{ka,\ell t}(ka)\right) \left(1 - d_{ka,\ell t}(\ell t)\right)$$

The measures at the learning stage of next period Let u_g and e_g denote the distribution of workers at different employment states at the learning stage. In this stage, only the human capital is updated.

The measure of unemployed workers remains unchanged as the model does not assume human capital depreciation, $u_g(ka) = u(ka)$. The measure of workers with no coworker, $e_g(p, ka, 0)$, is expressed blow. While working alone, human capital increases from s to k during learning by doing.

$$e_g(p, ka) = \sum_s g(k \mid s, 0) e(p, sa)$$

The measure of workers employed with a coworker of type ℓt, $e_g(p, ka, \ell t)$, is shown below. While working with a coworker, the worker with human capital s increases to k, and the worker with human capital i increases to ℓ.

$$e_g(p, ka, \ell t) = \sum_{s,i} g(k \mid s, it) g(\ell \mid i, sa) e(p, sa, it)$$

The measures at the entry and exit stage of the next period Finally, let u_+ and e_+ denote the distribution of workers across employment states at the entry and exit stage. The measures $u_+(ka), e_+(p, ka, 0)$ and $e_+(p, ka, \ell t)$ are given by

$$\begin{aligned}
u_+(p, ka) &= (1-\sigma)u_g(ka) + \sigma\, \Phi(a) \\
e_+(p, ka, 0) &= (1-\sigma)e_g(ka, 0) + \textstyle\sum_{\ell t} \sigma(t) e_g(ka, \ell t) \\
e_+(p, ka, \ell t) &= (1-2\sigma)e_m(ka, \ell t)
\end{aligned}$$

The equilibrium inflow-outflow condition The distribution u_+, e_+ is also the distribution of workers across employment states at the beginning of the search stage of the next period. The distribution is stationary *if and only if* the following inflow-outflow conditions hold:

$$\begin{aligned}
&u_+(ka) - u(ka) = 0, \\
&e_+(p, ka) - e(p, ka) = 0 \\
&e_+(p, ka, \ell t) - e(p, ka, \ell t) = 0
\end{aligned}$$