

Learning from Coworkers*

Gregor Jarosch

Ezra Oberfield

Princeton University

Princeton University

gregorjarosch@gmail.com

edo@princeton.edu

Esteban Rossi-Hansberg

Princeton University

erossi@princeton.edu

July 11, 2018

Preliminary and Incomplete

Abstract

We investigate learning at the workplace. To do so, we use German administrative data that contain information on the entire workforce of the sample establishments. We document that having more highly paid coworkers is strongly associated with future wage growth. We then show how wage growth is affected by other characteristics of the distribution of coworkers. We then develop a novel way to structurally estimate a coworker learning function within a benchmark model of idea flows in a competitive labor market. Our quantitative approach imposes minimal restrictions on firms' production functions, can be implemented on a very short panel, and allows for potentially rich and flexible coworker learning functions. We demonstrate the approach using several different specifications and find that, in line with our reduced form work, learning from more knowledgeable coworkers is significant and increases with knowledge gaps. In contrast, learning is less sensitive to less knowledgeable coworkers. Our estimated model can be used to examine the relationship between changes in the organization of work brought about by technological change and both earnings inequality and aggregate growth.

*We thank the IR Section at Princeton University for providing a secured facility to access the German data and the IES at Princeton University for financial support. We thank participants at seminars for useful comments.

1 Introduction

Social interactions are an essential part of an individual's life. These interactions are potentially an important source of learning. Furthermore, since working adults spend most of their time awake working, or interacting with coworkers, it is natural that most of this learning is the result of interactions with coworkers. It is plausible that this form of learning constitutes the largest and most important knowledge acquisition mechanism in society. One that transmits and diffuses the practical productive knowledge that individuals use every day in their productive endeavors. Little is known about this type of knowledge transfer in the workplace. Who learns from whom? How much? What is the labor market value of this learning? How does this learning change as we change the organization of production in the economy? We aim to provide answers to some of these questions.

In this paper we investigate learning at the workplace. We are interested in understanding how individuals learn from coworkers with different levels of knowledge and the implications of this form of learning for individual and aggregate outcomes. To do so, we first develop a benchmark model of idea flows in a competitive labor market. Workers produce in teams and, while doing so, learn from each other. The model has the key feature that a worker's pay reflects both her knowledge and a compensating differential for the opportunity to learn from her coworkers. In contrast, the labor market compensates those who provide their coworkers with learning opportunities.

Before structurally connecting the model with the data we start by analyzing the empirical relationship between the wage growth of an individual and the wages of her coworkers in a reduced form. To discipline the key features of this relationship we use German administrative data that contain the employment biographies of the entire workforce of the establishments in the sample. We use a variety of empirical specifications that allow us to understand which features of the distribution of wages are related to an individual's wage growth.

Our findings indicate that more highly paid coworkers substantially increase future wage growth. Furthermore, the transmission depends on particular features of the wage distribution. The data suggest little congestion from less well paid workers and a roughly symmetric

positive learning from those higher up in the wage distribution. We also show that the effects we find are present across the wage distribution.

Although suggestive of significant learning from coworkers, these findings could in principle also be consistent with other features of wage setting mechanisms in the labor market. To address these possibilities we offer a battery of checks which suggest that these findings do not purely reflect mean reversion, back-loading, or other firm-specific factors by separately analyzing switchers and stayers, plant closings, using information about worker tenure, and studying the nonlinearities in the empirical relationship between wage growth and the distribution of coworkers.

We then revisit our model which yields a mapping between the reduced-form evidence on wages and the underlying knowledge and learning of individuals. Hence, we structurally estimate a parametric version of our model using a simple learning function that can account for the most important reduced form patterns we document. We develop a novel way to estimate the parameters of this function using the matched employer-employee information for the German labor market. The result is a structurally estimated ‘learning function’ that maps an agent’s learning to the knowledge distribution of her coworkers.

Our model and estimation strategy rely on some standard, but relatively strong, assumptions like complete markets, perfect competition, and no aggregate shocks. However, it is quite general on the set of firm technologies, complementarities across workers, and the specification of the learning function. Furthermore, it can be implemented on very short panels requiring only two observations per worker. As such, we view the structure we impose, and the empirical results we obtain with it, as a natural benchmark.

Using this benchmark model we find that agents learn little from less knowledgeable workers and they learn significantly from those with more knowledge, particularly from the most knowledgeable members of their teams. On average, about 4.3% of the total compensation of workers comes in the form of learning from coworkers in the same establishment and occupation. We further document an apparent tension between firms’ production requirements—which are reflected in the equilibrium composition of teams—and coworker learning: Coworker knowledge flows would more than double if workers were to be grouped in teams randomly.

Our estimated model can be used to study a variety of phenomena that might affect team composition and therefore individual learning. For example, one can study how changes in the organization of work brought about by technological change affect earnings inequality, life cycle wage profiles, and the aggregate rate of growth.

There is a large literature in macroeconomics that has used learning from others as the key mechanism to generate aggregate growth. [Lucas \(2009\)](#) proposes a theory of growth based on random meetings between agents in the entire population. [Lucas and Moll \(2014\)](#) and [Perla and Tonetti \(2014\)](#) extend these models to add a time allocation choice, while [Jovanovic and Rob \(1989\)](#), [Jovanovic and MacDonald \(1994\)](#), and [König et al. \(2016\)](#) focus on the innovation/imitation margin. Other models like [Sampson \(2015\)](#), [Perla et al. \(2015\)](#), and [Luttmer et al. \(2014\)](#) also generate growth through adoption of ideas from others. As [Alvarez et al. \(2013\)](#) and [Buera and Oberfield \(2016\)](#) show, the selection of what particular ideas an individual or firm confronts, as determined for example by trade flows, is essential to shape the growth properties of these models. This literature considers random learning from the population, or a selected group of the population, but it has not incorporated learning from coworkers. The importance of studying this form of selection in learning is evident, but challenging. For starters, it requires modeling explicitly teams of coworkers that are heterogeneous across firms. [Caicedo et al. \(2016\)](#) introduce learning in an economy where production is organized in heterogeneous production hierarchies as in [Garicano and Rossi-Hansberg \(2006\)](#), but learning is not limited to agents within the organization. [Jovanovic \(2014\)](#) studies learning in teams of two, while [Burstein and Monge-Naranjo \(2009\)](#) study an environment in which a manager hires identical workers and imparts knowledge to those workers.¹ We go further than these papers in that we model learning within teams and provide direct evidence of its importance, its characteristics, as well as providing a structural estimation of the key parameters of the model.

While much of the empirical literature has focused on contemporaneous peer effects ([Mas and Moretti \(2009\)](#) and [Cornelissen et al. \(2017\)](#)), empirical studies of learning within teams is much more limited. [Nix \(2015\)](#) argues that increasing the average education of ones peers

¹[Anderson and Smith \(2010\)](#) study matching with dynamic types which can also be interpreted as a model of learning in teams of two.

raises one's earnings in subsequent years. [Akcigit et al. \(2018\)](#) argue that increasing one's exposure to star patenters raises the likelihood of patenting and the quality of one's patents.

In related and complementary work, [Herkenhoff et al. \(2018\)](#) build on a frictional sorting setup to investigate learning with production complementarities. Like us, they detect strong coworker spillovers. The main difference between the papers is that our empirical strategy builds on a competitive labor market and our analysis focuses in particular on the role of the within-team distribution of knowledge for the transmission process.

The remainder of this paper is organized as follows. Section 2 presents a general but simple model of an economy in which agents learn from their coworkers. The theory is useful in specifying exactly the concept of learning we have in mind and its implications. In Section 3 we introduce our German matched employer-employee data and present a number of reduced-form findings about the relationship between the wage of an individual, the wages of her co-workers and an individual's wage growth. Section 4 goes deeper and parameterizes the learning function in our theory from Section 2 in order to structurally estimate the model. Section 5 concludes.

2 Theory

Consider an economy populated by a unit mass of heterogenous individuals with knowledge $z \in \mathcal{Z} = [0, \bar{z}]$. Individuals have a probability δ of dying each period. Each period a mass δ of new individuals is born. Newborns start with a level of knowledge z drawn from a distribution $B_0(\cdot)$. Agents supply labor inelastically, consume, and discount the future according to a discount factor β . Agents are employed in firms where they obtain a wage and where they can learn from other coworkers. An agent z , working in a firm that employs the agent as well as a vector of coworkers $\tilde{\mathbf{z}}$ will draw her next period's knowledge from a distribution $G(z'|z, \tilde{\mathbf{z}})$. Markets are complete, so agents maximize the expected present value of income.

Since individuals learn from coworkers, the wage she is willing to accept depends on how much she might learn from coworkers. There will therefore be a wage schedule $w(z, \tilde{\mathbf{z}})$ that is paid to a worker with knowledge z that works with a vector of coworkers $\tilde{\mathbf{z}}$.

All firms produce the same consumption goods. Potential firms pay a fixed cost c in goods, after which they draw technology $a \in \mathcal{A}$ from a distribution $A(\cdot)$. A firm with technology a produces according to the production function $F(\mathbf{z}; a)$. Firms take the wage schedule as given and hire a vector of workers \mathbf{z} . Differences across technologies need not be Hicks-neutral or even factor augmenting; production technologies may also vary in their complementarities across workers with different levels of knowledge. Hence different firms will, in general, make different choices of \mathbf{z} .

2.1 Firms

Let $W(\mathbf{z})$ be the total wage bill of a firm that hires the vector of workers \mathbf{z} . If $\mathbf{z} = \{z_i\}_{i=1}^n$ for some n , then $W(\mathbf{z}) = \sum_{i=1}^n w(z_i, \tilde{\mathbf{z}}_{-i})$, where $\tilde{\mathbf{z}}_{-i}$ is the set of i 's coworkers. A firm chooses the set of workers to maximize profit

$$\pi(a) = \max_{\mathbf{z}} F(\mathbf{z}; a) - W(\mathbf{z}). \quad (1)$$

Let $\mathbf{z}(a) = \arg \max_{\mathbf{z}} F(\mathbf{z}; a) - W(\mathbf{z})$ denote a 's optimal choice.²

2.2 Individuals

Agents decide where to work each period given wages and the learning opportunities across firms. Let $\tilde{\mathbf{Z}}$ be the set of all possible vectors of coworkers. The expected present value of earnings for an agent with knowledge z is given simply by

$$V(z) = \max_{\tilde{\mathbf{z}} \in \tilde{\mathbf{Z}}} w(z; \tilde{\mathbf{z}}) + \beta \int_0^\infty V(z') dG(z'|z, \tilde{\mathbf{z}}). \quad (2)$$

Namely, individuals choose where to work to maximize their wage, plus the future stream of wages given their learning opportunities in the firm. In general it will be the case that equilibrium wages will adjust so that workers are indifferent about working in a set of firms. Furthermore, since firms take the wage schedule as given, it must be the case that if a firm

²Note that the firm is choosing both the type of workers, z_i , and the number of workers n . Together these choices determine the vector \mathbf{z} .

wants to hire a vector of workers $(z, \tilde{\mathbf{z}})$ then the wage schedule must capture what it would cost to hire those workers. The wage schedule must therefore satisfy

$$w(z; \tilde{\mathbf{z}}) = V(z) - \beta \int_0^\infty V(z') dG(z'|z, \tilde{\mathbf{z}}) \quad (3)$$

for any $z, \tilde{\mathbf{z}}$. A simple implication is that for any $\tilde{\mathbf{z}}, \tilde{\mathbf{z}}'$

$$w(z; \tilde{\mathbf{z}}) - w(z; \tilde{\mathbf{z}}') = -\beta \left[\int_0^\infty V(z') dG(z'|z, \tilde{\mathbf{z}}) - \int_0^\infty V(z') dG(z'|z, \tilde{\mathbf{z}}') \right]. \quad (4)$$

Namely, firms with distinct sets of employees pay different wages to identical individuals to compensate for differences in their learning. If an individual would learn a lot at a firm, the firm can pay a low wage and still attract her. In this sense wages incorporate compensating differentials.

2.3 Labor Market Clearing and Free Entry

Let $B(z)$ be the fraction of workers with knowledge no greater than z . For any vector \mathbf{z} , let $N(\mathbf{z}, z)$ denote the number of elements of \mathbf{z} that are weakly less than z . Labor market clearing requires that for each z ,

$$B(z) = m \int_a N(\mathbf{z}(a), z) dA(a), \quad (5)$$

where m denotes the mass of firms in the economy.

Free entry requires that

$$\int_a [\pi(a) - c] dA(a) = 0. \quad (6)$$

2.4 The Distribution of Knowledge

Given the choices of firms, we can define $O(\tilde{\mathbf{z}}|z) : \tilde{\mathbf{Z}} \times \mathcal{Z} \rightarrow [0, 1]$ to be the fraction of workers with knowledge z that, in equilibrium, have a vector of coworker knowledge that is strictly dominated by the vector $\tilde{\mathbf{z}}$. Then the fraction of workers with knowledge no greater than z next period are those who are born with knowledge weakly less than z , and those who did

not learn enough from coworkers to push them to knowledge above z . Namely,³

$$B(z) = \delta B_0(z) + (1 - \delta) \int_x \int_{\tilde{\mathbf{z}} \in \tilde{\mathbf{Z}}} G(z|x, \tilde{\mathbf{z}}) dO(\tilde{\mathbf{z}}|x) dB(x). \quad (7)$$

2.5 Equilibrium

A stationary competitive equilibrium consists of a wage schedule w , a value function V , a mass of firms m , firm choices $\mathbf{z}(a)$, a coworker vector $\tilde{\mathbf{Z}}$ set, and a distribution of worker knowledge B , such that

1. V and w satisfy (2) and (3);
2. $\mathbf{z}(a)$ solves (1), namely, maximizes the profit for a firm with technology a taking the wage schedule as given;
3. The labor market clears for each z , so (5) is satisfied;
4. The free entry condition (6) holds;
5. The law of motion for B in (7) is satisfied.

2.6 Characterizing Equilibrium

We now turn to characterize some basic properties of an equilibrium. To do so we need to impose some structure on the functions F and G . We state these properties in three assumptions. Throughout, when we compare two ordered vectors of the same length, $\mathbf{z}_1 < \mathbf{z}_2$ means that each element of \mathbf{z}_2 is weakly greater than the corresponding element in \mathbf{z}_1 , and at least one element is strictly greater.

Assumption 1 $F(\mathbf{z}, a)$ is strictly increasing in each element of \mathbf{z} : $\mathbf{z}_1 < \mathbf{z}_2$ implies $F(\mathbf{z}_1, a) < F(\mathbf{z}_2, a)$

³Note that if the solution to the maximization in (2) is unique, then $O(\tilde{\mathbf{z}}|x)$ would be degenerate with a mass point at $\tilde{\mathbf{z}}$ chosen by individual x , $\tilde{\mathbf{z}}(x)$, and so the integral in (7) would be $\int_x G(z|x, \tilde{\mathbf{z}}(x)) dB(x)$. Uniqueness of the solution of the maximization in (2) is neither guaranteed nor necessarily a desired property in our model.

Assumption 2 G is strictly decreasing in z and $\tilde{\mathbf{z}}$: $\tilde{\mathbf{z}}_1 < \tilde{\mathbf{z}}_2$ implies that $G(z'|z, \tilde{\mathbf{z}}_1) > G(z'|z, \tilde{\mathbf{z}}_2)$.

Assumption 3 There is free disposal of knowledge

The first assumption implies more knowledgeable individuals always have an absolute advantage in production. The second assumption is that if two individuals have the same coworkers, the one with more knowledge this period will have stochastically more knowledge next period. It also says that if two individuals have the same knowledge, the one with more knowledgeable coworkers will have stochastically more knowledge next period.

These assumptions are sufficient to deliver the following results:

Lemma 1 Suppose there is a firm with productivity a such that $(z, \tilde{\mathbf{z}}) = \mathbf{z}^*(a)$. Then for each $z_1 > z_2$ it must be that $w(z_1, \tilde{\mathbf{z}}) > w(z_2, \tilde{\mathbf{z}})$.

Proof. First, free disposal of knowledge ensures that V is weakly increasing. Second, the fact that G is decreasing in z implies that $w(z, \tilde{\mathbf{z}})$ is weakly decreasing in $\tilde{\mathbf{z}}$. Finally, toward a contradiction, suppose there was a $z_1 > z_2$ such that $w(z_1, \tilde{\mathbf{z}}) \leq w(z_2, \tilde{\mathbf{z}})$. Then the firm should hire z_1 instead of z_2 . It would strictly increase output, it could pay that worker a weakly lower wage, and it could weakly lower the wage of all other workers. ■

Proposition 1 $V(z)$ strictly increasing in z .

Proof. For any wage schedule, the operator $\mathcal{TV}(z) = \max_{\tilde{\mathbf{z}} \in \tilde{\mathbf{Z}}} w(z, \tilde{\mathbf{z}}) + \beta \int_0^\infty V(z') dG(z'|z, \tilde{\mathbf{z}})$ is a contraction because it satisfies Blackwell's sufficient conditions. To show that the V is strictly increasing, it is sufficient to show that if V is weakly increasing, \mathcal{TV} is strictly increasing. To see this, consider $z_1 < z_2$. Market clearing ensures that there is a firm that hires z_1 , and let $\tilde{\mathbf{z}}_1$ be the the coworkers of z_1 in at least one such firm. Then this along with

Lemma 1 imply

$$\begin{aligned}
\mathcal{TV}(z_1) &= w(z_1, \tilde{\mathbf{z}}_1) + \beta \int_0^\infty V(z') dG(z'|z_1, \tilde{\mathbf{z}}_1) \\
&< w(z_2, \tilde{\mathbf{z}}_1) + \beta \int_0^\infty V(z') dG(z'|z_1, \tilde{\mathbf{z}}_1) \\
&\leq w(z_2, \tilde{\mathbf{z}}_1) + \beta \int_0^\infty V(z') dG(z'|z_2, \tilde{\mathbf{z}}_1) \\
&\leq \max_{\tilde{\mathbf{z}} \in \tilde{\mathbf{Z}}} w(z_2, \tilde{\mathbf{z}}) + \beta \int_0^\infty V(z') dG(z'|z_2, \tilde{\mathbf{z}}) \\
&= \mathcal{TV}(z_2)
\end{aligned}$$

where the first inequality follows from Lemma 1 and the second inequality from the assumption that $G(\cdot|z, \tilde{\mathbf{z}})$ is decreasing in z and the presumption that V is weakly increasing

■

Proposition 2 $\tilde{\mathbf{z}}_1 < \tilde{\mathbf{z}}_2$ implies that $w(z, \tilde{\mathbf{z}}_1) > w(z, \tilde{\mathbf{z}}_2)$.

Proof. This follows directly from the assumption that G is decreasing in z , Proposition 1, and (4). ■

Proposition 3 *Within a team, a worker that earns a higher wage has more knowledge.*

Proof. Consider two workers in the same team, with respective knowledge $z_1 > z_2$. Let $\tilde{\mathbf{z}}$ denote the vector of the rest of their coworkers. Then we have that

$$w(z_2, (z_1, \tilde{\mathbf{z}})) < w(z_1, (z_1, \tilde{\mathbf{z}})) < w(z_1, (z_2, \tilde{\mathbf{z}}))$$

where the first inequality follows from Lemma 1 and the second inequality follows from Proposition 2. ■

Finally, we show how a worker's wage is related to her marginal product. Firms choose a vector of workers \mathbf{z} to maximize profits. Hence, they solve

$$\pi(a) = \max_{n, \{z_i\}_{i=1}^n} F(\mathbf{z}; a) - \sum_{j=1}^n w(z_j, \tilde{\mathbf{z}}_{-j})$$

If a firm wanted to choose a slightly more knowledgeable worker for the j^{th} position. Optimality implies

$$\frac{\partial}{\partial z_i} F(\mathbf{z}; a) - \sum_{j \neq i} \frac{w(z_j, \tilde{\mathbf{z}}_{-j})}{\partial z_i} = \frac{\partial w(z_i, \tilde{\mathbf{z}}_{-i})}{\partial z_i}$$

The marginal cost to a firm of having its i^{th} worker have a bit more knowledge is $\frac{\partial w(z_i, \tilde{\mathbf{z}}_{-i})}{\partial z_i}$. The marginal benefit equals the sum of its marginal product and the change in wages the firm must pay its other workers.

Since (3) must hold for any $\tilde{\mathbf{z}}$, we can differentiate with respect to coworker i 's knowledge to get

$$\frac{\partial w(z_j, \tilde{\mathbf{z}}_{-j})}{\partial z_i} = -\beta \frac{dE[V(z') | z_j, \tilde{\mathbf{z}}_{-j}]}{dz_i}$$

We can thus write the optimal condition for the firm to be

$$\frac{\partial w(z_i, \tilde{\mathbf{z}}_{-j})}{\partial z_i} = \frac{\partial}{\partial z_i} F(\mathbf{z}; a) + \beta \sum_{j \neq i} \frac{d}{dz_i} \mathbb{E}[V(z') | z_j, \tilde{\mathbf{z}}_{-j}].$$

Thus the marginal value of a worker's knowledge to the firm reflects both the marginal product of the knowledge and the marginal increase in coworkers' learning.

3 Reduced-Form Evidence

We next use German social security data to investigate the relationship between coworker (relative) wages and individual wage growth. Our goal is to provide empirical discipline on the learning function $G(z'|z, \tilde{\mathbf{z}})$.

To do so, we begin in a reduced-form fashion relating individual wage growth to the within-team wage distribution using various flexible reduced-form specifications. We are particularly interested in two questions: First, do future wages rise more steeply if one's coworkers are more highly paid. Second, if so, how does the response of wage growth co-move with key features of the wage distribution around a worker. We further offer various robustness checks with a particular focus on ruling out two alternatives to learning which could be driving our results, namely a back-loaded wage structure and mean reversion.

While none of the reduced-form specifications is tightly grounded in our theory, we nonetheless argue that the resulting picture is useful in guiding the more structural approach that follows. Of course, the lack of an empirical identification strategy implies that these results cannot be understood as causal and so remain simply a suggestive statement about equilibrium relationships. The next step is to use the structure of our theory and the suggested specification of the learning function to estimate the model.

Our structural approach builds on the baseline framework developed in Section 2. In particular, we specify a flexible but parsimonious parametric form for the learning function $G(z'|z, \tilde{\mathbf{z}})$ in light of our reduced form evidence. We then implement a simple routine to discipline the parameters of $G(\cdot)$ using only panel information on wages and peer groups.

3.1 Summary Statistics

We begin by briefly describing the dataset along key dimensions. The longitudinal version of the *Linked-Employer-Employee-Data of the IAB (LIAB)* contains information on the complete workforce of a subset of German establishments. The sample establishments are the ones selected—at least once—in an annually conducted survey between 2000 and 2008. The employee part of the dataset then contains the employment biographies from 1993 to 2010 of all individuals which were, for at least one day, employed at one of the sample establishments between 1999 and 2009. The employment biographies come in spell format and contain information, among other things, on a worker’s establishment, occupation, and average daily earnings along with a rich set of observables (age, gender, job and employment tenure, education, location,..). We organize the remaining dataset as an annual panel. Specifically, the annual observation recorded (employer, average daily wage,..) for each individual pertains to the spell which overlaps a particular *reference date* (January 31st).⁴ We will refer to the spell overlapping this reference date as the *reference spell*.

To construct a baseline sample, we then proceed as follows. We select only those establishments which were surveyed in each year between 2000 and 2008. We then include those individuals who were employed at one of those establishments at the reference date

⁴As an alternative for wages, we can also compute the average daily wage over the full year, potentially spanning multiple employers.

during at least one year between 2000 and 2008. This leaves us with the employment biographies (between 1993 and 2010) of the full workforce (at a reference date) of a large number of establishments. We further drop workers in apprenticeship, internships, and part-time employment.

Throughout, we work with the following two different ways of defining a peer group.

Team Definition 1: All workers in the same establishment.

Team Definition 2: All workers in the same establishment and occupation.

Size Distribution Figure 1 plots the unweighted size distribution for both team definitions for the year 2000. We restrict attention to teams that have size ≥ 2 . The team size distribution is naturally more compressed under the second, narrower, team definition, but for both definitions a sizable fraction of teams are fairly large. The remaining sample contains 4119 establishments with average size 143. When working with the second team definition we have a total of 21383 teams with an average size of 23.

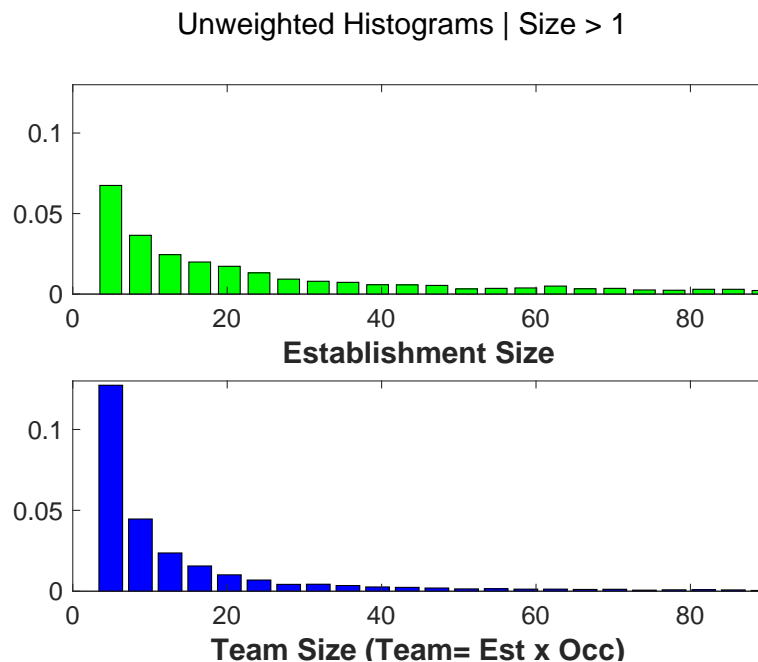


Figure 1 Size Distribution of Teams, two different team definitions.

Wage Distribution Figure 2 plots the histogram of the average daily earnings during the year-2000 reference spell in Euros.⁵ The two “mass-points” reflect that the earnings data are censored at the social security contribution ceiling (which is lower in Eastern Germany). As a consequence, roughly 5% of our wage observations are censored.⁶ A simple variance decomposition implies that the within team component accounts for 47.2% of the overall variance in wages under team definition 1 and 22.8% under team definition 2.

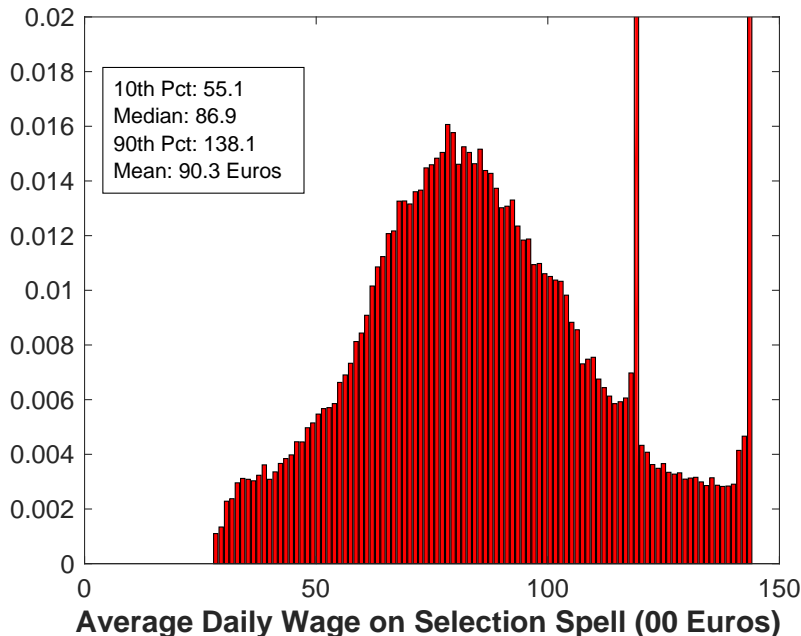


Figure 2 Wages during reference spell

Wage Gap to Coworkers We are interested in gauging how a worker’s future wage growth comoves with her coworker’s (relative) wages. Let the wage of individual i in year t be denoted by $w_{i,t}$ and the (leave-out) average wage of individual i ’s peers/team-members be denoted by $\bar{w}_{i,t}$. We then define $\hat{w}_{i,t} \equiv \log\left(\frac{\bar{w}_{i,t}}{w_{i,t}}\right)$. Figure 3 plots the histogram of wage gaps for each team definition. Under the first team definition, \hat{w} has mean .026 and -.26, .04, and .29 as the 10th, 50th, and 90th percentile when pooled across all years. Under the second team definition, \hat{w} has mean .011 and -.17, .00, and .20 as the 10th, 50th, and 90th percentile. Naturally, within-team wage dispersion is smaller under the narrower team

⁵We drop wage observations below the 1st and above the 99th percentile.

⁶While there exist imputation methods we instead treat the censored observations as actual wage observations and do not correct for the censoring.

definition.

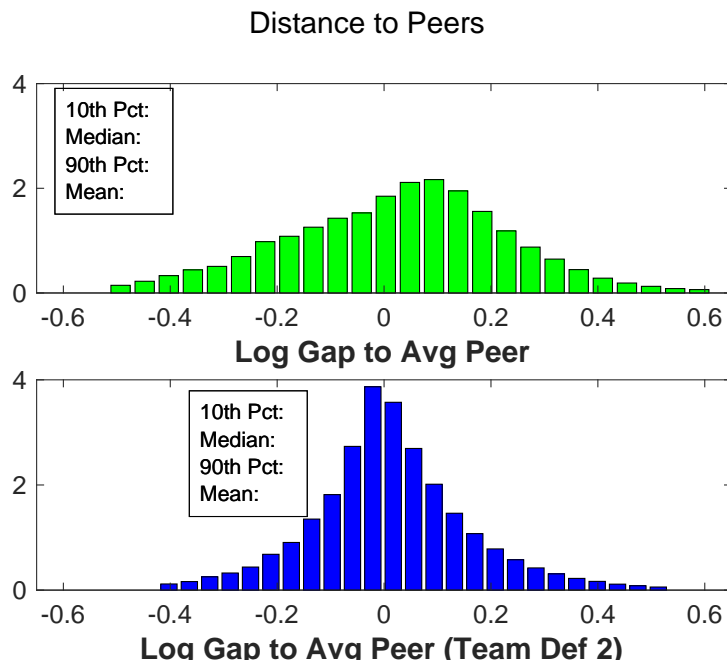


Figure 3 Distribution of wages gaps, two different team definitions.

Wage Growth We next construct n -year ahead annualized real wage growth (denoted $\Delta w_{i,t}^n$) which averages around 1% over all horizons when pooled across years t . The 10th, 50th, and 90th percentile for $n = 1$ are -6%, .6%, 8%, respectively, while the same deciles are -1.6%, .9%, and 3.6% for $n = 5$.

Correlations We compute a set of correlations of various wage moments at the team level. Specifically, Table I reports the correlation matrix of team average pay, team pay dispersion, team mean-median ratio (skewness), team size, and max wage at the team. All entries of the matrix are positive except the correlation between team average pay and the mean-median ratio.⁷

⁷A natural interpretation is that highly productive teams have a skewed wage distribution with very highly paid managers.

	Mean Wage	Wage sd	Mean/median	Team size	Max wage
Mean Wage	1				
Wage sd	0.51	1			
Mean/median	-0.01	0.20	1		
Team size	0.48	0.39	0.21	1	
Max wage	0.84	0.71	0.28	0.67	1

Table I Pairwise correlations at the team level, team definition 1. All variables *in logs*

3.2 Regression Framework

We begin with the following baseline specification which we implement separately for various horizons n ,

$$\Delta w_{i,t}^n = \alpha + \beta \hat{w}_{i,t} + \omega_o + \omega_a + \omega_e + \omega_g + \omega_t + \varepsilon_{i,t}. \quad (8)$$

In doing so, we pool the observations across all years t , controlling for a vector of time invariant occupation, age, education, gender, and year fixed effects Unless otherwise indicated that is the set of fixed effects used in all specifications. The results for $n > 2$ use only a subset of years t . For instance, for $n = 10$ we are restricted to exclusively use information about peers in the year 2000 (since we observe workers until 2010). Likewise, for $n = 5$ we can use all years between 2000 and 2005 .

We report the parameter estimates for team definition 2 in Table II, clustering standard errors at the establishment level. We first note that our findings suggest quantitatively large effects. \hat{w} has a standard deviation of .155 in the pooled sample, suggesting that a one standard deviation increase in \hat{w} is related to an increase in expected wage growth over the next year of one percentage point. At longer horizons, the average annualized growth rate is lower. This would be natural in the context of learning, as agents likely learn less as they gradually become more knowledgeable.

We next contrast these results with the corresponding results for team definition 1. As can be seen from Table III, the coefficients are somewhat smaller at all horizons for the broader team definition. These baseline results are thus consistent with learning from coworkers. The larger estimates when we use the second team definition are likely the result of more intense interactions between coworkers within occupations. From here on we restrict our attention

	Horizon in Years				
	1	2	3	5	10
\hat{w}	0.063*** (0.0030)	0.047*** (0.0022)	0.039*** (0.0020)	0.032*** (0.0017)	0.024*** (0.0015)
R^2	0.064	0.112	0.148	0.203	0.261
Observations	3544350	2911038	2384726	1576536	270837

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: $\hat{\beta}$ as estimated from specification (8). Team definition 2. Column titles indicate horizon n .

Table II Baseline Estimation Results

	Horizon in Years				
	1	2	3	5	10
\hat{w}	0.052*** (0.0024)	0.040*** (0.0019)	0.034*** (0.0017)	0.028*** (0.0015)	0.023*** (0.0014)
R^2	0.062	0.111	0.147	0.203	0.269
Observations	3672797	3013010	2465708	1626954	278289

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Identical to table II but applying team definition 1.

Table III Baseline Results, Broader Team Definition.

to team definition 2. We separately report all results for the alternative team definition in the Appendix.

We next offer several robustness checks. The first set of checks varies specification and sample selection. We also show that the effects we find are present and large across the wage distribution. We then show that there is a strong asymmetry in the patterns: wage growth is more strongly associated with the wage gap to higher paid workers than with the gap to lower paid workers. Finally, we aim to show that the results in Table II are not driven by back-loading or mean reversion. We first show that the patterns are robust to controlling for tenure. We then show that the patterns are still strong among those that switch firms.

	Different Subsamples					
	Benchmark	Size> 4	Size=2	Men Only	Age> 45	Age< 35
\hat{w}	0.039*** (0.0020)	0.042*** (0.0022)	0.014*** (0.00080)	0.046*** (0.0021)	0.028*** (0.0012)	0.056*** (0.0032)
R^2	0.148	0.153	0.091	0.153	0.095	0.155
Observations	2384726	2224894	60708	1700439	868537	534214

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: $\hat{\beta}$ as estimated from specification 8 under various sample selection modifications as indicated in column title. Team definition 2 at horizon $n = 3$ years.

Table IV Changing Sample Construction.

3.2.1 Robustness

We choose as benchmark for all our robustness results the coefficient estimate at $n = 3$ for team definition 2, $\hat{\beta} = .039$. We then run specification (8) for various modified sample selection criteria and report the results in table IV. For instance, the second column reports $\hat{\beta}$ when estimated on a sample of workers who are in teams of size > 4 at the reference date.

Our baseline results appear mostly insensitive to these modifications. However, it is clear that older workers seem to benefit less from more highly paid coworkers compared to younger workers.

We also check the robustness of our results to the inclusion of controls. To do so, we omit various fixed effects from the regression as indicated in table V. In addition to the differences in estimates by age category in IV, the sensitivity of wage growth to wage gap is larger when we omit age fixed effects. We revisit this issue when discussing wage back-loading below. The estimates seem robust with respect to other worker characteristics.

3.2.2 Across the Wage Distribution

We next show that the forces we document are present everywhere in the wage distribution. In particular, we run the same baseline specification separately for workers in different deciles of the population wage distribution.⁸

⁸Of course, we use the full peer group in the construction of the independent variable as always.

	(1)	(2)	(3)
	Age	Occupation	Education
\hat{w}	0.049*** (0.0021)	0.038*** (0.0019)	0.038*** (0.0021)
R^2	0.126	0.122	0.146
Observations	2384728	2384728	2481963

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: $\hat{\beta}$ as estimated from specification 8 under various fixed effect combinations. Column title indicate omitted fixed effects. Team definition 2 at horizon $n = 3$ years.

Table V Baseline results under varying fixed effects.

	Decile of the Wage Distribution									
	1	2	3	4	5	6	7	8	9	10
\hat{w}	0.039*** (0.0028)	0.051*** (0.0030)	0.045*** (0.0038)	0.036*** (0.0044)	0.033*** (0.0045)	0.034*** (0.0043)	0.036*** (0.0044)	0.037*** (0.0059)	0.049*** (0.0071)	0.012*** (0.0027)
R^2	0.185	0.228	0.254	0.218	0.189	0.195	0.170	0.172	0.198	0.088
Observations	13975	23508	27579	29574	30443	30699	31156	31101	30237	30641

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: $\hat{\beta}$ as estimated from specification (8) separately for different deciles of the wage distribution in the baseline sample in year t . Team definition 2 at horizon $n = 3$ years.

Table VI Baseline results for different deciles of the wage distribution.

The results are reported in Table VI. The results are fairly stable across the wage distribution.⁹ We conclude that, having more highly paid coworkers is associated with future wage growth, largely independent of the current level of wages. This is informative for the modeling choices we make below and also strongly suggests that our baseline findings do not reflect a form of economy-wide mean reversion in wages.

⁹The dropoff at the top decile is an artifact of the censoring. It also suggests that our treatment of the censored observation discussed above leads us to underestimate the strength of learning. Further, the fact that the number of observations is substantially smaller for the bottom deciles reflects that those workers are more marginally attached to the labor market and are therefore more frequently not employed at the reference spell n years ahead.

3.2.3 Asymmetries: The Role of the Distribution of Teammates

The previous results measure only whether the wage gap between a worker and her average coworkers is associated with wage growth. We are particularly interested in how the composition of one’s coworkers affects wage growth. Hence, we document next how the results presented in the previous subsection depend on characteristics of the wage distribution surrounding a worker. In particular, we try to shed light on whether these forces are symmetric from above and below, whether team size matters, and whether we can detect congestion and nonlinearities.

To this end, we employ various interactions, use flexible specifications that allow for nonlinearities, and measure the wage distribution in a team in more flexible parametric ways. Two broad features emerge: First, the level of wages of coworkers “below” in the wage distribution matters less for future wage growth. That is, most gains derive from those above in the wage distribution; we interpret this as evidence of little congestion effects in the learning process. Second, increasing the wage of those higher up in the wage distribution has a positive effect that declines somewhat as one moves to more distant coworkers. That is, marginally increasing the wage of a peer in close proximity benefits a worker more than increasing the wage of the top wage earner in the team.

Adding Interactions We begin by interacting \hat{w} with log team size, within-team wage dispersion, wage skewness, and the number as well as fraction of those more highly paid in the team. We report the results in Table VII.

We first highlight that size is insignificant. Thus, while there seems to be something special about very small teams as indicated in Table IV, overall learning seems to be independent of team size. This scale independence is an important input when choosing a parametric form for the learning process.

The remaining interactions suggest that, given \hat{w} , a less dispersed, less skewed distribution of wages increases future wage growth. Likewise, if there are more workers to learn from, that is workers higher up in the wage distribution, knowledge transmission also accelerates. We highlight, however, that these results could in principle be driven by two different forces. Either they could reflect the influence of the shape of the knowledge distribution a worker

	(1)	(2)	(3)	(4)	(5)
\hat{w}	0.026*** (0.0060)	0.067*** (0.0035)	0.042*** (0.0019)	0.019*** (0.0049)	0.0024 (0.0016)
$\hat{w} \times \text{Size}$	0.0032 (0.0019)				
$\hat{w} \times \text{Wage Dispersion}$		-0.13*** (0.012)			
$\hat{w} \times \text{Mean-Median}$			0.18*** (0.022)		
$\hat{w} \times \# \text{ Above}$				0.0048* (0.0020)	
$\hat{w} \times \% \text{ Above}$					0.044*** (0.0026)
R^2	0.149	0.151	0.151	0.151	0.156
Observations	2384726	2384726	2384726	2222579	2384726

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: $\hat{\beta}$ as estimated from specification (8) with the inclusion of additional interactions. Size is measured as log team size, wage dispersion is measured as the standard deviation of log wages, inverse skewness is measured as the log of the median/mean ratio. The forth column interacts with the log of the number of team-members above worker i in terms of wages, column (5) does so for the fraction. All regressions control for size, wage dispersion and skewness. Team definition 2 at horizon $n = 3$ years.

Table VII Baseline results interacted.

	(1)	(2)	(3)
\hat{w}^+	0.041*** (0.0028)		0.043*** (0.0034)
\hat{w}^-		0.020*** (0.0014)	0.018*** (0.0018)
R^2	0.131	0.121	0.136
Observations	2237206	2297090	2126790

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Column (3) reports the results from specification 9. Columns (2) and (1) only use information about those less well and those better paid, respectively. Team definition 2 at horizon $n = 3$ years.

Table VIII Splitting peers into those better and less well paid.

can learn from in which case they would suggest that workers learn more from those in close proximity. Alternatively, they may merely reflect that peers “above” and “below” yield distinct learning outcomes. In order to investigate further this possibility we next separately study the roles of agents that earn more than an individual and seem to play the role of her “teachers” and those that earn less and therefore play a role as her “students”.

Teammates Above and Below To study the differential role of teammates that earn more and those that earn less, we now split the peer group into those more highly and less well paid and run

$$\Delta w_{i,t}^n = \alpha + \beta_1 \hat{w}_{i,t}^+ + \beta_2 \hat{w}_{i,t}^- + \omega_o + \omega_a + \omega_e + \omega_g + \omega_t + \varepsilon_{i,t}. \quad (9)$$

Here $\hat{w}_{i,t}^+$ denotes the log wage distance of individual i to the mean wage of her more highly paid coworkers in period t while $\hat{w}_{i,t}^-$ denotes the corresponding distance to the mean wage of her less well paid coworkers.

Table XVI suggests that the marginal returns to improving peers above in the wage distribution far exceed the marginal returns to improving those below which suggests two main lessons.

This evidence suggests that learning from those with lower wages plays only a limited role.

It also suggest that having worse-paid coworkers does not hinder learning from others through congestion. Further, it suggests that our baseline results are driven primarily by learning from those higher up in the team-wage distribution. These observations are corroborated by the results we present next.

A More Flexible Specification We end the reduced-form exploration with an exercise that attempts to approximate the wage distribution surrounding a worker in a flexible way. To do so, we bin a worker’s peers into 11 brackets. The bottom bracket takes those peers with wage such that $\log(w_j) - \log(w_i) < -.5$ while the top bracket takes those peers with $\log(w_j) - \log(w_i) > .5$. All other workers are grouped into 9 equally spaced bins in-between. We then compute, for each individual i , the fraction of her coworkers in each bin and regress wage growth at various horizons on the bin-weights along with our standard controls and fixed effects. We present the results in figure 4 and report the underlying table in the appendix section A.1.¹⁰

Figure 4 plots the marginal response of n -year ahead annualized real wage growth to increasing the weight on each of the 10 bins (relative to increasing the weight on bin 1 which is the omitted category). To interpret, moving 10% of one’s peers from the bottom bin into the highest bin reduces 3-year ahead average wage growth by about .4 percentage points. That is, it represents the marginal effect of having more coworkers in that bin relative to bin 1. The figure does so separately for various other horizons n . The figure confirms the findings from the previous exercises: Those who are less well paid (those in bins 6 and under) have little effect on a worker’s future wage growth. In contrast, workers seem to benefit from additional highly paid workers in the peer group (those in bins 7 and higher). Note that, while workers benefit more from more highly paid peers the effects are less than proportional. This mimics our findings from the interaction specifications above which suggest that knowledge flows more efficiently from those in close proximity relative to those far above in the wage distribution. Nevertheless, we stress that the effects are monotonically increasing (almost everywhere) above implying that one learns more from workers further out in the wage distribution.

¹⁰We note that so far we only have those results for the first, broader team definition.

The bottom panel of Figure 4 also confirms the results from above which suggest that learning from higher earners is strongest over the one year horizon and declines monotonically as we extend the horizon of analysis. We conclude by pointing out that additional interaction specifications suggest that the results presented in Figure 4 are stable across agents in different percentiles of the wage distribution.

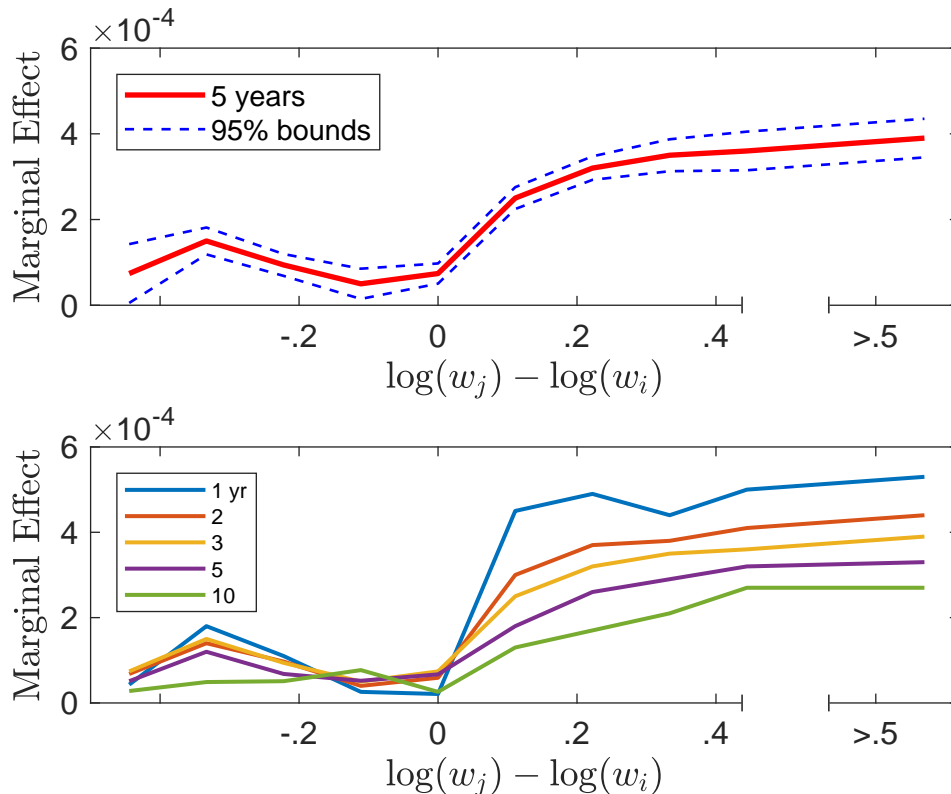


Figure 4 Dependent variable: n -year ahead average annualized wage growth expressed in percentage points. We plot the coefficients from a regression of wage growth on the fractions (multiplied by 100) of the team in the respective bins. The top panel only plots $n = 5$ along with the 95% confidence bands. The bottom panel plots estimates for different horizons. The lowest bin with $\log(w_j) - \log(w_i) < -0.5$ is the omitted category. All workers with $\log(w_j) - \log(w_i) > 0.5$ are in one single bin as indicated by the vertical dashed line. Standard errors clustered at the establishment level. The figure uses Team Definition 1.

3.2.4 Tenure and Backloading

Besides coworker learning there are two other prime mechanisms that could be driving the patterns uncovered in Table II. First, certain plausible wage back-loading patterns could

	Horizon in Years				
	1	2	3	5	10
\hat{w}	0.063*** (0.0013)	0.046*** (0.00095)	0.038*** (0.00085)	0.031*** (0.00078)	0.023*** (0.0011)
Tenure Quintile=2	-0.0035*** (0.00059)	-0.0031*** (0.00049)	-0.0027*** (0.00037)	-0.0022*** (0.00032)	-0.00068* (0.00031)
Tenure Quintile=3	-0.0032*** (0.00067)	-0.0034*** (0.00053)	-0.0030*** (0.00043)	-0.0023*** (0.00038)	-0.00046 (0.00041)
Tenure Quintile=4	-0.0036*** (0.00087)	-0.0031*** (0.00066)	-0.0026*** (0.00050)	-0.0022*** (0.00042)	-0.0031*** (0.00042)
Tenure Quintile=5	-0.0031** (0.00097)	-0.0040*** (0.00074)	-0.0038*** (0.00056)	-0.0033*** (0.00047)	-0.0035*** (0.00043)
R^2	0.064	0.113	0.149	0.204	0.267
Observations	3544350	2911038	2384726	1576536	270837

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: $\hat{\beta}$ as estimated from specification 8 controlling for tenure.

Table IX Baseline results controlling for job tenure.

give rise to the results. In particular, firms could attempt to retain workers by offering wage schedules that pay relatively more in the future. Firms could have incentives to do so if it is costly to hire new workers and workers search on the job. This would, for instance, be the case in an environment similar to [Burdett and Coles \(2003\)](#) or [Postel-Vinay and Robin \(2002\)](#).

To address this we add controls for job tenure into the baseline specification and report the results in [Table IX](#). The basic picture remains the same. Below, we also show that workers that switch teams benefit even more from having better coworkers. This alternative piece of evidence also helps rule out back-loading as a key driver of our findings.

3.2.5 Movers

ADD MOVER RESULTS HERE

4 Structural Estimation

We now turn to a structural estimation of the amount of learning within teams. One of the key problems interpreting the results in the previous section is that wages do not equal knowledge. In order to go beyond reduced-form relationships between the distribution of wages and wage growth and determine the implications of our findings for learning, we need a theory that allows us to map one into the other. We use the theory developed in Section 2 to do so. Our main objective is to estimate the “learning function” $G(\cdot)$. Below we describe a strategy to recover $G(\cdot)$ from panel data that includes teams’ wages, and implement our strategy using our German data.

4.1 Identifying Learning Parameters

Our identification strategy requires a panel of at least two years of matched employer-employee data that includes wages. We rely only on the worker’s Bellman equation,

$$V(z) = w(z, \tilde{z}) + \beta E[V(z') | z, \tilde{z}] \quad (10)$$

which is the result of the worker’s maximization. Equation (10) depends on the assumptions of stationarity, perpetual youth, perfect competition, and complete markets.¹¹ However, we do not need to place any assumptions on the set of firms that are active, or features of the technologies that firms use beyond those in Assumptions 1 to 3. The set of technologies and firms in the economy determine the set of teams we observe in equilibrium, but our strategy simply uses the set of observed teams.

We first note that z does not have a natural cardinality. We are therefore free to choose a convenient one: If $V(z)$ is the value function in the current equilibrium, we choose a

¹¹Our approach allows for a number of generalizations. For example, if markets are so incomplete that agents cannot save or borrow, we can simply replace the current return in (10) with a known increasing and concave function of the wage.

cardinality of z so that $V(z) = z$. Then, (10) becomes

$$\begin{aligned} z &= w(z, \tilde{z}) + \beta E[z'|z, \tilde{\mathbf{z}}] \\ &= w(z, \tilde{z}) + \beta \int_0^\infty z' dG(z'|z, \tilde{\mathbf{z}}) \end{aligned}$$

or

$$\begin{aligned} z_i &= w_i + \beta E[z'_i|z_i, \tilde{\mathbf{z}}_{-i}] \\ &= w_i + \beta \int_0^\infty z'_i dG(z'_i|z_i, \tilde{\mathbf{z}}_{-i}). \end{aligned} \tag{11}$$

Our strategy hinges on the following two observations. First, if we know, for each worker i , z'_i , z_i , and $\tilde{\mathbf{z}}_{-i}$, we can directly identify G . Conversely, if we know G , we can invert (11) and solve for z_i as a function of a worker's wage and the wages of her coworkers; for a team of size N , (11) for each of the N team members delivers a system of N equations in N unknowns (z_i for each team member). Together these equations provide several moment conditions that can be used to identify G using GMM.

Operationally, we choose a functional form for $G(z'|z, \tilde{z}; \theta)$, with parameters θ , and we calibrate β externally. Starting from period t , we can decompose next period's knowledge, z' , into expected and unexpected components. Namely,

$$z'_i = \mathcal{E}(z_i, \tilde{\mathbf{z}}_{-i}) + \varepsilon_i, \tag{12}$$

where $\mathcal{E}(z_i, \tilde{\mathbf{z}}_{-i})$ is the conditional expectation and ε_i is the expectational error. We then use the moment conditions built from

$$E[\varepsilon_i|z_i, \tilde{\mathbf{z}}_{-i}] = 0.$$

Below we specialize to the case where $\mathcal{E}(z_i, \tilde{\mathbf{z}}_{-i}) = E[z'_i|z_i, \tilde{\mathbf{z}}_{-i}]$ is a linear combination of several moments of $\{m_k(z_i, \tilde{\mathbf{z}}_{-i})\}_{k=1}^K$ so that $\mathcal{E}(z_i, \tilde{\mathbf{z}}_{-i}) = \sum_{k=1}^K \theta_k m_k(z_i, \tilde{\mathbf{z}}_{-i})$. In such a case,

we would have K parameters $\{\theta_k\}$ and K natural moment conditions

$$E [m_k(z_i, \tilde{\mathbf{z}}_{-i}) \varepsilon_i] = 0, \quad k = 1, \dots, K \quad (13)$$

Formally, if a team has N workers, then given θ and w , (11) provide N equations for the N unknowns of $\{z_i\}$. Therefore given the wages w_t and a vector of team assignments r_t , we can construct $Z(w_t, r_t, \theta)$ to be the $I \times 1$ vector of all workers' knowledge at t . Given this, we can construct $M(w_t, r_t, \theta)$ to be the $I \times k$ matrix of moments so that the i, k entry of $M(w_t, r_t, \theta)$ is $m_k(z_i, \tilde{\mathbf{z}}_{-i})$ where $z_i, \tilde{\mathbf{z}}_{-i}$ are the knowledge of i and her coworkers implied by the wages, w_t , the assignment r_t , and parameters θ . Then the k moments conditions (13) can be stacked as

$$E \left[M(w_t, r_t, \theta)^T (Z(w_{t+1}, r_{t+1}, \theta) - M(w_t, r_t, \theta) \theta) \right] = 0. \quad (14)$$

We solve for θ using an iterative two-step procedure that exploits the panel structure of our data along with the intertemporal restrictions inherent in the learning function (12). We first guess parameters θ^{guess} . Given this guess, we can back out the types z in a team solely from information on wages.¹² In other words, we invert (11) to solve for all workers' knowledge, $Z(w_t, r_t, \theta)$. We do this by finding a fixed point \mathbf{z} of the operator

$$T(\mathbf{z}) = \left\{ w_i + \beta \int z' dG(z' | z_i, \tilde{\mathbf{z}}_{-i}; \theta) \right\}_i$$

We can then use the wages at time $t + 1$ to solve for all workers knowledge at $t + 1$, $Z(w_{t+1}, r_{t+1}, \theta)$. With this, we have the implied values of $z_i, \tilde{\mathbf{z}}_{-i}$, and z'_i for each worker. We then use these knowledge levels to estimate θ using a linear regression

$$z_{it+1} = \sum_{k=1}^K \theta_k m_k(z_{it}, \tilde{\mathbf{z}}_{-it}) + \varepsilon_{it}.$$

If our estimated $\hat{\theta} = \theta^{guess}$ then we have found a fixed point. This fixed point is a solution

¹²As discussed in the theory section above, the vector of types \mathbf{z} is the solution to the firm problem in (1). Here, we simply use the composition of teams observed in the data.

to (14). In practice we use $\hat{\theta}$ to update our guess and iterate until we find a fixed point.

While we currently do not have a proof of identification, this method has always uncovered the true parameter values in Monte Carlo simulations and has always converged when implemented on the matched German data.

4.2 Results

Guided by our reduced form findings, we focus on the following parametric form for the conditional expectation, that implicitly determines $G(\cdot)$,

$$E[z'_i|z_i, \tilde{\mathbf{z}}_{-i}] = \int_0^\infty z'_i dG(z'_i|z_i, \tilde{\mathbf{z}}_{-i}; \theta) = \frac{1}{N-1} \sum_{j \neq i} z_j \Theta\left(\frac{z_j}{z_i}\right) \quad (15)$$

where N is the worker's team size and $\Theta(\cdot)$ is weakly increasing functions. Below we let $\Theta(\cdot)$ be piece-wise linear. We focus on the expected value because this is the only feature of the function G needed to invert the Bellman equation and recover the workers' knowledge. This functional form could be motivated by a variant of the model in Lucas (2009) in which a worker is equally likely to attain knowledge from any coworker, and the function Θ describes how the worker's learning depends on the gap between the worker and the coworker. In contrast to Lucas (2009), however, here agents only learn from coworkers, not from the whole population.

We begin by studying the parametric learning function $\Theta(x) = \begin{cases} 1 + \theta_0 + \bar{\theta}(x-1), & x \geq 1 \\ 1 + \theta_0 + \underline{\theta}(x-1), & x < 1 \end{cases}$,
or

$$E[z'_i - z_i|z_i, \tilde{\mathbf{z}}_{-i}] = \theta_0 z_i + \frac{1}{N-1} \left\{ \underline{\theta} \sum_{z_j < z_i} (z_j - z_i) + \bar{\theta} \sum_{z_j \geq z_i} (z_j - z_i) \right\}. \quad (16)$$

This learning function allows for asymmetric learning from types z_j on a worker z_i depending on whether $z_j > z_i$ or vice versa. It also sets allows for a constant time trend in skill growth, θ_0 .

In updating our guess for $\theta = \{\bar{\theta}, \underline{\theta}, \theta_0\}$, we make use of the linear structure of the learning function and regress $z'_i - z_i$ on z_i , $\frac{1}{N-1} \sum_{z_j < z_i} (z_j - z_i)$, and $\frac{1}{N-1} \sum_{z_j > z_i} (z_j - z_i)$. Note that all the information used in this regression is constructed from the purely cross-sectional

	Team Definition	
	1	2
$\bar{\theta}$.0702 (.0005)	.0932 (.0008)
$\underline{\theta}$.0076 (.0004)	.0395 (.0006)
θ_0	.0015 (.0001)	.0041 (.0001)
Observations	3327679	3198654

GMM standard errors in parentheses

Notes: GMM standard errors in parentheses.

Table X Parametric estimation results for the learning function (16).

dimension of the data in the first step.

For this baseline learning function, we report our parameter estimates along with the associated standard errors in Table XIII.¹³

Choosing the expected present value of earnings as the cardinality of z allows for a natural interpretation of these estimates. In particular, the point estimates suggest that raising the average expected present value of earnings of a worker’s more highly paid coworkers in the establishment by \$1 raises that workers expected present value of earnings by \$0.07 over the next year. In turn, doing so for the coworkers that are less well paid only increases expected present value of earnings by less than 1 cent. This implies that there are only very small learning effects coming from these workers, and there is no evidence of congestion. Naturally, we find somewhat larger effects for the narrower team definition.¹⁴ Clearly, these point estimates are very much consistent with the reduced form patterns discussed in the previous section.

Finally, while α is precisely estimated and strictly positive it is very small for both team definitions. One reason for why we find essentially no trend growth in wages beyond what

¹³The only other parameter we need to choose is β which we set to .95 (annual) here. The results presented here are not very sensitive to this choice. All results are currently for the second, narrower, team definition.

¹⁴One important observation across all specifications we have worked with is that $\underline{\theta}$ is substantially larger for team definition 2.

arises from learning is that the average real wage growth during the period covered in our dataset was very limited as discussed above.

The next step is to generalize the specification of G to allow for additional flexibility in order to capture potential nonlinearities in coworker learning. Hence, we specify the learning function to be defined by $\Theta(1) = \theta_0$ and

$$\Theta'(x) = \begin{cases} \bar{\theta}_2, & x \geq 1 + b \\ \bar{\theta}_1, & 1 \leq x < 1 + b \\ \underline{\theta}, & x < 1 \end{cases}$$

Θ is a continuous and piecewise linear function with kinks at $x = 1$ and $x = b$, and corresponds to the conditional expectation

$$E[z'_i - z_i | z_i, \tilde{\mathbf{z}}_{-i}] = \theta_0 z_i + \frac{1}{N-1} \left\{ \sum_{z_j < z_i} \underline{\theta} (z_j - z_i) + \sum_{z_j > z_i} [\bar{\theta}_1 (z_j - z_i) + \mathbf{1}_{z_j > (1+b)z_i} (\bar{\theta}_2 - \bar{\theta}_1) (z_j - z_i - bz_i)] \right\} \quad (17)$$

where $\mathbf{1}$ denotes the indicator function. This piece-wise linear function incorporates additional flexibility yet still allows us to linearly project $z' - z$ on the right-hand-side to update the four parameters of the learning function $\{\theta_0, \underline{\theta}, \bar{\theta}_1, \bar{\theta}_2\}$.¹⁵ When implementing this learning function, we set b such that it groups all workers with expected present value type $z_j > 1.1z_i$ into the bin that has slope $\bar{\theta}_2$.

So far, we only have results for team definition 2. They are reported in table [XIV](#). Our estimates for θ_0 and $\underline{\theta}$ are hardly changed by the modification of the learning function. That is, as before, changing the wages of those team members with lower type affects an individual's expected wage growth little in comparison with those team members with more knowledge. Likewise, the estimated trend growth remains minimal. Our results indicate that $\bar{\theta}_1 > \bar{\theta}_2$, and so the marginal returns (in terms of future wage growth) to improving the knowledge of those above in the wage distribution appears to be somewhat larger for those

¹⁵In light of our previous findings and due to computational limitations we have thus far restricted the learning function to take a single parameter for the group $z_j < z_i$. This restriction could, in principle, be relaxed.

Parametric Estimation Results	
$\bar{\theta}_1$.1148 (.001)
$\bar{\theta}_2$.0905 (.0008)
$\underline{\theta}$.037 (.0005)
θ_0	.0038 (.0001)
Observations	3198654

Standard errors in parentheses. Results for team definition 2 only.

Notes: GMM standard errors in parentheses.

Table XI Estimates for the learning function (17).

in closer proximity in the distribution of knowledge.

We conclude this subsection with two short exercises which cast light on the quantitative importance of coworker learning and its interplay with how teams are formed.

4.2.1 Investment in Knowledge

An individual receives compensation in two ways: with wages and with knowledge. We can use our estimated framework to gauge the quantitative importance of coworker learning in the economy by comparing the value of knowledge flows to the value of wage payments. Specifically, we compute the value of the annual flow of knowledge, $\beta(z'_i - z)$, where next period's knowledge z'_i is given by equation (12) and compute its simple pooled average across all individuals and years in our sample. We then subtract the pure trend component $\beta\theta_0z$ from this and contrast it with the pooled average of wages.

Doing so reveals that coworkers knowledge flows account for roughly 4.3% of the average flow value workers receive, with the remaining 95.6% given by the wage.¹⁶ In other words, agents invest on average 4.3% of their total compensation in learning from others at work. We note that this is likely a conservative estimate given the narrow team definition.

¹⁶We note that when we do not subtract the trend component this number rises to 12.4%.

Furthermore, there is naturally substantial heterogeneity in this breakdown across the knowledge distribution. This number rises to 7.8% for the bottom decile of the knowledge distribution reflecting the substantial room for learning at the bottom. In turn, it becomes negative at the top decile, dropping to -.9% reflecting the mild negative effect of having mostly less knowledgeable coworkers.

4.2.2 The Role of Sorting

In equilibrium, the team selected by a firm produces both output and knowledge. As a result, the sorting of workers across firms reflects both of these goals. How does the sorting that occurs in equilibrium affect the value of learning that occurs within teams? How much would the total value of learning change if teams were formed randomly?

To assess the role of coworker sorting for learning, we conduct a simple experiment where we randomly reshuffle workers across existing teams in the final coworker year in our sample, 2008. We leave the team size distribution unaltered and compute, for all workers, a counterfactual conditional expectation $\mathcal{E}(z_i, \tilde{\mathbf{z}}_{-i}^{\text{cf}})$ for z_i' where $\tilde{\mathbf{z}}_{-i}^{\text{cf}}$ is worker i 's counterfactual peer group.

We then contrast the average of the counterfactual conditional expectation with the average of the factual conditional expectation, $\mathcal{E}(z_i, \tilde{\mathbf{z}}_{-i})$. Doing so reveals that under random sorting the average growth in z rises by 113%. Thus, workers are allocated to teams in a way that hinders knowledge flows relative to a random sorting benchmark. We interpret this as reflecting supermodularity in the production function, which results in positive assortative matching of workers in teams. In other words, these findings seem to suggest a tension between the contemporaneous requirements on the production side and the dynamic returns from coworker learning.

4.2.3 An Alternative Interpretation

The equations characterizing an individual's value function centered around the idea that the change in one's knowledge depends on the distribution of knowledge among one's coworkers. Here we state there is an alternative interpretation of the equations that characterize learning that is consistent with our empirical specification of learning.

Suppose that z represents a vector of characteristics. For example these characteristics could reflect different dimensions of knowledge. In such an environment, one possibility is that learning is such that the change in one's value function depended on the composition of one's coworkers' value functions. Namely, in this case it is natural to assume that the learning function depends on values, not on knowledge directly. Similarly, since in this case knowledge cannot be equated to value, the procedure outlined above to obtain an individual's knowledge from a panel of wages simply recovers the values of all agents, as in equation (10). Under this assumption, our methodology can be used exactly as described, and our quantitative results would be unchanged. We would simply interpret the estimated learning function as determining how the value of other agents determines the change in value of a given individual.

5 Incorporating Other Observables

We now show how our methodology can be extended to a setting in which either production function or the learning function (or even the value placed on knowledge) depends on worker characteristics aside from knowledge. These characteristics may or may not evolve endogenously. We require that the characteristics are observable.

An individual is described by knowledge z and a vector of observable characteristics x . These evolve according to a joint Markov process. For example, x could consist of an individual's age, schooling, occupation, location, etc. The joint Markov process is characterized by

$$G(z', x' | z, x, \{\tilde{\mathbf{z}}, \tilde{\mathbf{x}}\})$$

The value function for an individual with state z, x is $V(z; x)$ satisfying

$$V(z, x) = w(z, x; \{\tilde{\mathbf{z}}, \tilde{\mathbf{x}}\}) + \beta \int \int V(z', x') G(dz', dx' | z, x, \{\tilde{\mathbf{z}}, \tilde{\mathbf{x}}\})$$

Thus both the production function and the learning function can depend on the individual characteristics and those of her coworkers. The realized value function of worker i at time t

is

$$v_{it} = w_{it} + \beta \int \int V(z', x') G(dz', dx' | z_{it}, x_{it}, \{\tilde{\mathbf{z}}_{it}, \tilde{\mathbf{x}}_{it}\}) \quad (18)$$

The key step is to transform the learning function from the knowledge space to the value space. We require that V is strictly increasing in z for each x , and therefore has a partial inverse $Z(v, x)$ that satisfies $v = V(Z(v, x), x)$ and $z = Z(V(z, x), x)$. Further, define the learning function in the value space as

$$\hat{G}(v', x' | v, x, \tilde{\mathbf{v}}, \tilde{\mathbf{x}}) \equiv G\left(Z(v', x'), x' | Z(v, x), x, \tilde{\mathbf{Z}}(\tilde{\mathbf{v}}, \tilde{\mathbf{x}}), \tilde{\mathbf{x}}\right)$$

where $\tilde{\mathbf{Z}}(\tilde{\mathbf{v}}; \tilde{\mathbf{x}})$ is the vector of coworkers' knowledge given their values and characteristics. With this, we can write (18) as

$$v_{it} = w_{it} + \beta \int \int v' \hat{G}(dv', dx' | v, x, \tilde{\mathbf{v}}, \tilde{\mathbf{x}}) \quad (19)$$

5.1 Algorithm

We now show that is straightforward to extend the methodology described in the previous section to estimate the function \hat{G} . Let \mathcal{E} be the conditional expectation:

$$\mathcal{E}(v, x, \tilde{\mathbf{v}}, \tilde{\mathbf{x}}) \equiv E[v' | v, x, \tilde{\mathbf{v}}, \tilde{\mathbf{x}}] = \int \int v' \hat{G}(dv', dx' | v, x, \tilde{\mathbf{v}}, \tilde{\mathbf{x}})$$

so that the realized Bellman equation (19) can be written as

$$v_{it} = w_{it} + \beta \mathcal{E}(v_{it}, x_{it}, \tilde{\mathbf{v}}_{-it}, \tilde{\mathbf{x}}_{-it}). \quad (20)$$

If we observed $\{v'_i, x'_i, v_i, x_i, \tilde{\mathbf{v}}_{-i}, \tilde{\mathbf{x}}_{-i}\}$ for each worker, we could directly estimate \mathcal{E} . And conversely, if we knew \mathcal{E} and we observed $\{w_{it}, x_{it}, \tilde{\mathbf{x}}_{-i}, x'_i\}$ for each worker, we could solve for $\{v_{it}\}$ for each team using the system of equations given by (20) for each team member.

We can again cast this in terms of GMM. For example, suppose we assume \mathcal{E} takes the

form of a linear combination of moments $\{m_n(v, x, \tilde{\mathbf{v}}, \tilde{\mathbf{x}})\}_{n=1}^N$

$$\mathcal{E}(v, x, \tilde{\mathbf{v}}, \tilde{\mathbf{x}}) = \sum_{k=1}^K \theta_k m_k(v, x, \tilde{\mathbf{v}}, \tilde{\mathbf{x}})$$

Next, define the expectational error term ε_{it+1} to be

$$\begin{aligned} \varepsilon_{it+1} &\equiv v_{it+1} - E[v_{it+1} | v_{it}, x_{it}, \tilde{\mathbf{v}}_{-it}, \tilde{\mathbf{x}}_{-it}] \\ &= v_{it+1} - \mathcal{E}(v_{it}, x_{it}, \tilde{\mathbf{v}}_{-it}, \tilde{\mathbf{x}}_{-it}) \end{aligned}$$

We can again build moment conditions to estimate θ from

$$E[\varepsilon_{it+1} | v_{it}, x_{it}, \tilde{\mathbf{v}}_{it}, \tilde{\mathbf{x}}_{it}] = 0.$$

Our natural moment conditions would be

$$E[\varepsilon_{it+1} m_k(v_{it}, x_{it}, \tilde{\mathbf{v}}_{it}, \tilde{\mathbf{x}}_{it})] = 0$$

Formally, given θ , we can solve for $\{v_{it}\}$ using (20). Given the entire vector of wages w , observable characteristics x , and team assignments r , let $\Upsilon(w, x, r, \theta)$ be the corresponding values that have been solved for using (20) so that

$$\{v_{it}\} = \Upsilon(w_t, x_t, r_t, \theta).$$

Given this, we can construct $M(w_t, x_t, r_t, \theta)$ to be the $I \times k$ matrix of moments so that the i, k entry of $M(w_t, x_t, r_t, \theta)$ is $m_k(v_i, x_i, \tilde{\mathbf{v}}_{-i}, \tilde{\mathbf{x}}_{-i})$ where $v_i, \tilde{\mathbf{v}}_{-i}$ are the values of i and her coworkers implied by the wages, w_t , the observable characteristics x_t , the assignment r_t , and parameters θ . Then the k moments conditions (13) can be stacked as

$$E \left[M(w_t, x_t, r_t, \theta)^T (\Upsilon(w_{t+1}, x_t, r_{t+1}, \theta) - M(w_t, x_t, r_t, \theta) \theta) \right] = 0.$$

5.2 Results

We illustrate this methodology by studying the differences in learning between young and old workers. Let $\{y, o\}$ indicate whether a worker is young or old. We implement the following parametric form for the conditional expectation

$$\begin{aligned}
 E_y [v'_i | v_i, \tilde{\mathbf{v}}_{-i}] &= \theta_{0,y} v_i + \bar{\theta}_{yy} \frac{1}{N-1} \sum_{v_j > v_i, j \text{ young}} (v_j - v_i) + \bar{\theta}_{yo} \frac{1}{N-1} \sum_{v_j > v_i, j \text{ old}} (v_j - v_i) \\
 E_o [v'_i | v_i, \tilde{\mathbf{v}}_{-i}] &= \theta_{0,o} v_i + \bar{\theta}_{oy} \frac{1}{N-1} \sum_{v_j > v_i, j \text{ young}} (v_j - v_i) + \bar{\theta}_{oo} \frac{1}{N-1} \sum_{v_j > v_i, j \text{ old}} (v_j - v_i).
 \end{aligned}
 \tag{21}$$

This specification allows knowledge flows to depend on both the age group of the worker and the age group of her coworker. For instance, $\bar{\theta}_{yo}$ captures the strength of the knowledge flows from old to young coworkers. Furthermore, the specification allows for age group specific trend growth. Note that this specification does not allow for any effects of coworkers j with $v_j < v_i$.

In practice, we do not regroup workers across age groups and instead label them y when younger than 40 in the year 2000. We then implement our routine on a short panel, using only information from the years 2000-2002. The rest of the implementation follows exactly the same routine outlined in the previous section.

We present our results in table [XV](#). In line with our previous findings we find little trend growth for either age group. We further find that each group learns more from itself, that is the young learn more from the young than from the old and vice versa for the old. Finally, the young learn substantially more than the old, likewise closely in line with our reduced form findings.

6 Conclusion

We set out to study learning from coworkers. We found strong evidence that in fact this form of learning is active. Our results are intuitive and natural. Workers learn only from agents that are more knowledgeable than they are. Other, less knowledgeable coworkers neither help

Parametric Estimation Results	
trend growth young: $\theta_{0,y}$.0104 (.)
learning from young to young: $\bar{\theta}_{yy}$.2077 (.)
learning from old to young: $\bar{\theta}_{yo}$.1702 (.)
trend growth old: $\theta_{0,o}$.0038 (.)
learning from young to old: $\bar{\theta}_{oy}$.0476 (.)
learning from old to old: $\bar{\theta}_{oo}$.1035 (.)
Observations	912558

Standard errors still under construction. Results for team definition 2 only.

Notes: Old: 40 and older in year 2000.

Table XII Estimates for the learning function (21).

nor hinder their learning. We also find that learning from more knowledgeable workers is increasing in the coworker's knowledge. That is, it is better to work with more knowledgeable teammates, and the benefits per unit of their knowledge seems to be increasing.

We hope that these findings are useful in encouraging work on more theories and empirical research with learning from coworkers at their core. Our theory, although general in its specification of technology and existing complementarities in production, does assume that workers are simply income maximizers and that labor markets are competitive. It would be valuable to refine estimates of the value of learning in the workplace by relaxing these assumptions.

Finally, the importance of learning from coworkers implied by our findings suggests large aggregate growth consequences of any economy-wide change that affects the composition of teams. Many such changes come to mind, like, for example, technological improvements in information and communication technology, other forms of skilled-biased technical change, as well increased spatial segregation. Our results underscore the importance of studying

these and other well-known trends in the economy from the point of view of their effect on team formation and the resulting learning from coworkers.

References

- Akcigit, Ufuk, Santiago Caicedo, Ernest Miguez, Stefanie Stantcheva, and Valerio Sterzi**, “Dancing with the Stars: Innovation Through Interactions,” Technical Report, National Bureau of Economic Research 2018.
- Alvarez, Fernando E, Francisco J Buera, and Robert E Lucas Jr**, “Idea Flows, Economic Growth, and Trade,” Technical Report, National Bureau of Economic Research 2013.
- Anderson, Axel and Lones Smith**, “Dynamic Matching and Evolving Reputations,” *The Review of Economic Studies*, 2010, *77* (1), 3–29.
- Buera, Francisco J and Ezra Oberfield**, “The Global Diffusion of Ideas,” Technical Report, National Bureau of Economic Research 2016.
- Burdett, Ken and Melvyn Coles**, “Equilibrium Wage-Tenure Contracts,” *Econometrica*, 2003, *71* (5), 1377–1404.
- Burstein, Ariel T and Alexander Monge-Naranjo**, “Foreign Know-how, Firm Control, and the Income of Developing Countries,” *The Quarterly Journal of Economics*, 2009, *124* (1), 149–195.
- Caicedo, Santiago, Robert E Lucas Jr, and Esteban Rossi-Hansberg**, “Learning, Career Paths, and the Distribution of Wages,” Technical Report, National Bureau of Economic Research 2016.
- Cornelissen, Thomas, Christian Dustmann, and Uta Schönberg**, “Peer Effects in the Workplace,” *American Economic Review*, 2017, *107* (2), 425–56.
- Garicano, Luis and Esteban Rossi-Hansberg**, “Organization and inequality in a knowledge economy,” *The Quarterly Journal of Economics*, 2006, *121* (4), 1383–1435.
- Herkenhoff, Kyle, Jeremy Lise, Guido Menzio, and Gordon Phillips**, “Worker Mobility and the Diffusion of Knowledge,” Technical Report, mimeo 2018.

- Jovanovic, Boyan**, “Misallocation and Growth,” *American Economic Review*, 2014, *104* (4), 1149–71.
- **and Glenn M MacDonald**, “Competitive Diffusion,” *Journal of Political Economy*, February 1994, *102* (1), 24–52.
- **and Rafael Rob**, “The Growth and Diffusion of Knowledge,” *Review of Economic Studies*, October 1989, *56* (4), 569–82.
- König, Michael D, Jan Lorenz, and Fabrizio Zilibotti**, “Innovation vs. imitation and the evolution of productivity distributions,” *Theoretical Economics*, 2016, *11* (3), 1053–1102.
- Lucas, Robert E.**, “Ideas and Growth,” *Economica*, 02 2009, *76* (301), 1–19.
- **and Benjamin Moll**, “Knowledge Growth and the Allocation of Time,” *Journal of Political Economy*, February 2014, *122* (1), pp. 1–51.
- Luttmer, Erzo GJ et al.**, “An Assignment Model of Knowledge Diffusion and Income Inequality,” *Federal Reserve Bank of Minneapolis Research Department Staff Report*, 2014, 715.
- Mas, Alexandre and Enrico Moretti**, “Peers at Work,” *American Economic Review*, 2009, *99* (1), 112–45.
- Nix, Emily**, “Learning Spillovers in the Firm,” Technical Report, mimeo 2015.
- Perla, Jesse and Christopher Tonetti**, “Equilibrium Imitation and Growth,” *Journal of Political Economy*, 2014, *122* (1), 52 – 76.
- , – , **and Michael E Waugh**, “Equilibrium Technology Diffusion, Trade, and Growth,” Technical Report, National Bureau of Economic Research 2015.
- Postel-Vinay, Fabien and Jean-Marc Robin**, “Equilibrium Wage Dispersion with Worker and Employer Heterogeneity,” *Econometrica*, 2002, *70* (6), 2295–2350.

Sampson, Thomas, “Dynamic Selection: An Idea Flows Theory of Entry, Trade and Growth*,” *The Quarterly Journal of Economics*, 2015.

Appendix

A Additional reduced form empirical results

A.1 Figure 4

	Horizon in Years				
	1	2	3	5	10
p_2	0.000043 (0.000037)	0.000068* (0.000032)	0.000074* (0.000035)	0.000051 (0.000034)	0.000028 (0.000021)
p_3	0.00018*** (0.000031)	0.00014*** (0.000019)	0.00015*** (0.000016)	0.00012*** (0.000015)	0.000049* (0.000019)
p_4	0.00011*** (0.000021)	0.000097*** (0.000015)	0.000094*** (0.000013)	0.000068*** (0.000013)	0.000051*** (0.000015)
p_5	0.000026 (0.000026)	0.000040 (0.000021)	0.000050** (0.000018)	0.000052** (0.000017)	0.000077*** (0.000016)
p_6	0.000021 (0.000015)	0.000059*** (0.000012)	0.000074*** (0.000012)	0.000067*** (0.000012)	0.000026* (0.000012)
p_7	0.00045*** (0.000017)	0.00030*** (0.000013)	0.00025*** (0.000013)	0.00018*** (0.000013)	0.00013*** (0.000015)
p_8	0.00049*** (0.000033)	0.00037*** (0.000022)	0.00032*** (0.000014)	0.00026*** (0.000013)	0.00017*** (0.000016)
p_9	0.00044*** (0.000025)	0.00038*** (0.000022)	0.00035*** (0.000019)	0.00029*** (0.000017)	0.00021*** (0.000015)
p_{10}	0.00050*** (0.000035)	0.00041*** (0.000026)	0.00036*** (0.000023)	0.00032*** (0.000021)	0.00027*** (0.000018)
p_{11}	0.00053*** (0.000028)	0.00044*** (0.000023)	0.00039*** (0.000023)	0.00033*** (0.000022)	0.00027*** (0.000021)
R^2	0.068	0.118	0.154	0.210	0.278
Observations	3680375	3018284	2469598	1629217	278415

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Each row reports the coefficient on the weight of bins 2 through 11 where the weight on the bottom bin is the omitted category. Each column corresponds to one line in figure 4. Team definition 1.

Table XIII Results for figure 4.

	Different Subsamples					
	Benchmark	Size> 4	Size=2	Men Only	Age> 45	Age< 35
\hat{w}	0.034*** (0.0017)	0.034*** (0.0017)	0.0077* (0.0033)	0.040*** (0.0020)	0.024*** (0.0010)	0.048*** (0.0029)
R^2	0.147	0.148	0.175	0.154	0.094	0.154
Observations	2465708	2458090	2215	1754938	907704	545245

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: $\hat{\beta}$ as estimated from specification 8 under various sample selection modifications as indicated in column title. Team definition 1 at horizon $n = 3$ years.

Table XIV Changing Sample Construction—Counterpart to table IV

	(1) Age	(2) Occupation	(3) Education
\hat{w}	0.042*** (0.00078)	0.025*** (0.00067)	0.032*** (0.0008)
R^2	0.124	0.112	0.145
Observations	2465710	2465710	2568620

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: $\hat{\beta}$ as estimated from specification 8 under various fixed effect combinations. Column title indicate omitted fixed effects. Last column *includes* establishment fixed effects. Team definition 1 at horizon $n = 3$ years.

Table XV Baseline results under varying fixed effects—Counterpart to table V

A.2 Tables for alternative team definition

This subsection reports all empirical results from the main body of the paper for the alternative team definition 1.

	Horizon in Years				
	1	2	3	5	10
\hat{w}	0.051*** (0.0012)	0.039*** (0.00088)	0.033*** (0.00076)	0.028*** (0.00071)	0.022*** (0.0011)
Tenure Quintile=2	-0.0037*** (0.00057)	-0.0031*** (0.00046)	-0.0027*** (0.00036)	-0.0022*** (0.00031)	-0.00072* (0.00030)
Tenure Quintile=3	-0.0035*** (0.00065)	-0.0035*** (0.00051)	-0.0031*** (0.00042)	-0.0024*** (0.00036)	-0.00068 (0.00039)
Tenure Quintile=4	-0.0039*** (0.00084)	-0.0032*** (0.00064)	-0.0027*** (0.00048)	-0.0022*** (0.00041)	-0.0028*** (0.00041)
Tenure Quintile=5	-0.0031** (0.00095)	-0.0038*** (0.00072)	-0.0035*** (0.00055)	-0.0030*** (0.00046)	-0.0032*** (0.00042)
R^2	0.063	0.112	0.149	0.204	0.274
Observations	3672797	3013010	2465708	1626954	278289

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: $\hat{\beta}$ as estimated from specification 8 controlling for tenure.

Table XVI Baseline results controlling for job tenure—Counterpart to table IX

	Horizon in years				
	1	2	3	5	10
\hat{w}	0.13*** (0.0091)	0.07*** (0.0026)	0.056*** (0.0017)	0.047*** (0.0014)	0.034*** (0.002)
R^2	0.172	0.132	0.172	0.257	0.390
Observations	4801	36588	30357	20464	3443

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: $\hat{\beta}$ as estimated from specification (8) on a sample of establishment switchers who experienced an interim unemployment spell. See definition in the main text.

Table XVII Baseline results for establishment switchers with an interim unemployment spell. Counterpart to table X

	Horizon in years				
	1	2	3	5	10
\hat{w}	0.099 (0.074)	0.086*** (0.001)	0.055*** (0.0075)	0.044*** (0.0066)	0.073 (0.027)
R^2	0.726	0.255	0.276	0.433	0.652
Observations	144	1371	1139	692	64

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: $\hat{\beta}$ as estimated from specification (8) on a sample of establishment switchers who experienced an interim unemployment spell. See definition in the main text.

Table XVIII Baseline results for establishment mass-layoff switchers. Counterpart to table XI

	Decile of the Wage Distribution									
	1	2	3	4	5	6	7	8	9	10
\hat{w}	0.034*** (0.0025)	0.046*** (0.0028)	0.041*** (0.0033)	0.037*** (0.0043)	0.033*** (0.0046)	0.030*** (0.0047)	0.031*** (0.0055)	0.035*** (0.0066)	0.037*** (0.0074)	0.0071* (0.0029)
R^2	0.180	0.226	0.251	0.218	0.188	0.190	0.165	0.170	0.185	0.079
Observations	15237	24877	28705	30506	31234	31505	31932	31930	31133	31006

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: $\hat{\beta}$ as estimated from specification (8) separately for different deciles of the wage distribution in the baseline sample in year t . Team definition 2 at horizon $n = 3$ years.

Table XIX Baseline results for different deciles of the wage distribution. Counterpart to table VI

	(1)	(2)	(3)	(4)	(5)
\hat{w}	0.0058 (0.0063)	0.053*** (0.0041)	0.038*** (0.0021)	0.0053 (0.0041)	0.011*** (0.0021)
$\hat{w} \times \text{Size}$	0.0046*** (0.0013)				
$\hat{w} \times \text{Wage Dispersion}$		-0.072*** (0.012)			
$\hat{w} \times \text{Mean-Median}$			0.074*** (0.022)		
$\hat{w} \times \# \text{ Above}$				0.0049*** (0.00090)	
$\hat{w} \times \% \text{ Above}$					0.025*** (0.0017)
R^2	0.153	0.151	0.151	0.154	0.154
Observations	2465708	2465708	2465708	2440370	2465708

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: $\hat{\beta}$ as estimated from specification (8) with the inclusion of additional interactions. Size is measured as log team size, wage dispersion is measured as the standard deviation of log wages, inverse skewness is measured as the log of the median/mean ratio. The forth column interacts with the log of the number of team-members above worker i in terms of wages, column (5) does so for the fraction. All regressions control for size, wage dispersion and skewness. Team definition 1 at horizon $n = 3$ years.

Table XX Baseline results interacted—Counterpart to table VII

	(1)	(2)	(3)
\hat{w}^+	0.042*** (0.0027)		0.041*** (0.0028)
\hat{w}^-		0.027*** (0.0019)	0.020*** (0.0020)
R^2	0.134	0.124	0.140
Observations	2459982	2483277	2449107

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Column (3) reports the results from specification 9. Columns (2) and (1) only use information about those less well and those better paid, respectively. Team definition 1 at horizon $n = 3$ years.

Table XXI Splitting peers into those better and less well paid—Counterpart to table VIII

	(1)	(2)
\hat{w}^+	0.056*** (.0028)	0.048*** (.012)
\hat{w}^-	0.0160*** (.0028)	0.017 (.011)
Observations	29900	1131

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Column (1) reports the results from specification 9 restricting the sample to establishment movers (within the year following the observation) with an intermittent unemployment spell. Column (2) requires a simultaneous mass-layoff event at the employer. Team definition 1 at horizon $n = 3$ years.

Table XXII Splitting peers into those better and less well paid - movers only. —Counterpart to table [XII](#)