Learning from Coworkers*

Gregor Jarosch

Princeton University

gregorjarosch@gmail.com

Ezra Oberfield

Princeton University

edo@princeton.edu

Esteban Rossi-Hansberg

Princeton University

erossi@princeton.edu

November 8, 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 a sample of establishments. We document that having more highly paid coworkers is strongly associated with future wage growth, particularly if those workers earn more. Motivated by this fact, we propose a dynamic theory of a competitive labor market where firms produce using teams of heterogenous workers that learn from each other. We develop a methodology to structurally estimate knowledge flows using the full-richness of the German employer-employee matched data. 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. In line with our reduced form results, learning from coworkers is significant, particularly from more knowledgeable coworkers, although the amount learned from having marginally better coworkers decreases with knowledge gaps. We show that, on average, at least 4.3% of a worker's compensation is in the form of learning and that inequality in total compensation is significantly lower than inequality in wages.

*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 a large fraction 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.

Our goal is to take advantage of the structure of the model, together with detailed micro data on individual wages in production teams, to study learning on the job. However, before structurally connecting the model with the data, we start by analyzing the reduced-form empirical relationship between the wage growth of an individual and the wages of her coworkers. To measure 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 small effects from less-well-paid workers and large, and 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 matched employer-employee data and the underlying knowledge and learning of individuals. We structurally estimate a variety of parametric versions of the learning function, motivated by the most important reduced form patterns we document. We develop a novel way to estimate the parameters of this function using the micro data for the German labor market. Our methodology uses only information on workers wages and the wages of her coworkers. Given an initial guess of a learning function, we show that we can invert individual Bellman equations to recover the underlying knowledge of individuals from the full set of wages of the members of each production team in a single cross section. Doing this for several years yields a panel of individuals' knowledge. Our identification of the learning function comes from the restriction that the evolution of individual knowledge must be consistent with our initial guess of the learning function. We find the fixed point of this GMM procedure using an iterative algorithm, resulting in 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, although perhaps strong, assumptions like complete financial markets (or linear utility), perfect labor market competition, and no aggregate shocks. However, since the estimation strategy relies solely on the individual Bellman equations together with observed wages and team composition, it allows for a very general on the set of firm technologies, complementarities across workers, and the specification of the learning function. In fact, our methodology proves that these characteristics of the production function are not needed to estimate learning functions. 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 also show that inequality in wages is roughly one quarter larger than inequality in compensation because workers with different levels of knowledge vary in how their compensation is divided between wages and learning. 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 interactions do not happen exclusively 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 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 competitive labor market model allows us to structurally estimate the model and thereby take advantage of the full richness of the matched employer-employee data. Furthermore, because our model features teams with arbitrary numbers of heterogenous workers, our analysis focuses on the role of the within-team distribution of knowledge for the coworker learning 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 coworkers 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.

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

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 $G(z'|z, \tilde{\mathbf{z}})$. Financial markets are complete, or utility is linear, so agents maximize the expected present value of income.

Since individuals learn from coworkers, the wage they are willing to accept depends on how much they might learn from coworkers. Thus, the wage schedule, $w(z, \tilde{\mathbf{z}})$, paid to a worker with knowledge z depends also on the 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 $F(\mathbf{z}; a)$, where $\mathbf{z}$ is the vector of worker it hires. Firms take the wage schedule as given. We purposely impose minimal structure on the production function. In particular, 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, 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

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

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 $\tilde{\mathbf{Z}}$ denote the set of all possible vectors of coworkers. The expected present value of earnings for an agent with knowledge z is given simply by

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

Namely, each period 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 is the case that equilibrium wages adjust so that workers are indifferent about working in a set of firms. The competitive labor market assumption implies that workers with a given z will obtain the same value, $V(z)$, independent of where they work. Hence, the present value of earnings of a worker does not depend on her current coworkers. Furthermore, since firms take the wage schedule as given, it must be the case that if a firm 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

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

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

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

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

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

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,

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

where m denotes the mass of firms in the economy.

Free entry requires that

$$\int_a [\pi(a) - c] \, dA(a) = 0. \tag{6}$$

2.4 The Distribution of Knowledge

Given the choices of firms, we can define $O(\tilde{\mathbf{z}}|z) : \tilde{\mathbf{Z}} \times \mathcal{Z} \to [0,1]$ 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 whose interactions with coworkers leaves them with knowledge weakly less than z. Namely,[3]

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

[3] Note 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.

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 set $\tilde{\mathbf{Z}}$, 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 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 $F(\mathbf{z}_1, a) < F(\mathbf{z}_2, a)$.*

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 $\mathcal{T}V(z) = \max_{\tilde{\mathbf{z}} \in \tilde{\mathbf{Z}}} w(z, \tilde{\mathbf{z}}) + \beta \int_0^\infty V(z') \, dG(z'|z, \tilde{\mathbf{z}})$ 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, $\mathcal{T}V$ 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, implies

$$\begin{aligned}
\mathcal{T}V(z_1) &= w(z_1, \tilde{\mathbf{z}}_1) + \beta \int_0^\infty V(z') \, dG(z'|z_1, \tilde{\mathbf{z}}_1) \\
&< w(z_2, \tilde{\mathbf{z}}_1) + \beta \int_0^\infty V(z') \, dG(z'|z_1, \tilde{\mathbf{z}}_1) \\
&\leq w(z_2, \tilde{\mathbf{z}}_1) + \beta \int_0^\infty V(z') \, dG(z'|z_2, \tilde{\mathbf{z}}_1) \\
&\leq \max_{\tilde{\mathbf{z}} \in \tilde{\mathbf{Z}}} w(z_2, \tilde{\mathbf{z}}) + \beta \int_0^\infty V(z') \, dG(z'|z_2, \tilde{\mathbf{z}}) \\
&= \mathcal{T}V(z_2),
\end{aligned}$$

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

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

Hence, 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 $G\left(z'|z, \tilde{\mathbf{z}}\right)$.

To do so, we relate individual wage dynamics to wages in the peer group using various flexible reduced-form specifications. We are particularly interested in three questions: First, do future wages rise more steeply if one's coworkers are more highly paid? Second, if so, does this relationship depend on *which* coworkers are more highly paid? That is, is it those below in the within-team wage distribution or those above that matter? Third, how do these reduced form patterns change with team size, tenure, age, and the current wage level? We further offer various robustness checks with the particular focus on ruling out two alternatives to learning which could be driving our initial findings, namely a back-loaded wage structure and mean reversion.

While none of the reduced-form specifications are 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. In the next section we use the structure of our theory, and the suggested specification of the learning function, to structurally 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 $G\left(z'|z, \tilde{\mathbf{z}}\right)$ 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 LM 9310)* 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.[4] 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, among others). We organize the resulting dataset as an annual panel. Specifically, the annual observation recorded (employer, average daily wage, etc.) for each individual pertains to the spell which overlaps a particular *reference date* (January 31st).

To construct a baseline sample, we then proceed as follows. We select panel case establishments for the year 2000 to 2008 since those are the establishments where we obtain information on the full workforce. We then include those individuals who were employed at one of those establishments at the reference date 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. The Data Appendix A provides more detailed information on the construction of the baseline sample.

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

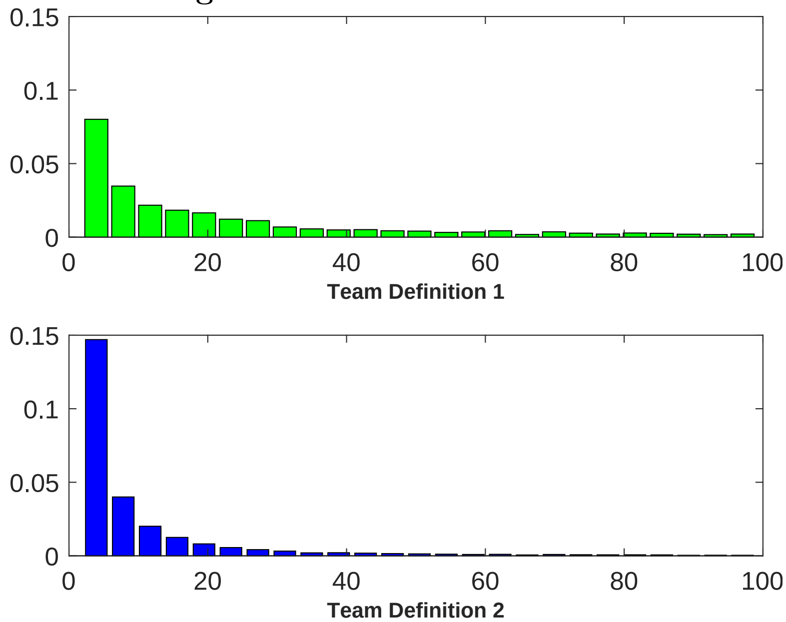
- Team Definition 1: All workers in the same establishment.
- Team Definition 2: All workers in the same establishment and occupation.

We next document the team size distribution along with the wage distribution, both economy-wide, and within teams. We report all those statistics for the year 2000.

[4] For more detail, see Klosterhuber et al. (2014).

Team 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 sample contains 4538 establishments with average size 116. When working with the second team definition we have a total of 28685 teams with an average size of 18.

Figure 1 Team Size Distribution.



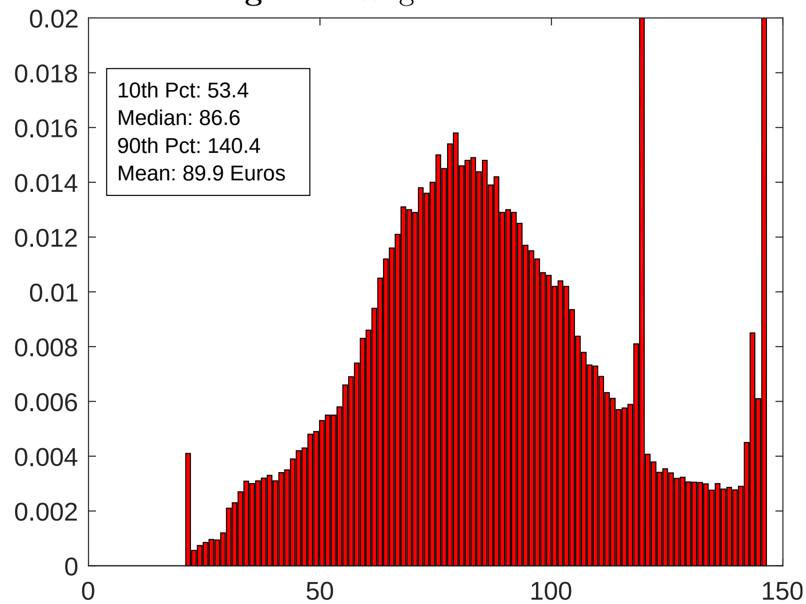
Notes: Top panel plots the unweighted team size distribution in the year 2000 for teams of size 2-99 for team definition 1 (so it corresponds to the establishment size distribution). The bottom panel plots the unweighted team size distribution in the year 2000 for teams of size 2-99 for team definition 2.

Wage Distribution Figure 2 plots the histogram of the average daily earnings during the year-2000 reference spell in Euros. The "mass-points" reflect top-coding of the earnings data at the social security contribution ceiling (which is lower in Eastern Germany). As a consequence, roughly 5% of our wage observations are top coded.[5] A simple variance decomposition implies that the within-team component accounts for 47.5% of the overall

[5]While there exist imputation methods to address the truncation, we instead treat the top-coded observations as actual wage observations and do not correct for the top-coding. We have experimented with various ways of treating the top coded observations and found our main empirical results to be insensitive. We show below how our reduced form results change when omitting all teams with top coded wage observations. For more detail regarding the construction of wages, see Data Appendix A. There, we also describe how we winsorize extreme wage observations which leads to the bunching at the bottom of the wage distribution.

variance in wages under team definition 1 and 22.8% under team definition 2. Finally, there is fairly little wage growth in the decade covered by our dataset. The cohort whose wages are depicted in Figure 2 experienced average annual wage growth of .98% over the next 5 years. The annual growth rate drops to .4% in the second half of the decade.

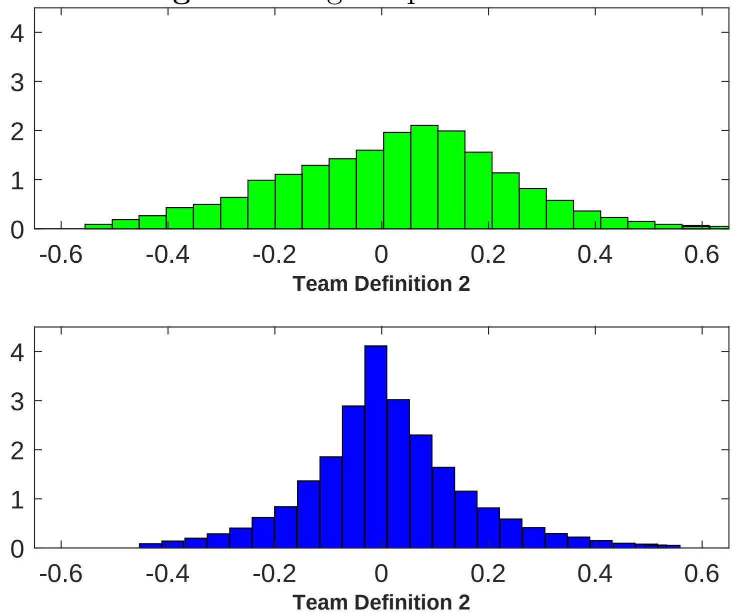
Figure 2 Wage Distribution



Notes: Distribution of mean daily wages during spell overlapping 01/31/2000 for full time employees working subject to social security.

Wage Gap to Coworkers We are interested in how a worker's future wage growth relates to her coworker's (relative) wages. To gauge the extent of wage differences across peers, Figure 3 plots the histogram of wage gaps, defined as the log difference between an individual's wage and the mean wage of her peers, for each team definition. Under the first team definition, the gap has mean .023 and amounts to -.26, .04, and .28 at the 10th, 50th, and 90th percentiles in the year 2000. Under the second team definition, $\hat{w}$ has mean .01 and -.17, .00, and .20 at the 10th, 50th, and 90th percentiles. Naturally, within-team wage dispersion is smaller under the narrower team definition.

Figure 3 Wage Gap Distribution



Notes: Top panel plots the distribution of wage gaps as defined in the main text in the year 2000 for team definition 1. Bottom panel: team definition 2.

3.2 Regression Framework

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

$$w_{i,t+h} = \alpha + \beta \bar{w}_{-i,t} + \gamma w_{i,t} + \omega_{age} + \omega_{tenure} + \omega_{gender} + \omega_{educ} + \omega_{occ} + \omega_t + \varepsilon_{i,t}. \tag{8}$$

$w_{i,t+h}$ is individual i's log wage in year $t+h$, which we project on the log mean wage of her peers in year t, $\bar{w}_{-i,t}$, controlling for her own log wage in year t, $w_{i,t}$ along with fixed effects for age decile, tenure decile, gender, education, occupation, and year. Unless otherwise indicated that is the set of fixed effects used in all specifications. Further, we omit observations that fall into the top and bottom percentile in terms of wage growth from t to $t+h$ in all reduced form specifications.

All our regressions pool the observations across all years t. Since we observe the full peer groups for ten years, the results for $h > 2$ use only a subset of years t. For instance, for $h = 10$ we are restricted to exclusively use information about peers in the year 2000 (since we observe workers until 2010). Likewise, for $h = 5$ we can use all years between 2000 and

	Panel A: Narrow Team Definition				
Horizon in Years	1	2	3	5	10
$\bar{w}$	0.054***	0.077***	0.10***	0.14***	0.18***
	(0.0022)	(0.0030)	(0.0034)	(0.0046)	(0.0094)
Within R^2	0.92	0.86	0.82	0.73	0.54
Observations	3888638	3402968	2925264	2121201	504292
	Panel B: Broad Team Definition				
Horizon in Years	1	2	3	5	10
$\bar{w}$	0.047***	0.070***	0.094***	0.14***	0.19***
	(0.0021)	(0.0028)	(0.0033)	(0.0045)	(0.0098)
Within R^2	0.92	0.86	0.82	0.74	0.54
Observations	4026431	3525811	3034413	2203212	518845

Standard errors in parentheses

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

Notes: $\hat{\beta}$ as estimated from specification (8). Column titles indicate horizon n. Standard errors clustered at the establishment-year level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year.

Table I Estimation Results for Specification (8).

2005.

We report the parameter estimates for each team definition in Table I, clustering standard errors at the establishment-year level. Panel A reports the results using team definition 2. We first note that our findings suggest quantitatively large effects: They imply that doubling the mean wage of individual i's peers raises, in expectation, i's wage next year by over 5%. These effects are naturally larger as the horizon extends further into the future but not linearly. This is natural in the context of learning, as agents likely learn less as they gradually become more knowledgeable. Over a 10 year horizon, doubling peers' wages results in 18% higher wages.

We next contrast these results with the corresponding results for the wider team definition 1. Comparing Panel A and Panel B of Table I, the coefficients tend to be larger for the narrower team definition. These results are consistent with learning from coworkers if interactions between coworkers within occupations are more intense. Thus, in the rest of this section we restrict our attention to team definition 2. We separately report all results

	Horizon in Years				
	1	2	3	5	10
$\bar{w}^+$	0.072***	0.10***	0.13***	0.18***	0.22***
	(0.0042)	(0.0055)	(0.0066)	(0.0089)	(0.020)
$\bar{w}^-$	0.016***	0.025***	0.036***	0.055***	0.089***
	(0.0028)	(0.0036)	(0.0043)	(0.0059)	(0.010)
Within R^2	0.91	0.85	0.81	0.72	0.52
Observations	3442091	3018546	2603447	1892288	446099

Standard errors in parentheses

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

Notes: $\hat{\beta}^+$ and β^- as estimated from specification (9). Team definition 2. Column titles indicate horizon n. Standard errors clustered at the establishment-year level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year.

Table II Estimation Results for Specification (9).

for the alternative team definition in the Appendix.

We next turn to an alternative specification where we split a peer group into those with higher and lower wages. In particular, we let $\bar{w}^+_{-i,t}$ ($\bar{w}^-_{-i,t}$) denote the log of the mean wage of i's peers with higher (lower) wages. We then run the otherwise unaltered specification,

$$w_{i,t+h} = \alpha + \beta^+ \bar{w}^+_{-i,t} + \beta^- \bar{w}^-_{-i,t} + \gamma w_{i,t} + \omega_{age} + \omega_{tenure} + \omega_{gender} + \omega_{educ} + \omega_{occ} + \omega_t + \varepsilon_{i,t} \ , \quad (9)$$

and report our findings in Table II.

The table documents a stark asymmetry. It suggests that the peers higher up in the team wage distribution matter far more for future wage outcomes than the peers below. While increasing the average wage of either group comes with a significant increase in the expected wage of an individual, the peer group above has an impact three to four times larger at all horizons. These findings are consistent with larger knowledge flows from more highly skilled peers and little congestion from less skilled peers. Since we find this stark asymmetry to be robust across all specifications we henceforth build on specification (9).

3.2.1 Across the Wage, Age, Tenure, and Size Distributions

We show next that the forces we document are present across the labor market. In particular, we run the same baseline specification separately for workers in different deciles of the wage distribution.[6] We next repeat the exercise for different deciles of the age, tenure, and team size distribution. As we discuss further below, the findings in this subsection suggest that the patterns we have described thus far are not entirely driven by mean-reversion, a backloaded wage structure, or returns to tenure or experience.

Our first set of results are reported in Panel A of Table III which reports the regression coefficients for specification (9) which we run separately for each decile of the wage distribution.[7]

The results are fairly stable across the wage distribution.[8] 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. Furthermore, since the coefficients do not decline as we focus on higher wage deciles, the results strongly suggests that our baseline findings do not reflect a form of economy-wide mean reversion in wages.

Our next set of results is reported in Panel B of Table III where we cut the sample into different deciles of the (pooled) sample age distribution. Our findings suggest that the effects we document are substantially stronger for young workers. Furthermore, while we find positive and significant effects from more highly paid peers above and below for all segments of the wage distribution, the asymmetry vanishes, and perhaps reverses, at the top.

As a consequence of these strong age patterns, we entertain an age-specific learning function in the structural estimation below. The results we find there starkly mimic these reduced form patterns.

Our next set of results is reported in Panel C of Table III and shows that the patterns we

[6] Of course, we use the full peer group in the construction of the independent variable as before.

[7] Specifically, we assign a worker-year observation to decile i if the worker's wage in that year falls into decile i of the wage distribution during that year.

[8] The sharp increase in the coefficient estimate for the two top deciles is likely a consequence of the top-coding since that group has an artificially compressed distribution of $\bar{w}^+$. 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 h years ahead.

	1	2	3	4	5	6	7	8	9	10
	Panel A: Decile of the Wage Distribution									
$\bar{w}^+$	0.100***	0.11***	0.13***	0.13***	0.13***	0.11***	0.11***	0.10***	0.32***	0.32***
	(0.0055)	(0.0073)	(0.0095)	(0.0099)	(0.011)	(0.013)	(0.015)	(0.016)	(0.015)	(0.045)
$\bar{w}^-$	0.017***	0.025***	0.028***	0.027***	0.030***	0.033***	0.062***	0.055***	0.047***	0.019***
	(0.0045)	(0.0060)	(0.0063)	(0.0080)	(0.0079)	(0.0081)	(0.0084)	(0.0068)	(0.0068)	(0.0032)
Within R^2	0.56	0.11	0.067	0.049	0.045	0.047	0.059	0.92	0.19	0.079
Observations	247633	262248	264113	263406	261165	260231	260130	260297	261434	262708
	Panel B: Decile of the Age Distribution									
	1	2	3	4	5	6	7	8	9	10
$\bar{w}^+$	0.22***	0.20***	0.14***	0.10***	0.094***	0.078***	0.056***	0.038***	0.022**	0.010
	(0.010)	(0.010)	(0.0088)	(0.0078)	(0.0078)	(0.0071)	(0.0065)	(0.0061)	(0.0062)	(0.0057)
$\bar{w}^-$	0.0063	0.029***	0.028***	0.034***	0.039***	0.042***	0.046***	0.053***	0.051***	0.044***
	(0.0059)	(0.0061)	(0.0059)	(0.0059)	(0.0057)	(0.0052)	(0.0050)	(0.0047)	(0.0048)	(0.0043)
Within R^2	0.70	0.76	0.79	0.81	0.82	0.82	0.83	0.85	0.84	0.85
Observations	309309	254804	252464	292890	204510	300823	283837	260617	221482	222664
	Panel C: Decile of the Tenure Distribution									
	1	2	3	4	5	6	7	8	9	10
$\bar{w}^+$	0.20***	0.18***	0.16***	0.15***	0.096***	0.077***	0.059***	0.047***	0.032**	0.011
	(0.010)	(0.011)	(0.015)	(0.0094)	(0.0082)	(0.0093)	(0.0081)	(0.0093)	(0.010)	(0.011)
$\bar{w}^-$	0.0087	0.012*	0.033***	0.047***	0.041***	0.051***	0.053***	0.058***	0.044***	0.049***
	(0.0054)	(0.0050)	(0.0057)	(0.0064)	(0.0057)	(0.0064)	(0.0064)	(0.0071)	(0.0073)	(0.0083)
Within R^2	0.75	0.79	0.81	0.82	0.83	0.82	0.82	0.80	0.76	0.76
Observations	255987	258485	264894	258362	310146	218113	295387	231593	259911	250518
	Panel D: Decile of the Size Distribution									
	1	2	3	4	5	6	7	8	9	10
$\bar{w}^+$	0.051***	0.089***	0.12***	0.12***	0.14***	0.17***	0.16***	0.16***	0.11***	0.086
	(0.0031)	(0.0049)	(0.0068)	(0.0083)	(0.011)	(0.012)	(0.015)	(0.023)	(0.027)	(0.069)
$\bar{w}^-$	0.019***	0.033***	0.042***	0.044***	0.039***	0.022*	-0.0094	-0.058***	-0.039	-0.100*
	(0.0022)	(0.0037)	(0.0048)	(0.0061)	(0.0080)	(0.0098)	(0.011)	(0.015)	(0.023)	(0.039)
Within R^2	0.83	0.81	0.80	0.79	0.77	0.76	0.76	0.72	0.67	0.68
Observations	261571	274913	254324	255309	264897	261547	259216	256569	257512	257584

Standard errors in parentheses

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

Notes: $\hat{\beta}^+$ and β^- as estimated from specification (9) for separate deciles of the wage, age, tenure, and team size distributions. We include observation i in the decile k in t if i falls into the k'th decile of the distribution in year t. Team definition 2 at horizon $h = 3$ years. Standard errors clustered at the establishment-year level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year (whenever possible).

Table III Baseline results for different deciles of the wage, age, tenure, and team size distribution.

describe are present across the (job) tenure distribution. In particular, the table shows that more highly paid coworkers are associated with larger wage growth for workers everywhere in the tenure distribution. In addition, increasing the wages of more highly paid peers has a larger effect on one's future wage growth compared with the effect from less-well-paid peers. The asymmetry is stark at the bottom of the tenure distribution and, similar to the results for age, vanishes towards the top.

Our final set of results in this subsection gauges the role of team size and is reported in Panel D of Table III. It shows that the patterns we describe are present in small and large teams alike. The estimated relationships are somewhat weaker for the smallest teams and become less precise at the very top.[9] These results suggest that learning is, to some extent, independent of team size and happens throughout the team size distribution. As a consequence, when we have to take a stance on parametric forms in the structural estimation, we work with scale-independent learning functions.

3.2.2 Movers

We now consider whether our baseline results may be driven by mean reversion in wages. To do so, we run specification (9) for a sample of workers that leave their establishment after the reference spell. Specifically, we restrict the sample to workers who leave their job after the reference date in year t and regain employment at a different employer by the reference date in year $t + 1$.[10]

The results are reported in Panel A of Table IV. The table corroborates the asymmetry we have found so far using the mover design and shows that more highly paid peers comes with higher wage growth even for those who move to a different establishment.

Clearly, the effects remain large and significant which implies that our results do not purely reflect team level mean-reversion in wages. In fact, the results indicate that the effect of having better coworkers is somewhat larger for agents that switch jobs than for agents that remain in the same job. This can be the result of selection. Switchers might be the

[9] The reason is that we cluster standard errors at the establishment year level and there are very few establishments in the largest team decile.

[10] Recall that we assign the employer pertaining to the spell overlapping January 31th of any given year as the annual observation. We further note that our data do not allow us to observe the firm's other establishments so we cannot rule out that some of the team-switchers move within the same firm.

ones that learned the most, which might give them an incentive to leave the team if there is no room for their newly acquired skills in their current organization. Alternatively, it maybe the result of the fact that switchers tend to be young, and young worker's wage growth is more sensitive to peers' wages.

To account for this we present next results for only those switchers who experience an interim unemployment spell between jobs. That is, we restrict the sample to movers who experience a period of joblessness in year t and report the corresponding results in Panel B of Table IV.

To go even further, we conclude this section by restricting the sample to switchers with an interim spell of nonemployment whose reference spell employer also experienced a mass layoff event in year t.[11] Arguably, constructing the sample this way partially controls for selection and, as can be seen in Panel C of Table IV our corresponding estimates tend to fall somewhat. Further, the results are somewhat more noisy because we lose most observations. Nevertheless, the results remain large and significant which strongly suggests that within-team mean reversion is not a key driver of our baseline results.

3.2.3 Discussion and Robustness

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

We argue, however, that the additional results we have shown thus far are hard to reconcile with those mechanisms. First, the mover results suggest that the forces we document are present and similar in size for team switchers, including those who arguably leave their

[11] We identify such an event at establishments who had at least 50 employers two years prior and have since contracted employment by at least 30%. We further require that they did not build up employment by more than 30% between year $t-3$ and $t-2$ and that in year $t+1$ employment remains at least 10% below its year $t-2$ level.

	Panel A: All Switchers				
Horizon in Years	1	2	3	5	10
$\bar{w}^+$	0.083***	0.12***	0.14***	0.18***	0.27***
	(0.0091)	(0.011)	(0.012)	(0.012)	(0.023)
$\bar{w}^-$	-0.029***	-0.026***	-0.018**	-0.0044	0.017
	(0.0058)	(0.0057)	(0.0069)	(0.0086)	(0.016)
Within R^2	0.77	0.70	0.64	0.53	0.35
Observations	180183	217140	194975	154877	42717

	Panel B: Switchers with Unemployment Spell				
Horizon in Years	1	2	3	5	10
$\bar{w}^+$	0.050***	0.10***	0.12***	0.16***	0.22***
	(0.012)	(0.0092)	(0.010)	(0.014)	(0.029)
$\bar{w}^-$	-0.032**	-0.0062	0.0059	0.0090	0.021
	(0.011)	(0.0074)	(0.0088)	(0.011)	(0.023)
Within R^2	0.67	0.67	0.59	0.47	0.29
Observations	16920	66251	63576	53939	15771

	Panel C: Switchers, Mass Layoff Event				
Horizon in Years	1	2	3	5	10
$\bar{w}^+$	0.044	0.062*	0.077*	0.14**	0.35***
	(0.039)	(0.030)	(0.038)	(0.042)	(0.095)
$\bar{w}^-$	0.014	-0.029	-0.042	0.0048	-0.15
	(0.043)	(0.027)	(0.033)	(0.038)	(0.083)
Within R^2	0.64	0.58	0.47	0.36	0.21
Observations	1860	4674	4936	4524	1533

Standard errors in parentheses

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

Notes: $\hat{\beta}^+$ and β^- as estimated from specification (9) on a sample of establishment switchers. Team Definition 2. Column titles indicate horizon n. Standard errors clustered at the establishment-year level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year.

Table IV Establishment switchers.

establishments involuntarily. This suggests back-loading or establishment specific factors cannot be the only force behind our baseline results. This is further corroborated by the results which show that our baseline effects are present across the tenure and age distribution. Second, the sharp asymmetry between $\hat{\beta}^{+}$ and $\hat{\beta}^{-}$, for which we provide additional evidence in the next subsection, is hard to reconcile with an explanation which broadly builds on within-team mean reversion in wages. Third, the asymmetry, jointly with the results for movers and job tenure rule out that the basic results are exclusively driven by learning-by-doing or on-the-job search as in Postel-Vinay and Robin (2002)

Finally, we point to additional robustness results which are relegated to Appendix B.2. There, we show that the patterns documented in this section are robust to modified sample selection criteria and reduced-form specifications. We show results when we omit teams with top-coded wages, when we omit teams with apprentices, when we exclusively work with establishments that neither have a collective bargaining agreement nor benchmark their wages with one.

As can be seen from the corresponding tables in Appendix B.2, our baseline results appear mostly insensitive to these modifications and the basic pattern highlighted throughout this section remains stable.

We conclude this section by offering the results for a specification which splits individuals working in the same establishment as worker i into two groups: Those working in the same occupation and those working in different occupations. We again split each of those two groups into those paid more than w_i and those paid less. We include those four variables on an otherwise unchanged specification (9) and report the results in Table V.

The results indicate that individuals learn more from higher-wage peers in the same occupation than in other occupations. The asymmetry between $\hat{\beta}^{+}$ and $\hat{\beta}^{-}$ is also much larger in the same than in other occupations. These results are natural if we interpret them as resulting from learning. In their own occupation, individuals learn mostly from more knowledgeable peers. In contrast, when they interact with peers in occupations that use different knowledge, they learn from everyone since they know less of the topic themselves.

	Horizon in Years				
	1	2	3	5	10
$\bar{w}^{+}_{same\ occ}$	0.054***	0.072***	0.092***	0.13***	0.15***
	(0.0040)	(0.0052)	(0.0061)	(0.0081)	(0.019)
$\bar{w}^{-}_{same\ occ}$	0.0070**	0.011**	0.016***	0.025***	0.051***
	(0.0026)	(0.0035)	(0.0040)	(0.0054)	(0.0089)
$\bar{w}^{+}_{other\ occ}$	0.040***	0.059***	0.074***	0.11***	0.13***
	(0.0040)	(0.0053)	(0.0062)	(0.0085)	(0.021)
$\bar{w}^{-}_{other\ occ}$	0.019***	0.030***	0.044***	0.064***	0.090***
	(0.0027)	(0.0036)	(0.0046)	(0.0063)	(0.014)
Within R^2	0.91	0.85	0.80	0.71	0.52
Observations	3299076	2894765	2498943	1818950	429837

Standard errors in parentheses

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

Notes: $\hat{\beta}^{+}$ and β^{-} as estimated from specification (9) with separate coefficients for peers in the same occupation and in other occupations. Column titles indicate horizon n. Standard errors clustered at the establishment-year level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year.

Table V Peers in the same and in other occupations.

3.2.4 A More Flexible Specification

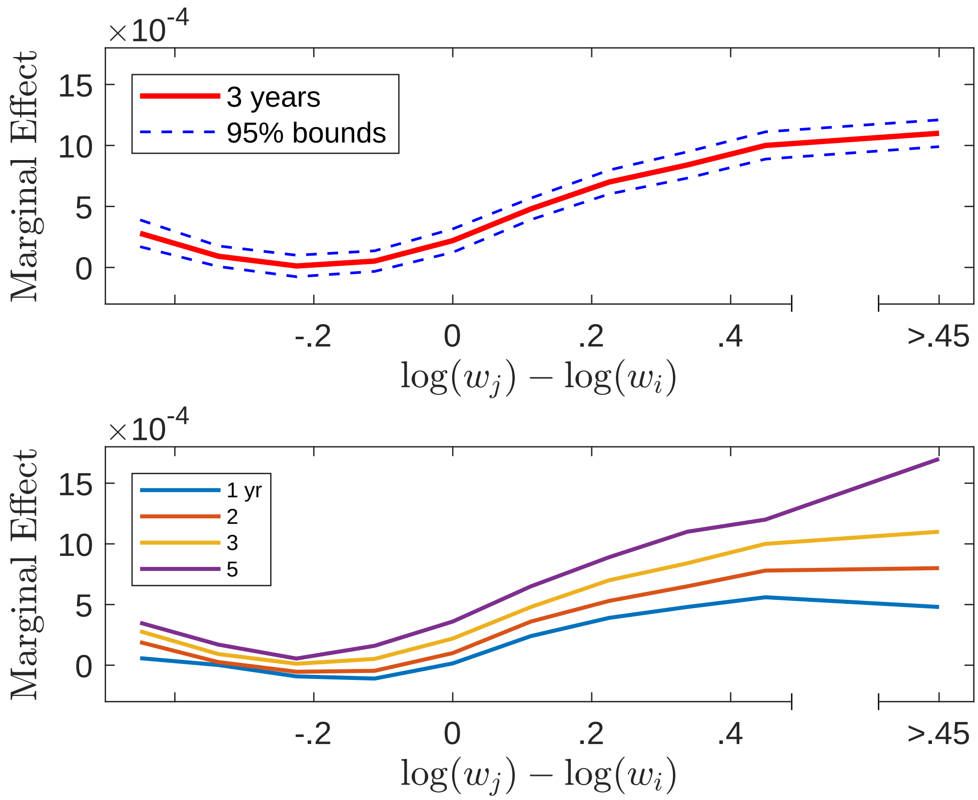
We end the reduced-form exploration with an exercise that attempts to approximate the wage distribution surrounding a worker in a flexible way. To do so, we divide a worker's peers into 11 bins. The bottom bin takes peers j with wage such that $\log(w_j) - \log(w_i) < -.45$ while the top bin takes those peers with $\log(w_j) - \log(w_i) > .45$. All other workers are grouped into 9 equally spaced bins in-between. We then compute, for each individual i and year t, the fraction of her coworkers in each bin k, $p_{i,k,t}$, and run the following regression,

$$w_{i,t+h} = \sum_{k=2}^{11} \beta_k p_{i,k,t} + \gamma w_{i,t} + \omega_{age} + \omega_{tenure} + \omega_{gender} + \omega_{educ} + \omega_{occ} + \omega_t + \varepsilon_{i,t}. \tag{10}$$

That is, we project log wages h years ahead on the current log wage and a non-parametric approximation of the current peer wage distribution around a worker along with our standard controls and fixed effects. We present the results in Figure 4 and report the underlying table in the appendix section B.1.

Figure 4 plots the marginal response of log wages h years ahead to increasing the weight on each of the 10 bins (relative to increasing the weight on bin 1 which is the omitted category). The figure shows that moving 10% of one's peers from the bottom bin into the highest bin increases wages 3 years ahead by slightly more than 1%. It also shows that the effects are naturally larger for longer horizons, but exhibit similar patterns. The figure confirms the findings from the previous exercises: Those who are less well paid (those in bins 5 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 structural estimation below 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), suggesting that individuals learn more from coworkers that are further out in the wage distribution.

The bottom panel of Figure 4 also confirms that learning from higher earners is accumulates over time, presumably because the composition of teams is highly persistent, but not in a linear fashion. We report the table underlying Figure 4 in Appendix B.1 along with the corresponding results for the other team definition. In addition, we report the results, for the horizon $h = 3$, for specifications restricted to workers above (or below) the median wage in their team and to specifications restricted to workers selected from particular deciles of the wage distribution. The basic patterns in Figure 4 are generally confirmed.



Notes: We plot the coefficients $\hat{\beta}_k$ from regression specification (10) with weights p_k scaled such that they add up to 100. The bin $k = 1$ (which takes the peers i such that $\log(w_j) - \log(w_i) < -.45$) is the ommitted category. The top panel only plots $h = 3$ along with the 95% confidence bands. Standard errors clustered at the establishment year level. The bottom panel plots estimates for different horizons. All workers with $\log(w_j) - \log(w_i) > .45$ are in one single bin as indicated by break in the axis. The figure uses Team Definition 2. Standard errors clustered at the establishment-year level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year.

Figure 4 Approximating the Wage Distribution

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.

Heuristically, our method exploits two dimensions of the data. First, it uses the within-team coworker pay differentials we observe in repeated cross-sections to back out the worker types z which are consistent with a particular $G(.)$. That is, compensating differentials reveal z given a learning function $G(.)$. Second, it uses the intertemporal dimension of the resulting panel dataset of worker types to estimate $G(.)$. That is, the dynamic evolution of z_i projected on z_{-i} identifies the learning function $G(.)$.

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,

$$V(z) = w(z, \tilde{z}) + \beta E\left[V(z') \,|z, \tilde{\mathbf{z}}\right] \tag{11}$$

which is the result of the worker's maximization. Equation (11) depends on the assumptions of stationarity, perpetual youth, competitive labor markets, and complete financial markets (or linear utility).[12] 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

[12]Our 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 (11) with a known increasing and concave function of the wage.

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 cardinality of z so that $V(z) = z$. Then, (11) becomes

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

or

$$\begin{aligned} z_i &= w_i + \beta E\left[z_i' | z_i, \tilde{\mathbf{z}}_{-i}\right] \\ &= w_i + \beta \int_0^\infty z_i' dG\left(z_i' | z_i, \tilde{\mathbf{z}}_{-i}\right). \end{aligned} \tag{12}$$

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 (12) 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, (12) 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,

$$z_i' = \mathcal{E}\left(z_i, \tilde{\mathbf{z}}_{-i}\right) + \varepsilon_i, \tag{13}$$

where $\mathcal{E}\left(z_i, \tilde{\mathbf{z}}_{-i}\right)$ is the conditional expectation and ε_i is the expectational error. We then use the moment conditions built from $E\left[\varepsilon_i | z_i, \tilde{\mathbf{z}}_{-i}\right] = 0$. Below we specialize to the case where $\mathcal{E}\left(z_i, \tilde{\mathbf{z}}_{-i}\right) = E\left[z_i' | z_i, \tilde{\mathbf{z}}_{-i}\right]$ is a linear combination of several moments of $\{m_k\left(z_i, \tilde{\mathbf{z}}_{-i}\right)\}_{k=1}^K$, so that $\mathcal{E}\left(z_i, \tilde{\mathbf{z}}_{-i}\right) = \sum_{k=1}^K \theta_k m_k\left(z_i, \tilde{\mathbf{z}}_{-i}\right)$. In such a case, we would have K parameters $\{\theta_k\}$ and

K natural moment conditions

$$E\left[m_k\left(z_i, \tilde{\mathbf{z}}_{-i}\right)\varepsilon_i\right] = 0, \quad k = 1, ..., K. \tag{14}$$

Formally, if a team has n workers, then given θ and w, (12) 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 $Z\left(w_t, r_t, \theta\right)$ to be the $I \times 1$ vector of all workers' knowledge at t. Given this, we can construct $M\left(w_t, r_t, \theta\right)$ to be the $I \times k$ matrix of moments so that the i, k entry of $M\left(w_t, r_t, \theta\right)$ is $m_k\left(z_i, \tilde{\mathbf{z}}_{-i}\right)$ 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 (14) can be stacked as

$$E\left[M\left(w_t, r_t, \theta\right)^T\left(Z\left(w_{t+1}, r_{t+1}, \theta\right) - M\left(w_t, r_t, \theta\right)\theta\right)\right] = 0. \tag{15}$$

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 (13). We first guess parameters θ^{guess}. Given this guess, we can back out the types z in a team solely from information on wages.[13] In other words, we invert (12) to solve for all workers' knowledge, $Z\left(w_t, r_t, \theta\right)$. We do this by finding a fixed point $\mathbf{z}$ of the operator

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

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

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

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

[13] As 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 (15). In practice we use $\hat{\theta}$ to update our guess and iterate until we find a fixed point.

A proof of identification then amounts to guaranteeing that this procedure has a unique fixed point. While we currently do not have such a proof, 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 $G(\cdot)$,

$$E\left[z_i'|z_i,\tilde{\mathbf{z}}_{-i}\right] = \int_0^\infty z_i' dG\left(z_i'|z_i,\tilde{\mathbf{z}}_{-i};\theta\right) = \frac{1}{n-1}\sum_{j\neq i} z_i \Theta\left(\frac{z_j}{z_i}\right) \tag{16}$$

where n is the worker's team size and $\Theta\left(\cdot\right)$ is weakly increasing functions. Below we let $\Theta\left(\cdot\right)$ 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

$$\Theta\left(x\right) = \begin{cases} 1+\theta^0+\theta^+\left(x-1\right), & x\geq 1 \\ 1+\theta^0+\theta^-\left(x-1\right), & x<1 \end{cases},$$

or

$$E\left[z_i'-z_i|z_i,\tilde{\mathbf{z}}_{-i}\right] = \theta^0 z_i + \frac{1}{n-1}\left\{\theta^- \sum_{z_j<z_i}\left(z_j-z_i\right) + \theta^+ \sum_{z_j\geq z_i}\left(z_j-z_i\right)\right\}. \tag{17}$$

This learning function allows for asymmetric learning from types z_j for a worker z_i depending on whether $z_j > z_i$ or vice versa. It also allows for a constant time trend in skill growth, θ^0. It is also scale-invariant (apart from the constant) since we divide the second term by $n-1$.

	Team Definition	
	1	2
θ^+	.0772	.0968
	(.0005)	(.0007)
θ^-	.0112	.0349
	(.0003)	(.0005)
θ^0	.0006	.0029
	(.00005)	(.00004)
Observations	3355921	3229902

GMM standard errors in parentheses

Table VI Parametric estimation results for the learning function (17).

In updating our guess for $\theta = \{\theta^+, \theta^-, \theta^0\}$, 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 purely from the cross-sectional 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 VI.[14]

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 \$100 raises that workers expected present value of earnings by \$8-10 times the share of more-highly-paid workers 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 \$1-3 times the share of less-well-paid workers. This implies that there are only very small learning effects coming from these workers, and there is no evidence of congestion. Furthermore, the results imply that learning for individuals at the bottom of the distribution of knowledge in a team is large relative to learning for individuals at the top of the distribution. Naturally, we find somewhat larger effects for the narrower team definition.[15] Clearly, these point

[14] The 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.

[15] One important observation across all specifications we have worked with is that θ^0 is substantially larger

estimates are very much consistent with the reduced form patterns discussed in the previous section.

Finally, while θ^0 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 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

$$\Theta'(x) = \begin{cases} \theta^{++}, & x \geq 1+b \\ \theta^{+}, & 1 \leq x < 1+b \\ \theta^{-}, & x < 1 \end{cases} .$$

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

$$E\left[z_i' - z_i | z_i, \tilde{\mathbf{z}}_{-i}\right] = \theta^0 z_i + \frac{1}{n-1} \left\{ \sum_{z_j < z_i} \theta^- \left(z_j - z_i\right) + \sum_{z_j > z_i} \left[\theta^+ \left(z_j - z_i\right) + \mathbf{1}_{z_j > (1+b) z_i} \left(\theta^{++} - \theta^+\right) \left(z_j - z_i - b z_i\right)\right] \right\} \tag{18}$$

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 $\{\theta^0, \theta^-, \theta^+, \theta^{++}\}$.[16] When implementing this learning function, we set $b = 10\%$.

The results are reported in Table VII. Our estimates for θ^0 and θ^- are hardly changed by the modification of the learning function for either team definition. That is, as before, changing the wages of those team members with lower type affects an individual's expected

for team definition 2.

[16] In 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.

	Team Definition	
	1	2
θ^{+}	.1207	.1148
	(.0016)	(.0010)
θ^{++}	.0765	.0905
	(.0005)	(.0008)
θ^{-}	.0087	.0370
	(.0003)	(.0006)
θ^{0}	.0001	.0038
	(.00005)	(.00004)
Observations	3355921	3198654

GMM standard errors in parentheses.

Table VII Estimates for the learning function (18).

wage growth little in comparison with those team members with more knowledge. Likewise, the estimated trend growth remains minimal. Our results indicate that $\theta^{+} > \theta^{++}$, 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 in closer proximity in the distribution of knowledge. Just like in Table VI, we find that the effects are stronger for the narrower team definition. Workers appear to benefit more from those coworkers that work in the same occupation. Similarly, improving those below in the skill distribution has far larger positive effects when they also work in the same occupation.

We conclude this subsection with three short exercises which cast light on the quantitative importance of coworker learning and its interplay with how teams are formed. We run all three exercises in the context of the piece-wise linear learning function in (16) and team definition 2.

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, $\beta\left(z_i' - z\right)$, where next period's knowledge z_i' is given by equation (13) and compute its simple pooled average across all individuals and years in our sample. We then subtract the pure trend component $\beta\theta^0 z$ 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 95.7% given by the wage. This is the total value of knowledge flows each worker receives *relative to the knowledge flows she would attain from working on a team of identical workers*. In other words, agents invest on average 4.3% of their total compensation in learning from others at work.[17]

Naturally, there is substantial heterogeneity in this breakdown across the knowledge distribution. Knowledge flows amount to 7.9% 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 -1%, 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 equilibrium sorting affect the value of learning 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 $\mathcal{E}\left(z_i, \tilde{\mathbf{z}}_{-i}^{\text{cf}}\right)$ 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, $\mathcal{E}\left(z_i, \tilde{\mathbf{z}}_{-i}\right)$. Doing so reveals that under random

[17] If individuals do, in fact, learn from coworkers with the same knowledge, then some of the trend component would also represent knowledge flows. We note that when we do not subtract the trend component this number rises to 12.4%. The effect of the trend component $\beta\theta^0 z$ is large since average knowledge levels are high relative to the amount of dispersion in knowledge within teams.

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. Intuitively, since the learning function is increasing and convex over much of its domain, learning benefits from large differences between coworkers. In contrast, production benefits from small differences between team members due to knowledge complementarity in the production function. In sum, 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 Inequality

Two individuals with the same knowledge and same present value of earnings might earn different wages because they work on teams with different opportunities to learn. In other words some of wage differences reflect compensating differentials for learning rather than unequal compensation.

We now ask how variation in log wages compares to variation in log compensation, where compensation is measured as $(1-\beta)z_i = w_i + \beta\mathbb{E}[z_i' - z_i|z_i, \tilde{\mathbf{z}}_{-i}]$. The variance of log wages in our data is .138, while the variance in log compensation is .106, about 23% smaller.

4.2.4 An Alternative Interpretation

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

Suppose that z, instead of being a scalar, represents a vector of individual 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 depends 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, since values provide a relevant summary of the vector of characteristics z. Then, 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 (11). 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 the 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 the characteristics to be 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. Denote the joint Markov process by

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

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

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

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

$$v_{it} = w_{it} + \beta \int \int V\left(z', x'\right) G\left(dz', dx' | z_{it}, x_{it}, \{\tilde{\mathbf{z}}_{it}, \tilde{\mathbf{x}}_{it}\}\right), \tag{19}$$

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 $Z\left(v, x\right)$ that satisfies $v = V\left(Z\left(v, x\right), x\right)$ and $z = Z\left(V\left(z, x\right), x\right)$. Further, define the learning function in the value space as

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

where $\tilde{\mathbf{Z}}\left(\tilde{\mathbf{v}}; \tilde{\mathbf{x}}\right)$ is the vector of coworkers' knowledge given their values and characteristics. With this, we can write (19) as

$$v_{it} = w_{it} + \beta \int \int v' \hat{G}\left(dv', dx'|v, x, \tilde{\mathbf{v}}, \tilde{\mathbf{x}}\right) \tag{20}$$

5.1 Algorithm

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

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

so that the realized Bellman equation (20) can be written as

$$v_{it} = w_{it} + \beta \mathcal{E}\left(v_{it}, x_{it}, \tilde{\mathbf{v}}_{-it}, \tilde{\mathbf{x}}_{-it}\right). \tag{21}$$

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 (21) 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 $\{m_\ell\left(v, x, \tilde{\mathbf{v}}, \tilde{\mathbf{x}}\right)\}_{\ell=1}^{n}$

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

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

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

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 then 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 (21). 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 (21) 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 (14) can be stacked as

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

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

$$\begin{aligned} E_y[v_i' - v_i|v_i, \tilde{\mathbf{v}}_{-\mathbf{i}}] &= \theta_y^0 v_i + \theta_{yy}^+ \frac{1}{n-1} \sum_{v_j > v_i, j \text{ young}} (v_j - v_i) + \theta_{yo}^+ \frac{1}{n-1} \sum_{v_j > v_i, j \text{ old}} (v_j - v_i) \qquad (22) \\ E_o[v_i' - v_i|v_i, \tilde{\mathbf{v}}_{-\mathbf{i}}] &= \theta_o^0 v_i + \theta_{oy}^+ \frac{1}{n-1} \sum_{v_j > v_i, j \text{ young}} (v_j - v_i) + \theta_{oo}^+ \frac{1}{n-1} \sum_{v_j > v_i, j \text{ old}} (v_j - v_i). \end{aligned}$$

This specification allows knowledge flows to depend on both the age group of the worker and the age group of her coworker. For instance, θ_{yo}^+ captures the strength of the knowledge flows from old to young coworkers. Furthermore, the specification allows for age group specific trend growth. For simplicity we do 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

	Team Definition	
	1	2
trend growth young: θ_y^0	.0097	.0110
	(.0003)	(.0003)
learning of young from young: θ_{yy}^+	.2961	.1801
	(.0115)	(.0126)
learning of young from old: θ_{yo}^+	.3730	.1494
	(.0355)	(.0200)
trend growth old: θ_o^0	-.0022	.0040
	(.0012)	(.0005)
learning of old from young: θ_{oy}^+	-.0014	.0437
	(.0016)	(.0030)
learning of old from old: θ_{oo}^+	.0546	.0867
	(.0073)	(.0071)
Observations	459427	423118

GMM standard errors in parentheses.

Notes: Old: 40 and older in year 2000.

Table VIII Estimates for the learning function (22).

the same routine outlined in the previous section.

We present our results in Table VIII. In line with our previous findings we find little trend growth for either age group. For the first team definition we find that both the young and the old learn more from the old. In fact, the old only seem to learn from the old, with insignificant trend growth or learning from the young. For the second, narrower team definition we 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. Regardless of the team definition, the young learn far more than the old; 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 nor hinder their learning. We also find that learning from more knowledgeable workers is increasing in the coworker's knowledge. That is, workers learn more if they work with more knowledgeable teammates, but learning per unit of their coworkers knowledge seems to be decreasing.

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 Miguelez, 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.

Davis, Steven J. and Till von Wachter, "Recessions and the Costs of Job Loss," *Brookings Papers on Economic Activity*, 2011, *43* (2), 1–72.

Eberle, Johanna, Alexandra Schmucker, and Stefan Seth, "Example programs for data preparation of the Sample of Integrated Labour Market Biographies for Stata - Creating cross-sectional data and biographical variables," Technical Report 2013.

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.

Klosterhuber, Wolfram, Joerg Heining, and Stefan Seth, "Linked-employer-employee-data from the IAB: LIAB longitudinal model 1993-2010 (LIAB LM 9310)," FDZ Datenreport. Documentation on Labour Market Data 201308_en, Institut fuer Arbeitsmarkt- und Berufsforschung (IAB) February 2014.

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 Data Appendix

Construction of Basic Annual Panel

The basic dataset comes in spell format where a spell can correspond either to employment or to a period of benefit receipt. We implement the publicly available code by Eberle, Schmucker and Seth (2013) to convert the spells into monthly cross sections which we then merge into a monthly panel covering 1993-2010.[18] For each spell that runs through an entire calendar year we see one observation per variable (occupation, employment status, average hourly wage, type of benefit receipt,...) per year. For all other spells, we see one observation per variable per spell.

Our main analysis is carried out on an annual panel and we select the spell overlapping January 31st of a given year as the observation for the year. This implies that the peer groups we study are the full workforce of the sample establishments on January 31st of each year from 2000 to 2008. To construct real prices we deflate using a CPI provided by the data provider.

When assigning a wage observation we assign the daily wage during spells of full time employment unless otherwise noted.[19] As a consequence, we ignore information on earnings from part time employment and construct peer groups only from full time employees for full time employees. The

[18] Download link accessed under http://doku.iab.de/fdz/reporte/2013/MR_04-13_EN.pdf.

[19] We follow the routine in the aforementioned code to select a main employment spell in case individuals hold several jobs.

reason is that, while we observe a part time flag, we do not have good information on hours which blurs the mapping between daily earnings—as reported to the social security administration—and the wage for non full time employees.

Construction of Regression Sample for Section 3

Wages For the reduced form work reported in section 3, we winsorize the annual wage observations at the bottom percentile during any calendar year and do the same at the top .1 percentile.[20] We flag observations which are top-coded due to the social security ceiling when they fall into one of the two masspoints which are easily identified in the wage distribution in a given year. We further omit any observations from the regressions in section 3 where wage growth over the corresponding horizon n falls into the top or bottom percentile of the pooled sample.

Team Construction We identify all workers that work at one of the sample establishments and construct teams with minimum size 2 as the collection of workers employed full time subject to social security during the reference spell for a given calendar year. We exclude workers in vocational training and interns. Thus, our reduced form exploration projects log wages for individual i working full time subject to social security in year $t+n$ on the wages of her full-time coworkers in year t if she worked in one of our sample establishments in year t.

Mass Layoffs and Job Loss To identify a mass layoff event at an establishment, we use information from the IAB establishment panel, which is the annual survey from which the panel cases in our dataset are sampled. In particular, we identify a mass layoff event if the following is true: The establishment reduces full time employment by at least 25% since two years prior, still has a strictly positive number of full time employees, had more than 25 employees two years prior, did not build up employment by more than 30% between three and two years prior, does not rebuild to more than 90% of employment two years prior within the next year, and was surveyed each years from three years prior to one year past.[21]

We register a job loss for individual i in year t if there is at least one instance where she is employed subject to social security at the end of a month but not anymore at the end of the following month.[22] We register a job loss in the context of a mass layoff event in year t if we register a job loss at the individual level during year t and a mass layoff event at her ascribed establishment, that is the one she works at during the spell overlapping January 31st of that year, during the same year t.

Construction of Estimation Sample for Section 4

To construct the sample for the structural estimations in section 4, we build on the same annual worker panel underlying our reduced form work.

[20] The reason for the asymmetry is that almost no wage observations lie above the social security ceiling. A small number of observations have a wage that is above the ceiling because the ceiling applies to annual earnings, and if a worker works for part of the year, her average wage while employed will be above the ceiling.

[21] These criteria closely follow Davis and von Wachter (2011).

[22] We do so when we actually have information on that worker for the following month, that is when we see her receiving benefits. If we do not have any information for the subsequent month we only register job loss if the worker disappears for more than two months from the dataset.

We drop the bottom percentile and the top .1 percentile of wage observations and, since our structural approach requires information on wages, we only use information on full time workers who work subject to social security. As in the reduced form work, we assign the workforce on January 31 during any year to the establishment in that year. Furthermore, we only keep establishments which were surveyed every year between 2000 and 2010.[23]

Correlations We compute a set of correlations of various wage moments at the team level. Specifically, Table IX 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. [24]

	Mean Wage	SD Wage	Mean/median	Team size	Max wage
Mean Wage	1				
Wage sd	0.54	1			
Mean/median	−0.04	0.20	1		
Team size	0.48	0.40	0.21	1	
Max wage	0.84	0.72	0.26	0.67	1

Table IX Pairwise correlations at the team level, team definition 2. All variables *in logs*

B Additional reduced form empirical results

B.1 Figure 4

This section offers the regression output underlying Figure 4. It then offers the same table for the alternative team definition followed by the results for various restricted samples at horizon $n = 3$.

[23] This is clearly not necessary. We constructed the sample in this way so as to be able to supplement the analysis with firm level information from the BHP establishment survey.

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

	Horizon in Years				
	1	2	3	5	10
Bin 2	0.000058	0.00019***	0.00028***	0.00035***	0.00091***
	(0.000032)	(0.000046)	(0.000056)	(0.000074)	(0.00018)
Bin 3	0.0000019	0.000026	0.000092*	0.00017**	0.00075***
	(0.000027)	(0.000036)	(0.000043)	(0.000065)	(0.00015)
Bin 4	-0.000092**	-0.000054	0.000012	0.000055	0.00048***
	(0.000029)	(0.000040)	(0.000045)	(0.000063)	(0.00014)
Bin 5	-0.00011***	-0.000046	0.000052	0.00016*	0.00078***
	(0.000029)	(0.000038)	(0.000043)	(0.000064)	(0.00015)
Bin 6	0.000015	0.00010*	0.00022***	0.00036***	0.00082***
	(0.000029)	(0.000041)	(0.000049)	(0.000069)	(0.00013)
Bin 7	0.00024***	0.00036***	0.00048***	0.00065***	0.0012***
	(0.000032)	(0.000040)	(0.000045)	(0.000065)	(0.00015)
Bin 8	0.00039***	0.00053***	0.00070***	0.00089***	0.0015***
	(0.000035)	(0.000046)	(0.000050)	(0.000071)	(0.00016)
Bin 9	0.00048***	0.00065***	0.00084***	0.0011***	0.0017***
	(0.000038)	(0.000050)	(0.000055)	(0.000075)	(0.00017)
Bin 10	0.00056***	0.00078***	0.0010***	0.0012***	0.0017***
	(0.000040)	(0.000051)	(0.000057)	(0.000082)	(0.00017)
Bin 11	0.00048***	0.00080***	0.0011***	0.0017***	0.0026***
	(0.000037)	(0.000049)	(0.000056)	(0.000083)	(0.00017)
Within R^2	0.91	0.85	0.81	0.72	0.52
Observations	3336570	2928665	2525353	1837294	434527

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 2. Column titles indicate horizon n. Standard errors clustered at the establishment-year level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year.

Table X Results for figure 4 using specification (10).

	All	Above Team-Median	Below Team-Median	2nd Pct.	4th Pct.	7th Pct.	9th Pct.
p_2	0.00028***	0.00051	0.00023***	0.00036	-0.00025	0.0011***	0.00011
	(0.000056)	(0.00030)	(0.000051)	(0.00025)	(0.00028)	(0.00016)	(0.000073)
p_3	0.000092*	0.00036	0.000084*	0.00034	-0.00028	0.00054***	-0.000033
	(0.000043)	(0.00028)	(0.000040)	(0.00022)	(0.00023)	(0.00011)	(0.000066)
p_4	0.000012	0.00033	0.000022	0.00014	-0.000064	0.00041***	-0.00017**
	(0.000045)	(0.00025)	(0.000042)	(0.00019)	(0.00020)	(0.00011)	(0.000059)
p_5	0.000052	0.0000081	0.00013**	0.00021	-0.000015	0.00042***	-0.000032
	(0.000043)	(0.00024)	(0.000041)	(0.00017)	(0.00020)	(0.00011)	(0.000062)
p_6	0.00022***	0.00035	0.00034***	0.00034*	0.00016	0.00046***	0.00020***
	(0.000049)	(0.00024)	(0.000042)	(0.00017)	(0.00018)	(0.00011)	(0.000056)
p_7	0.00048***	0.00065**	0.00050***	0.00056**	0.00027	0.00094***	0.00048***
	(0.000045)	(0.00023)	(0.000053)	(0.00017)	(0.00021)	(0.00012)	(0.000054)
p_8	0.00070***	0.00088***	0.00071***	0.00093***	0.00056**	0.00091***	0.00070***
	(0.000050)	(0.00024)	(0.000078)	(0.00017)	(0.00020)	(0.00012)	(0.000061)
p_9	0.00084***	0.0011***	0.00041***	0.00100***	0.00049*	0.00099***	0.00096***
	(0.000055)	(0.00024)	(0.000087)	(0.00018)	(0.00020)	(0.00012)	(0.000090)
p_10	0.0010***	0.0012***	0.00038***	0.0016***	0.00040	0.0013***	0.0025*
	(0.000057)	(0.00024)	(0.00011)	(0.00020)	(0.00024)	(0.00013)	(0.00097)
p_11	0.0011***	0.0014***	0.00011	0.0017***	0.0012***	0.0016***	0
	(0.000056)	(0.00024)	(0.00011)	(0.00017)	(0.00020)	(0.00012)	(.)
Within R^2	0.81	0.78	0.81	0.22	0.095	0.096	0.18
Observations	2525353	1296730	1228620	102577	256035	338031	315113

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. Columns 2 and 3 report the results when the sample is restricted to workers above (below) the team median wage. The remaining columns restrict the sample to workers from particular parts of the wage distribution. Team definition 2. Column titles indicate horizon n. Standard errors clustered at the establishment-year level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year.

Table XI Results from specification (10) under team definition 2 for various restricted samples.

B.2 Robustness

This sections evaluates the robustness of the main reduced form empirical results reported in section 3.2. We do so for team definition 2 at the horizon $n = 3$ years and report the corresponding tables for team definition 1 below.

To do so, we begin by contrasting our baseline results for specification (9) when omitting teams that have any apprentices.[25] We then restrict the sample exclusively to teams without any top-coded wage observations.[26] In a third exercise we restrict the sample exclusively to men. We report the corresponding results, contrasted with our benchmark results, for $n = 3$ in table XII and, for

[25] We highlight that even our baseline results do not use any wage information on workers in apprenticeship.

[26] The ceiling varies from year to year and differs between former Eastern and Western Germany. Furthermore the data display a certain amount of bunching in a small interval around the officially reported ceiling levels. To identify workers with top-coded wages we thus simply group workers into 50 Euro-cent wide bins in each year and flag the two bins with the most mass.

	Baseline	Teams w/o Apprentices	Teams w/o Top Coded Wages	Men Only
$\bar{w}^+$	0.13***	0.13***	0.10***	0.16***
	(0.0066)	(0.0072)	(0.0047)	(0.0078)
$\bar{w}^-$	0.036***	0.045***	0.043***	0.046***
	(0.0043)	(0.0046)	(0.0037)	(0.0050)
Within R^2	0.81	0.81	0.78	0.79
Observations	2603447	1658118	1311475	1827269

Standard errors in parentheses

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

Notes: $\hat{\beta}^+$ and $\hat{\beta}^-$ as estimated from specification (9). Team definition 2. Column (1): Baseline. Column (2): Sample restricted to teams without workers in apprenticeship. Column (3): Sample restricted to teams without top-coded wages. Column (4): Sample restricted to men. Standard errors clustered at the establishment-year level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year (whenever possible).

Table XII Subsamples. Team Definition 2.

the alternative team definition, in table XVII. While the results vary across the samples, the main takeaway from our reduced form exercises is robust.

We next restrict the sample to workers in teams that are not restricted by collective bargaining agreements. To that end, we use IAB establishment panel for the year 2000 which asks establishments whether a binding collective bargaining agreement exists and if so if they pay above the applicable collective bargaining agreement. The survey also asks whether firms benchmark their wages with a collective bargaining agreement in case they are not subject to a binding agreement. The second columns of table XIII reports the results if we restrict the sample to workers in establishments paying, on average, at least 10% above their collective bargaining agreement in the year 2000. The third column restricts the sample to establishments that are neither subject to a collective bargaining agreement nor report to benchmark their pay structure with one.

B.3 Tables for alternative team definition 1

This subsection reports all empirical results from the main body of the paper when we define a peer group to be all workers at an establishment.

	All 2000	>10% CB	No CB, No Benchmarking
$\bar{w}^+$	0.078***	0.075***	0.064**
	(0.018)	(0.021)	(0.020)
$\bar{w}^-$	0.038***	0.040***	0.026
	(0.011)	(0.012)	(0.019)
Within R^2	0.79	0.80	0.74
Observations	334329	274471	10468

Standard errors in parentheses

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

Notes: $\hat{\beta}^+$ and $\hat{\beta}^-$ as estimated from specification (9). Team definition 2. Column (1): Benchmark results for year 2000 at horizon $n = 3$ years. Column (2): Restrict sample to establishments which report to pay at least 10% above their collective bargaining agreement. Column (3): Restrict sample to establishments which neither have a collective bargaining agreement nor benchmark their wage structure with one. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year.

Table XIII Collective Bargaining. Team Definition 2.

	Horizon in Years				
	1	2	3	5	10
$\bar{w}^+$	0.072***	0.11***	0.14***	0.20***	0.27***
	(0.0039)	(0.0051)	(0.0061)	(0.0082)	(0.018)
$\bar{w}^-$	0.023***	0.036***	0.052***	0.079***	0.12***
	(0.0026)	(0.0034)	(0.0042)	(0.0061)	(0.013)
Within R^2	0.92	0.86	0.82	0.73	0.54
Observations	3999631	3502018	3014078	2183722	511878

Standard errors in parentheses

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

Notes: $\hat{\beta}^+$ and β^- as estimated from specification (9). Team definition 1. Column titles indicate horizon n. Standard errors clustered at the establishment-year level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year.

Table XIV Estimation Results for Specification (9). Team Definition 1. Counterpart to table II.

	Panel A: All Switchers				
Horizon in Years	1	2	3	5	10
$\bar{w}^+$	0.074***	0.10***	0.13***	0.18***	0.27***
	(0.0091)	(0.0098)	(0.010)	(0.011)	(0.022)
$\bar{w}^-$	-0.017**	-0.0063	0.0051	0.028***	0.035*
	(0.0053)	(0.0056)	(0.0067)	(0.0082)	(0.017)
Within R^2	0.78	0.71	0.65	0.55	0.37
Observations	218356	261162	234310	186512	50919
	Panel B: Switchers with Unemployment Spell				
Horizon in Years	1	2	3	5	10
$\bar{w}^+$	0.053***	0.096***	0.12***	0.15***	0.23***
	(0.011)	(0.0079)	(0.0085)	(0.011)	(0.023)
$\bar{w}^-$	-0.025*	0.018*	0.026**	0.029**	0.0036
	(0.011)	(0.0071)	(0.0084)	(0.010)	(0.022)
Within R^2	0.69	0.68	0.61	0.48	0.30
Observations	21596	80346	77344	65882	18936
	Panel C: Switchers, Mass Layoff Event				
Horizon in Years	1	2	3	5	10
$\bar{w}^+$	0.083	0.059*	0.11***	0.098**	0.31***
	(0.043)	(0.026)	(0.028)	(0.037)	(0.067)
$\bar{w}^-$	0.045	0.068**	0.013	0.038	-0.051
	(0.038)	(0.026)	(0.030)	(0.035)	(0.071)
Within R^2	0.67	0.61	0.50	0.39	0.22
Observations	2359	5855	6195	5665	1873

Standard errors in parentheses

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

Notes: $\hat{\beta}^+$ and β^- as estimated from specification (9) on a sample of establishment switchers. Team Definition 1. Column titles indicate horizon n. Standard errors clustered at the establishment-year level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year.

Table XV Establishment switchers. Counterpart to table IV.

	Decile of the Wage Distribution									
	1	2	3	4	5	6	7	8	9	10
$\bar{w}^+$	0.11***	0.12***	0.15***	0.15***	0.16***	0.16***	0.14***	0.10***	0.43***	0.36***
	(0.0045)	(0.0072)	(0.0094)	(0.012)	(0.014)	(0.019)	(0.023)	(0.021)	(0.016)	(0.060)
$\bar{w}^-$	0.025***	0.046***	0.053***	0.063***	0.072***	0.077***	0.098***	0.081***	0.053***	0.010***
	(0.0047)	(0.0067)	(0.0078)	(0.0092)	(0.010)	(0.010)	(0.0097)	(0.0077)	(0.0066)	(0.0030)
Within R^2	0.56	0.12	0.072	0.054	0.051	0.054	0.066	0.099	0.24	0.077
Observations	286037	302162	305121	304840	302659	301559	301603	301866	310894	297243

	Decile of the Age Distribution									
	1	2	3	4	5	6	7	8	9	10
$\bar{w}^+$	0.22***	0.21***	0.15***	0.13***	0.11***	0.094***	0.082***	0.071***	0.053***	0.037***
	(0.0094)	(0.0094)	(0.0080)	(0.0073)	(0.0070)	(0.0071)	(0.0064)	(0.0063)	(0.0062)	(0.0059)
$\bar{w}^-$	0.022**	0.064***	0.055***	0.050***	0.047***	0.052***	0.057***	0.063***	0.059***	0.056***
	(0.0073)	(0.0063)	(0.0056)	(0.0054)	(0.0049)	(0.0049)	(0.0047)	(0.0046)	(0.0046)	(0.0040)
Within R^2	0.71	0.77	0.80	0.82	0.83	0.84	0.85	0.86	0.86	0.86
Observations	342894	282749	284374	334394	353421	231950	332531	308393	264835	278482

	Decile of the Tenure Distribution									
	1	2	3	4	5	6	7	8	9	10
$\bar{w}^+$	0.18***	0.16***	0.16***	0.16***	0.13***	0.087***	0.092***	0.079***	0.065***	0.077***
	(0.0089)	(0.0091)	(0.013)	(0.0087)	(0.0082)	(0.0090)	(0.0093)	(0.010)	(0.012)	(0.016)
$\bar{w}^-$	0.037***	0.053***	0.057***	0.062***	0.052***	0.059***	0.065***	0.072***	0.046***	0.048***
	(0.0055)	(0.0051)	(0.0059)	(0.0058)	(0.0055)	(0.0062)	(0.0073)	(0.0067)	(0.0085)	(0.0088)
Within R^2	0.76	0.80	0.82	0.83	0.84	0.84	0.83	0.83	0.78	0.78
Observations	296334	297561	301590	303271	304405	306798	305249	305519	298630	294676

	Decile of the Size Distribution									
	1	2	3	4	5	6	7	8	9	10
$\bar{w}^+$	0.055***	0.087***	0.10***	0.12***	0.090***	0.10***	0.15***	0.12***	0.16***	0.51***
	(0.0038)	(0.0069)	(0.0097)	(0.012)	(0.016)	(0.019)	(0.023)	(0.033)	(0.044)	(0.069)
$\bar{w}^-$	0.021***	0.043***	0.044***	0.036***	0.0090	0.0061	0.027	-0.035	0.036	0.0077
	(0.0029)	(0.0048)	(0.0067)	(0.0099)	(0.011)	(0.011)	(0.015)	(0.027)	(0.027)	(0.026)
Within R^2	0.82	0.80	0.79	0.77	0.76	0.76	0.74	0.68	0.67	0.72
Observations	288006	299753	299104	309208	307503	310360	309883	310238	309299	270645

Standard errors in parentheses

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

Notes: $\hat{\beta}^+$ and β^- as estimated from specification (9) for separate deciles of the wage, age, tenure, and team size distributions. We include observation i in the decile k in t if i falls into the k'th decile of the distribution in year t. Team definition 1 at horizon $n = 3$ years. Standard errors clustered at the establishment-year level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year (whenever possible).

Table XVI Baseline results for different deciles of the wage distribution. Counterpart to Table III.

	Baseline	Teams w/o Apprentices	Teams w/o Top Coded Wages	Men Only
$\bar{w}^+$	0.14***	0.070***	0.076***	0.17***
	(0.0061)	(0.014)	(0.0071)	(0.0075)
$\bar{w}^-$	0.052***	0.036***	0.044***	0.055***
	(0.0042)	(0.0070)	(0.0049)	(0.0048)
Within R^2	0.82	0.81	0.80	0.81
Observations	3014078	301826	296416	2133465

Standard errors in parentheses

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

Notes: $\hat{\beta}^+$ and $\hat{\beta}^-$ as estimated from specification (9). Team definition 2. Column (1): Baseline. Column (2): Sample restricted to teams without workers in apprenticeship. Column (3): Sample restricted to teams without top-coded wages. Column (4): Sample restricted to men. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year (whenever possible).

Table XVII Subsamples. Team Definition 1. Counterpart to table XII.

	All 2000	>10% CB	No CB, No Benchmarking
$\bar{w}^+$	0.079***	0.072***	0.050**
	(0.016)	(0.018)	(0.019)
$\bar{w}^-$	0.035***	0.046***	0.016
	(0.011)	(0.012)	(0.017)
Within R^2	0.80	0.81	0.78
Observations	386815	316739	14068

Standard errors in parentheses

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

Notes: $\hat{\beta}$ and $\hat{b}eta^-$ as estimated from specification (9). Team definition 1. Column (1): Benchmark results for year 2000 at horizon $n = 3$ years. Column (2): Restrict sample to establishments which report to pay at least 10% above their collective bargaining agreement. Column (3): Restrict sample to establishments which neither have a collective bargaining agreement nor benchmark their wage structure with one. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year.

Table XVIII Collective Bargaining. Team Definition 1. Counterpart to table XIII.

	Horizon in Years				
	1	2	3	5	10
Bin 2	0.000025	0.00016**	0.00027***	0.00039***	0.00066***
	(0.000031)	(0.000054)	(0.000070)	(0.000087)	(0.00013)
Bin 3	0.00015***	0.00027***	0.00044***	0.00068***	0.00097***
	(0.000034)	(0.000046)	(0.000055)	(0.000071)	(0.00011)
Bin 4	0.00010***	0.00024***	0.00039***	0.00055***	0.00086***
	(0.000025)	(0.000040)	(0.000047)	(0.000058)	(0.00010)
Bin 5	-0.0000010	0.00011**	0.00025***	0.00050***	0.0010***
	(0.000023)	(0.000036)	(0.000044)	(0.000061)	(0.00012)
Bin 6	0.000024	0.00015**	0.00029***	0.00049***	0.00065***
	(0.000034)	(0.000048)	(0.000061)	(0.000073)	(0.00011)
Bin 7	0.00032***	0.00045***	0.00061***	0.00084***	0.0013***
	(0.000025)	(0.000036)	(0.000041)	(0.000058)	(0.00013)
Bin 8	0.00044***	0.00064***	0.00087***	0.0012***	0.0016***
	(0.000027)	(0.000040)	(0.000045)	(0.000060)	(0.00013)
Bin 9	0.00050***	0.00075***	0.0010***	0.0014***	0.0019***
	(0.000034)	(0.000049)	(0.000054)	(0.000070)	(0.00013)
Bin 10	0.00057***	0.00087***	0.0012***	0.0016***	0.0021***
	(0.000036)	(0.000056)	(0.000061)	(0.000083)	(0.00015)
Bin 11	0.00056***	0.00091***	0.0013***	0.0019***	0.0027***
	(0.000030)	(0.000046)	(0.000054)	(0.000075)	(0.00013)
Within R^2	0.92	0.87	0.82	0.74	0.54
Observations	4038521	3535129	3040881	2202276	515689

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. Column titles indicate horizon n. Standard errors clustered at the establishment-year level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year.

Table XIX Results from specification (10) under team definition 1. Counterpart to table X.

	All	Above Team-Median	Below Team-Median	2nd Pct.	4th Pct.	7th Pct.	9th Pct.
Bin 2	0.00027***	0.00020	0.00027***	0.00020	0.00010	0.00042***	0.00026***
	(0.000070)	(0.00044)	(0.000062)	(0.00018)	(0.00019)	(0.00012)	(0.000059)
Bin 3	0.00044***	0.0015***	0.00036***	0.00059***	-0.00015	0.00061***	0.00030***
	(0.000055)	(0.00040)	(0.000050)	(0.00017)	(0.00014)	(0.00010)	(0.000066)
Bin 4	0.00039***	0.0011***	0.00029***	0.00026	0.00014	0.00050***	0.00020*
	(0.000047)	(0.00035)	(0.000045)	(0.00014)	(0.00014)	(0.00010)	(0.000089)
Bin 5	0.00025***	0.0011***	0.00026***	0.00031*	0.00032*	0.00045***	0.00048***
	(0.000044)	(0.00033)	(0.000050)	(0.00013)	(0.00014)	(0.00012)	(0.00011)
Bin 6	0.00029***	0.0011***	0.00029***	0.00027*	0.00035**	0.00017	0.00018*
	(0.000061)	(0.00030)	(0.000063)	(0.00013)	(0.00011)	(0.00014)	(0.000072)
Bin 7	0.00061***	0.0015***	0.00054***	0.00047***	0.00044**	0.0012***	0.00055***
	(0.000041)	(0.00031)	(0.000049)	(0.00013)	(0.00016)	(0.00019)	(0.000054)
Bin 8	0.00087***	0.0017***	0.0011***	0.00092***	0.00074***	0.00066***	0.00088***
	(0.000045)	(0.00030)	(0.000075)	(0.00013)	(0.00015)	(0.00015)	(0.000065)
Bin 9	0.0010***	0.0019***	0.0012***	0.0011***	0.00065***	0.0010***	0.0013***
	(0.000054)	(0.00031)	(0.000100)	(0.00014)	(0.00017)	(0.00011)	(0.00011)
Bin 10	0.0012***	0.0021***	0.0011***	0.0012***	0.00062**	0.0013***	0.0043**
	(0.000061)	(0.00031)	(0.00013)	(0.00016)	(0.00024)	(0.00013)	(0.0015)
Bin 11	0.0013***	0.0022***	0.0014***	0.0016***	0.0013***	0.0014***	0
	(0.000054)	(0.00031)	(0.00013)	(0.00012)	(0.00012)	(0.00011)	(.)
Within R^2	0.82	0.79	0.82	0.24	0.098	0.095	0.18
Observations	3040881	1463699	1577179	146055	314846	393874	377783

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. Columns 2 and 3 report the results when the sample is restricted to workers above (below) the team median wage. The remaining columns restrict the sample to workers from particular parts of the wage distribution. Team definition 1. Column titles indicate horizon n. Standard errors clustered at the establishment-year level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year.

Table XX Results from specification (10) under team definition 1 for various restricted samples. Counterpart to table XI