ECONOMETRICA

JOURNAL OF THE ECONOMETRIC SOCIETY

An International Society for the Advancement of Economic Theory in its Relation to Statistics and Mathematics

https://www.econometricsociety.org/

Econometrica, Vol. 89, No. 2 (March, 2021), 647–676

LEARNING FROM COWORKERS

GREGOR JAROSCH
Department of Economics and SPIA, Princeton University

EZRA OBERFIELD
Department of Economics, Princeton University

ESTEBAN ROSSI-HANSBERG
Department of Economics and SPIA, Princeton University

Econometrica, Vol. 89, No. 2 (March, 2021), 647–676

LEARNING FROM COWORKERS

GREGOR JAROSCH
Department of Economics and SPIA, Princeton University

EZRA OBERFIELD
Department of Economics, Princeton University

ESTEBAN ROSSI-HANSBERG
Department of Economics and SPIA, Princeton University

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 heterogeneous 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. The methodology builds on the observation that a competitive labor market prices coworker learning. 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. We show that between 4 and 9% of total worker compensation is in the form of learning and that inequality in total compensation is significantly lower than inequality in wages.

KEYWORDS: Knowledge diffusion, production in teams, growth, income distribution, peer effects.

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 working, 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 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.

Gregor Jarosch: gregorjarosch@gmail.com
Ezra Oberfield: edo@princeton.edu
Esteban Rossi-Hansberg: erossi@princeton.edu

We thank the IR Section at Princeton University for providing a secured facility to access the data and the IES at Princeton University for financial support. We thank our discussant Hugo Hopenhayn and participants at various seminars and conferences for useful comments. We also thank Sara Casella and Mingduo Zhao for research assistance. This study uses the longitudinal model of the Linked-Employer-Employee Data (LIAB) (Version LM 9310, Years 1993–2010) from the IAB. Data access was provided via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and subsequently remote data access.

 https://doi.org/10.3982/ECTA16915

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 start by exploring 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 a small sensitivity of wage growth to the wages of less-well-paid workers and larger sensitivity to the wages of those higher up in the wage distribution. We also show that the effects we find are present across the wage distribution, for workers that switch plants, and for workers that leave their employers as a result of a mass layoff. In addition, we show how the effects vary with worker tenure, age, establishment size, and with the occupation of the team of coworkers. We also show that they are robust to controlling for employment and wage bill growth at the worker's plant.

Our results are strongly suggestive of significant learning from coworkers, particularly from workers that earn more. Furthermore, although the battery of checks and different specifications we present can never discard all alternative interpretations, we argue that our evidence can rule out many of them. In particular, these findings cannot purely reflect most forms of back-loading, mean reversion, sorting, rent-sharing, as well as certain alternative forms of worker learning unrelated to coworkers.[1]

Motivated by these reduced-form facts, we 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. The labor market also 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 measure the magnitude and characteristics of learning on the job. Our model yields a mapping between the matched employer-employee data and the underlying knowledge and learning of individuals. Thus, 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 micro data for the German labor market. Our methodology uses only information on each worker's wage and the wages of her coworkers.

A first step in implementing our approach is to choose a cardinality for knowledge. We show that the expected present value of income is a natural and useful choice for the units of knowledge. It simplifies the empirical implementation of our approach and allows for a natural interpretation of the estimated learning functions. We proceed by showing that, given the learning function, we can invert individual Bellman equations to recover the knowledge of individuals from the full set of wages of the members of each production

[1]To further show that wage growth is related to the composition of coworkers, in Appendix C.1 of the Supplemental Material (Jarosch, Oberfield, and Rossi-Hansberg (2021)) we also present an exercise where we build an instrument for changes in coworker composition using workers that die or otherwise disappear unexpectedly from the data set.

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 the learning function we had imposed. 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 strong but standard assumptions like complete financial markets (or linear utility), perfect labor market competition, and stationarity. However, since the estimation strategy relies solely on the individual Bellman equations together with observed wages and team composition, it imposes minimal restrictions on the set of firm technologies and complementarities across workers. 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.

We implement this structural estimation strategy using the German data and find that agents' learning is relatively insensitive to the knowledge of less knowledgeable workers and significantly more sensitive to those with more knowledge, particularly from the most knowledgeable members of their teams. On average, between 4 and 9% of the total compensation of workers comes in the form of learning from coworkers in the same team (either same establishment or same establishment and occupation). We also show that inequality in wages is between one third and one fifth larger than inequality in compensation because workers with different levels of knowledge differ 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 almost double if workers were to be grouped in teams randomly. The finding suggests the presence of knowledge complementarities in production.

Our methodology to estimate the learning function can be extended to incorporate other observable worker characteristics beyond their level of knowledge. In particular, we explore the role of age, and show that agents that are younger than 40 learn more rapidly, particularly from other young peers. We also generalize our methodology in order to incorporate search frictions in the labor market, rent-sharing with heterogeneous firms, imperfect information, and incomplete markets. An empirical implementation of these extensions requires a richer data set than the one we have available, but they should be helpful to guide researchers with access to dynamic employer-employee panel data.

There is a large literature in macroeconomics that has used learning from others as the key mechanism to generate aggregate growth. Lucas (2009) proposed a theory of growth based on random meetings between agents in the entire population. Lucas and Moll (2014) and Perla and Tonetti (2014) extended these models to add a time allocation choice, while Jovanovic and Rob (1989), Jovanovic and MacDonald (1994), and König, Lorenz, and Zilibotti (2016) focused on the innovation/imitation margin. Other models like Sampson (2015), Perla, Tonetti, and Waugh (2021), and Luttmer (2014) also generate growth through adoption of ideas from others. As Alvarez, Buera, and Lucas (2013) and Buera and Oberfield (2020) showed, 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, Lucas, and Rossi-Hansberg (2019) introduced 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) studied learning in teams of two, while Burstein and Monge-Naranjo (2009) studied an environment in which a manager hires identical workers and imparts knowledge to those workers.[2] We go further than these papers in that we model learning within teams and provide direct evidence of its importance and its characteristics. We also provide 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, Dustmann, and Schönberg (2017)), empirical work of learning within teams is much more scarce. Nix (2015) argued that increasing the average education of one's peers raises one's earnings in subsequent years. Akcigit, Caicedo, Miguelez, Stantcheva, and Sterzi (2018) argued 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, Lise, Menzio, and Phillips (2018) built on a frictional sorting setup to investigate learning with production complementarities. Like us, they detected strong coworker spillovers. The main difference between both papers is that our competitive labor market model allows us to structurally estimate the model without imposing restrictions on the production technology. Our strategy is thus well-suited for utilizing the full richness of the matched employer-employee data. Furthermore, our model features teams with arbitrary numbers of heterogeneous workers, which allows us to study the role of the within-team distribution of knowledge for the coworker learning process. A limitation of our analysis is that wages reflect only knowledge and compensating differential from learning. While we view this as an important benchmark, it certainly misses some components of wages, such as rent-sharing or adjustment frictions. An important advantage of Herkenhoff et al. (2018) is that, in addition to knowledge and compensating differentials, wages reflect rent-sharing. A limitation is that, to inform the division of wages into these three components, they have to lean heavily on the rest of the structure of the model, and in particular on auxiliary assumptions on the production function.

The remainder of this paper is organized as follows. In Section 2, we present a number of reduced-form findings about the relationship between the wage of an individual, the wages of her coworkers, and the individual's subsequent wage growth. Section 3 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. Section 4 contains our main results. We present and implement an algorithm to structurally estimate the learning function introduced in Section 3, and describe the implications for individual investment in knowledge and inequality. Section 5 concludes. The Supplemental Material (Jarosch, Oberfield, and Rossi-Hansberg (2021)) discusses multiple extensions and alternative interpretations of our model, presents results when we condition on a worker's age, introduces our German matched employer-employee data set, and presents additional reduced-form results. A Supplementary Appendix discusses the detailed construction of our data set and presents results for an alternative team definition.

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

2. REDUCED-FORM EVIDENCE

We start our analysis using German social security data to investigate the relationship between coworker (relative) wages and individual wage growth. Our goal is to provide empirical evidence of workers learning from their teammates. To do so, we relate individual wage dynamics to wages in the peer group using various flexible reduced-form specifications. We explore three empirical relationships that could indicate learning from coworkers. 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 alternatives to learning which could be driving our initial findings, like a back-loaded wage structure, mean reversion in wages, sorting, or rent-sharing within firms. At the end of the section we take stock and discuss what we can conclude from the evidence we present.

We use a German data set that contains the complete set of workers at a sample of establishments from 1999 to 2009. The data set contains information on a worker's establishment, occupation, and average daily earnings along with a rich set of other worker characteristics (age, gender, job and employment tenure, education, location, among others). Throughout, we work with two different ways of defining a peer group. In Team Definition 1, a team is defined by all workers in the same establishment. In Team Definition 2, a team is defined as all workers in the same establishment and occupation. Data Appendix B of the Supplemental Material presents summary statistics, and Supplementary Appendix D describes the construction of the data set.

2.1. *Regression Framework*

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

$$w_{i,t+h} = \alpha + \beta \bar{w}_{-i,t} + \gamma w_{i,t} + \omega_{\text{age}} + \omega_{\text{tenure}} + \omega_{\text{gender}} + \omega_{\text{educ}} + \omega_{\text{occ}} + \omega_t + \varepsilon_{i,t}. \quad (1)$$

$w_{i,t+h}$ 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 $h > 1$ use only a subset of years t. For instance, for $h = 10$ we are restricted to exclusively use information about peers in the year 1999 and 2000 (since we observe workers until 2010). Likewise, for $h = 5$ we can use all years between 1999 and 2005.

We report the parameter estimates for each team definition in Table I, clustering standard errors at the establishment 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 almost 7%. These effects are naturally larger as the horizon extends further into the fu-

TABLE I

ESTIMATION RESULTS FOR SPECIFICATION (1)[a]

Horizon in Years	Narrow Team Definition				
	1	2	3	5	10
$\bar{w}$	0.068 (0.0036)	0.094 (0.0052)	0.12 (0.0071)	0.16 (0.0094)	0.21 (0.014)
Within R^2	0.89	0.82	0.77	0.68	0.47
Observations	3,969,166	3,473,642	2,989,524	2,167,851	508,504

Horizon in Years	Broad Team Definition				
	1	2	3	5	10
$\bar{w}$	0.058 (0.0031)	0.083 (0.0046)	0.11 (0.0063)	0.15 (0.0086)	0.21 (0.014)
Within R^2	0.89	0.83	0.78	0.68	0.48
Observations	4,112,284	3,595,957	3,092,504	2,239,223	523,683

[a]*Notes*: $\hat{\beta}$ as estimated from specification (1). Column titles indicate horizon n. Standard errors clustered at the establishment level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year. Standard errors in parentheses.

ture, but they do not grow 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 21% 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 or more useful. Thus, in the rest of this section, we restrict our attention to Team Definition 2. We separately report all results for the alternative team definition in Supplementary Appendix E.

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:

$$w_{i,t+h} = \alpha + \beta^+ \bar{w}^+_{-i,t} + \beta^- \bar{w}^-_{-i,t} + \gamma w_{i,t} + \omega_{\text{age}} + \omega_{\text{tenure}} + \omega_{\text{gender}} + \omega_{\text{educ}} + \omega_{\text{occ}} + \omega_t + \varepsilon_{i,t}, \quad (2)$$

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 five times larger at all horizons. These findings are consistent with larger knowledge flows from more-highly-skilled peers and relatively little sensitivity to the knowledge of less-skilled peers. Table III shows similar results for the alternative Team Definition 1 that includes all workers in an establishment. Since we find this stark asymmetry to be robust throughout, we henceforth build on specification (2).

TABLE II

ESTIMATION RESULTS FOR SPECIFICATION (2)[a]

	Horizon in Years				
	1	2	3	5	10
$\bar{w}^+$	0.090 (0.0064)	0.13 (0.010)	0.16 (0.015)	0.22 (0.021)	0.28 (0.032)
$\bar{w}^-$	0.019 (0.0048)	0.029 (0.0065)	0.041 (0.0081)	0.060 (0.011)	0.097 (0.016)
Within R^2	0.88	0.81	0.76	0.66	0.46
Observations	3,462,305	3,034,301	2,617,097	1,903,104	448,560

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

Table C.VIII in Appendix C.3 of the Supplemental Material presents additional results when we use the same specification but, in addition, control for either establishment wage bill growth or establishment employment growth. Specifically, we control for the growth in total wage bill (or total count) of the full time employed at the establishment between $t-2$ and $t-1$, $t-1$ and t, and t and $t+1$. The results remain virtually identical, indicating that the relationship between wage growth and coworker composition is not driven by changes in these establishment-level characteristics.[3]

TABLE III

ESTIMATION RESULTS FOR SPECIFICATION (2)—TEAM DEFINITION 1[a]

	Horizon in Years				
	1	2	3	5	10
$\bar{w}^+$	0.090 (0.0058)	0.13 (0.0089)	0.17 (0.012)	0.23 (0.017)	0.32 (0.026)
$\bar{w}^-$	0.025 (0.0044)	0.039 (0.0060)	0.057 (0.0081)	0.085 (0.012)	0.12 (0.019)
Within R^2	0.89	0.82	0.77	0.68	0.48
Observations	4,026,321	3,522,994	3,032,228	2,197,932	515,017

[a]*Notes*: $\hat{\beta}^+$ and β^- as estimated from specification (2). Team Definition 1. Column titles indicate horizon h. Standard errors clustered at the establishment level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year. Standard errors in parentheses.

[3]We point to additional robustness results which are relegated to Appendix C.3. There, we show that the patterns documented 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, and we show results when we split the sample into prior and after the Hartz labor market reforms. We also report results when we include more high-dimensional fixed effects and when we control for many lags of the individual log wage. Most importantly, we show our baseline results under the inclusion of team fixed effects under both team definitions. As can be seen from the corresponding tables in Appendix C.3, our baseline results appear mostly insensitive to these modifications.

2.1.1. *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.[4] We also repeat the exercise for different deciles of the age, tenure, and team size distribution.

Our first set of results are reported in Panel A of Table IV which presents the regression coefficients for specification (2) which we run separately for each decile of the wage distribution. The results are fairly stable across the wage distribution.[5] We conclude that having more-highly-paid coworkers is associated with future wage growth, largely independent of the current level of wages.

Our next set of results is reported in Panel B of Table IV 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.

Panel C of Table IV also shows that the patterns we 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.

Panel D of Table IV gauges the role of team size. It shows that the patterns we describe are present in small and large teams alike. The estimated relationships are smaller for teams with fewer workers and become somewhat less precise at the very top.[6] The results are consistent with knowledge flows per worker that increase with team size, but that are significant throughout the team size distribution.

2.1.2. *Switchers*

In a competitive labor market, if the relationship between wage growth and team composition is the result of learning, the resulting knowledge has to be valued outside the establishment. We now explore whether the relationship we have uncovered is also present for workers that switch teams. Thus, we run specification (2) 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 $t+1$.[7]

The results are reported in Panel A of Table V. The table shows that more-highly-paid peers are associated with higher wage growth even for those who move to a different

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

[5]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 $\bar{w}^{+}$.

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

[7]As we discuss in Supplementary Appendix D, 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.

TABLE IV

BASELINE RESULTS FOR DIFFERENT DECILES OF THE WAGE, AGE, TENURE, AND TEAM SIZE DISTRIBUTION[a]

	Panel A: Decile of the Wage Distribution									
	1	2	3	4	5	6	7	8	9	10
$\bar{w}^+$	0.18	0.13	0.14	0.13	0.13	0.11	0.11	0.10	0.32	0.35
	(0.012)	(0.015)	(0.018)	(0.018)	(0.021)	(0.026)	(0.028)	(0.023)	(0.021)	(0.060)
$\bar{w}^-$	0.043	0.041	0.045	0.047	0.048	0.046	0.071	0.069	0.057	0.024
	(0.0097)	(0.010)	(0.012)	(0.014)	(0.015)	(0.015)	(0.015)	(0.0096)	(0.010)	(0.0058)
Within R^2	0.43	0.086	0.053	0.040	0.037	0.039	0.051	0.080	0.17	0.062
Observations	248,920	263,528	265,462	264,823	262,632	261,579	261,499	261,687	262,796	264,087

	Panel B: Decile of the Age Distribution									
	1	2	3	4	5	6	7	8	9	10
$\bar{w}^+$	0.31	0.27	0.16	0.12	0.097	0.089	0.058	0.035	0.018	0.0092
	(0.021)	(0.026)	(0.017)	(0.014)	(0.013)	(0.012)	(0.010)	(0.0092)	(0.0093)	(0.0088)
$\bar{w}^-$	0.023	0.032	0.030	0.038	0.045	0.046	0.056	0.061	0.065	0.056
	(0.014)	(0.011)	(0.0089)	(0.0084)	(0.0088)	(0.0083)	(0.0082)	(0.0073)	(0.0079)	(0.0069)
Within R^2	0.61	0.70	0.75	0.77	0.78	0.79	0.80	0.82	0.82	0.82
Observations	313,804	256,870	253,754	294,113	205,242	301,874	284,811	261,404	221,992	223,184

	Panel C: Decile of the Tenure Distribution									
	1	2	3	4	5	6	7	8	9	10
$\bar{w}^+$	0.26	0.25	0.19	0.16	0.10	0.077	0.062	0.050	0.035	0.012
	(0.023)	(0.027)	(0.025)	(0.017)	(0.014)	(0.014)	(0.012)	(0.013)	(0.014)	(0.014)
$\bar{w}^-$	0.012	0.022	0.038	0.059	0.051	0.061	0.062	0.065	0.050	0.059
	(0.011)	(0.0086)	(0.0090)	(0.0091)	(0.0085)	(0.0093)	(0.0089)	(0.010)	(0.013)	(0.016)
Within R^2	0.66	0.73	0.77	0.79	0.80	0.79	0.79	0.78	0.74	0.74
Observations	259,787	258,018	261,847	264,724	314,261	216,646	298,387	231,182	261,285	250,908

	Panel D: Decile of the Size Distribution									
	1	2	3	4	5	6	7	8	9	10
$\bar{w}^+$	0.057	0.097	0.13	0.16	0.16	0.23	0.24	0.25	0.16	0.16
	(0.0049)	(0.0081)	(0.013)	(0.017)	(0.020)	(0.027)	(0.043)	(0.054)	(0.040)	(0.11)
$\bar{w}^-$	0.027	0.040	0.050	0.056	0.042	0.021	−0.022	−0.084	−0.056	−0.11
	(0.0036)	(0.0063)	(0.0080)	(0.010)	(0.014)	(0.016)	(0.018)	(0.030)	(0.044)	(0.074)
Within R^2	0.78	0.77	0.75	0.74	0.72	0.70	0.70	0.66	0.61	0.65
Observations	263,527	276,795	256,020	256,637	266,652	259,741	260,926	259,466	257,506	259,821

[a]*Notes*: $\hat{\beta}^+$ and β^- as estimated from specification (2) 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 kth decile of the pooled distribution. Team Definition 2 at horizon $h = 3$ years. Standard errors clustered at the establishment level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year (whenever possible). Standard errors in parentheses.

TABLE V

ESTABLISHMENT SWITCHERS[a]

Horizon in Years	Panel A: All Switchers				
	1	2	3	5	10
$\bar{w}^+$	0.092	0.17	0.20	0.25	0.35
	(0.018)	(0.022)	(0.024)	(0.027)	(0.032)
$\bar{w}^-$	−0.052	−0.031	−0.019	−0.0082	0.026
	(0.011)	(0.0099)	(0.012)	(0.014)	(0.022)
Within R^2	0.59	0.55	0.49	0.40	0.26
Observations	194,848	228,110	203,726	160,495	43,609

Horizon in Years	Panel B: Switchers With Nonemployment Spell				
	1	2	3	5	10
$\bar{w}^+$	0.094	0.19	0.20	0.24	0.31
	(0.020)	(0.019)	(0.021)	(0.023)	(0.032)
$\bar{w}^-$	−0.039	0.024	0.044	0.036	0.061
	(0.022)	(0.014)	(0.016)	(0.018)	(0.030)
Within R^2	0.37	0.46	0.39	0.31	0.19
Observations	21,084	72,223	68,781	57,331	16,224

Horizon in Years	Panel C: Switchers, Mass Layoff Event				
	1	2	3	5	10
$\bar{w}^+$	0.11	0.14	0.14	0.22	0.34
	(0.056)	(0.046)	(0.051)	(0.055)	(0.089)
$\bar{w}^-$	0.032	0.032	0.064	0.049	−0.016
	(0.069)	(0.042)	(0.048)	(0.054)	(0.11)
Within R^2	0.34	0.34	0.28	0.21	0.14
Observations	2264	5258	5453	4904	1545

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

establishment. Furthermore, we find the same marked asymmetry we documented using all workers. 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 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 might be the result of the fact that switchers tend to be young, and young workers' wage growth is more sensitive to peers' wages.

To account for some of these forms of selection, we also present results for only those switchers who experience an interim nonemployment 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 V.

We go further and restrict the sample to switchers with an interim spell of nonemployment whose reference-spell employer also experienced a mass layoff event in year t.[8] Arguably, constructing the sample this way controls more fully for selection. As can be seen in Panel C of Table V, our corresponding estimates remain large and significant, although they tend to fall somewhat. Naturally, the results are less precise since we lose many observations.

2.1.3. *Coworker Occupation*

Learning could depend on the occupation of the coworkers with whom one interacts. Hence, we offer results for specifications which split individuals working in the same establishment as worker i into groups. The first exercise explores whether wage growth is more strongly related to the wages of teammates that are employed in managerial occupations.[9] Table VI presents the results for our baseline specification in (1) after weighting for the share of each group. Perhaps surprisingly, it shows that the relationship between wage growth and peer wages is almost identical, and positive, for peers that are managers or those that are workers. Under our interpretation, the managerial classification does not seem to be essential for worker learning.[10]

Perhaps workers learn more from their peers in the same occupation. Hence, we go further and divide coworkers into those working in the same occupation and those working in different occupations. We explore this using the specification in (2) where we 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 (2) and report the results in Table VII.

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 occupation than in alternative ones. 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

TABLE VI
MANAGER AND WORKER PEERS[a]

	Horizon in Years				
	1	2	3	5	10
$\bar{w}_w$	0.062 (0.0039)	0.090 (0.0058)	0.12 (0.0077)	0.16 (0.010)	0.21 (0.017)
$\bar{w}_m$	0.058 (0.0036)	0.084 (0.0053)	0.11 (0.0076)	0.16 (0.010)	0.22 (0.019)
Within R^2	0.88	0.82	0.77	0.67	0.47
Observations	3,517,568	3,086,436	2,661,936	1,934,856	458,302

[a]*Notes*: Variables are weighted: To construct $\bar{w}_m$, we compute the mean wage of workers in managerial occupation(s) and then weight by the fraction of overall establishment employment of managers. Column titles indicate horizon h. Standard errors clustered at the establishment level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year. Standard errors in parentheses.

[8]See Supplementary Appendix D for details on the mass layoff definition.
[9]We classify as managers workers in 3-digit KldB_88 occupations 751–753 and 761–763.
[10]We obtain similar conclusions if we use the wage gap to workers and managers that earn more.

TABLE VII

PEERS IN THE SAME AND IN OTHER OCCUPATIONS[a]

	Horizon in Years				
	1	2	3	5	10
$\bar{w}^{+}_{\text{same occ}}$	0.12 (0.013)	0.16 (0.019)	0.21 (0.026)	0.29 (0.037)	0.39 (0.048)
$\bar{w}^{-}_{\text{same occ}}$	−0.00039 (0.012)	0.0076 (0.017)	0.013 (0.023)	0.022 (0.033)	0.043 (0.047)
$\bar{w}^{+}_{\text{other occ}}$	0.086 (0.0068)	0.12 (0.011)	0.15 (0.015)	0.21 (0.022)	0.30 (0.037)
$\bar{w}^{-}_{\text{other occ}}$	0.038 (0.0064)	0.054 (0.0089)	0.076 (0.012)	0.11 (0.018)	0.15 (0.027)
Within R^2	0.88	0.81	0.76	0.66	0.46
Observations	3,315,351	2,907,077	2,509,645	1,827,701	431,782

[a]*Notes*: $\hat{\beta}^{+}$ and β^{-} as estimated from specification (2) with separate coefficients for peers in the same occupation and in other occupations. Variables are weighted: To construct $\bar{w}^{+}_{\text{same occ}}$ ($\bar{w}^{+}_{\text{other occ}}$), we compute the mean wage of workers above in the wage distribution in the same (other) occupation(s) and then weight by the fraction of overall establishment employment the (other) occupation(s) accounts for. Column titles indicate horizon h. Standard errors clustered at the establishment level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year. Standard errors in parentheses.

that use different knowledge, they learn from everyone since they may know less of the topic themselves.

2.1.4. *A More Flexible Specification*

The results above, based on specification (2), exploit only two moments of the distribution of coworker wages, namely, the gap to coworkers with wages above and below. We continue the reduced-form exploration with an exercise that attempts to approximate the wage distribution surrounding a worker in a more flexible way. To do so, we divide a worker's peers into 11 bins. The bottom bin takes peers j with wage such that $\log(w_j) - \log(w_i) < -0.45$, while the top bin takes those peers with $\log(w_j) - \log(w_i) > 0.45$. 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:

$$w_{i,t+h} = \sum_{k=2}^{11} \beta_k p_{i,k,t} + \gamma w_{i,t} + \omega_{\text{age}} + \omega_{\text{tenure}} + \omega_{\text{gender}} + \omega_{\text{educ}} + \omega_{\text{occ}} + \omega_t + \varepsilon_{i,t}. \tag{3}$$

That is, we project log wages h years ahead on the current log wage and a nonparametric approximation of the current peer wage distribution around a worker, along with our standard controls and fixed effects. We present the results in Figure 1.

Figure 1 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 1.5%. The figure confirms the findings from the previous exercises: Those who are less well paid (those in bins 5 and under) have similar effects 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

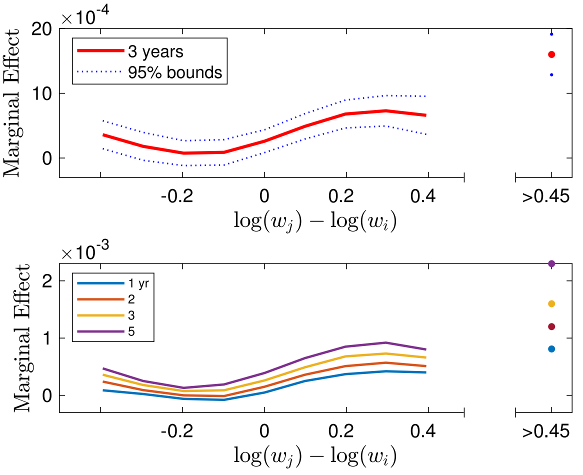

FIGURE 1.—Approximating the wage distribution. *Notes*: We plot the coefficients $\hat{\beta}_k$ from regression specification (3) with weights p_k scaled such that they add up to 100. The bin $k = 1$ (which takes the peers i such that $\log(w_j) - \log(w_i) < -0.45$) is the omitted category. The top 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 $\log(w_j) - \log(w_i) > 0.45$ are in one single bin as indicated by the break in the axis and the lines. The figure uses Team Definition 2. Standard errors clustered at the establishment level. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year.

higher). Note that, while workers benefit more from more-highly-paid peers, the effects are less than proportional.[11] This suggests 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 1 also confirms that learning from higher earners accumulates over time presumably because the composition of teams is highly persistent. We report the table underlying Figure 1 in Appendix C.2 of the Supplemental Material along with the corresponding results for the other team definition. In addition, we report the results, for 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 1 are generally confirmed.

2.1.5. *Taking Stock*

Put together, the results in this section are, we believe, strongly consistent with the view that workers learn from their peers. In a labor market in which wages are monotone in knowledge conditional on team composition, wage growth is naturally associated with learning. Learning is naturally the result of interactions with those that possess knowledge, so having more knowledgeable peers should lead to more learning. We show that, in fact, wage growth is positively associated with the wage gap to coworkers, and particularly with the wage gap to coworkers that earn more and are in similar occupations.

[11]The top bin collects all peers j such that $\log(w_j) - \log(w_i) > 0.45$ and thus does not directly compare with the other groups.

The resulting increase in knowledge is embedded in workers, and therefore reflected in wages, even for workers that leave the establishment for, arguably, exogenous reasons.

Besides coworker learning, there are other mechanisms that might potentially drive some of the patterns uncovered above. The entirety of the evidence presented above allows us to rule out many of them. We group alternative mechanisms in five main classes and discuss each in turn.

Wage Back-Loading. Certain plausible wage back-loading patterns could give rise to some of our 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).

Note, however, that the patterns we document are present and similar in size for team-switchers, including those who arguably leave their 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 evidence showing that our baseline effects are present across the tenure and age distribution.

Mean Reversion. There are two main forms of mean reversion in wages that could potentially affect our results: economy-wide and team-specific. Economy-wide mean reversion is inconsistent with Table IV Panel A. There, we show that the magnitude of the relationship between wage growth and the wage gap does not decline as we focus on higher wage deciles.

Furthermore, the sharp asymmetry between $\hat{\beta}^{+}$ and $\hat{\beta}^{-}$, and the pattern of marginal effects in Figure 1, are hard to reconcile with an explanation which builds on within-team mean reversion in wages. For example, suppose deviations of one's wage from the average are idiosyncratic and not due to productivity differences, and therefore may be unrelated to future wages within one's own team or other teams. Then, differences in the timing of raises unrelated to changes in productivity could give rise to a positive relationship between wage gaps and future wage growth. Workers that have not received raises might later seek other offers, or might later get raises due to equity concerns, or might quit to find a better-paying job. This mechanism, however, does not lead to effects that are stronger for the gap to those that earn more than for the gap to those that earn less. In addition, the patterns should not persist when we focus on those that switch jobs due to mass layoffs.

Rent-Sharing. The simplest forms of rent-sharing, in which a firm gets a windfall or increases profitability and, through bargaining, raises the wages of all workers, cannot explain our results. The reason is that the shock does not affect the gap to coworkers, and would therefore not affect the coefficient of interest. In particular, this mechanism does not provide any reason why a contemporaneous gap between one's wage and coworkers' wages would predict future wage growth.

It is conceivable that our results could in part reflect a form of staggered rent-sharing. Specifically, suppose a firm's marginal product rises, which increases rents to be shared, but some workers receive raises before others. We would expect the firms whose marginal product rises to also increase employment or raise wages. However, as we show in Appendix C.3, the baseline results hardly change when we include establishment-level growth controls.

It could also be that a firm experiences windfalls that lead to asymmetric raises among employees, and the pattern of which coworkers get larger raises (due to greater rent-sharing) happens to covary with the distribution of coworkers in just the right way. However, the fact that the relationship is similar in magnitude among those who switch firms due to a mass layoff is inconsistent with this mechanism.

Worker Sorting. In principle, worker sorting could also be a mechanism behind our findings. Workers could have heterogeneous income profiles: some workers will experience faster wage growth than others regardless of their coworkers (conditional on observables such as age, experience, etc.). It is possible that our estimates of learning from coworkers are confounded by the sorting of workers across teams if sorting is such that high growth individuals happen to be at the lower end of the within-team wage distribution. To check if this is driving our results, we use a specification that controls for the worker's wage growth over the last five years in Table C.VII of Appendix C.3.[12] The resulting relationship between future wage growth and coworker wages is similar, albeit a bit smaller, than our baseline specification, suggesting that our results are not driven by worker sorting.

Other Forms of Learning. Beyond learning from coworkers, individual learning in the workplace can result from learning-by-doing or from diffusion of a firm's embedded knowledge. For example, our results could also perhaps be explained by team-wide events that lead employees to learn, with the new knowledge spreading to workers in a way that improves their labor market prospects elsewhere. However, the fact that our results barely change when we include establishment-level employment growth and wage bill growth controls limits the importance of this alternative mechanism.[13]

More generic stories in which learning is sufficiently flexible and arbitrarily covaries with flexible moments of the distribution of wages within or outside of the firm are hard to rule out in general, particularly if they do not lead to overall employment or wage bill growth or large differences in learning across the size distribution (which we also do not find in Table IV Panel D). Therefore, to move forward, we need to isolate the causal effect of team composition on wage growth. In Appendix C.1, we propose and execute such an identification strategy using (pseudo-)random exit events where a full-time employed, prime-age worker permanently leaves employment (as in Jäger and Heining (2019)). Although noisy, these results are consistent with our interpretation.

The reasoning above suggests that learning is a natural and parsimonious rationalization of the facts we have uncovered, and that it is not obvious how to rationalize them using alternative mechanisms. Of course, except for the somewhat noisier instrumental variable results in Appendix C.1, the rest of our reduced-form results cannot be understood as causal and so provide simply a suggestive statement about equilibrium relationships. In the next section, we propose a theory of learning motivated by this evidence. We

[12]Specifically, we control for the log of the wage in each of the last five years to allow for more flexible time series processes in income profiles.

[13]It could also be that some firms engender learning more than others, and that team formation is such that, on average, those in high-learning firms tend to have more-highly-paid coworkers than those in low-learning firms. To assess this, we incorporate team fixed effects into our baseline specification. Such a specification cannot identify the impact of having better coworkers on average, as this is approximately co-linear with one's own wage and the team fixed effect. Nevertheless, we can still identify the asymmetry between the gap to more-highly-paid coworkers and the gap to less-well-paid coworkers. Table C.VI of Appendix C.3 shows that this asymmetry is similar to our baseline specification.

use the suggested specification of the learning function to structurally estimate the model using the same data.

3. A BENCHMARK MODEL

Motivated by the evidence in the previous section, we develop a theory of worker learning from coworkers in a competitive labor market. Consider an economy populated by a unit mass of heterogeneous 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 workers 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 $\mathbf{z}$.

3.1. *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

$$\pi(a) = \max_{\mathbf{z}} F(\mathbf{z}; a) - W(\mathbf{z}). \tag{4}$$

Let $\mathbf{z}(a) = \arg\max_{\mathbf{z}} F(\mathbf{z}; a) - W(\mathbf{z})$ denote a's optimal choice.[14]

3.2. *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

$$V(z) = \max_{\tilde{\mathbf{z}} \in \tilde{\mathbf{Z}}} w(z; \tilde{\mathbf{z}}) + \beta \int_0^\infty V(z')\, dG(z'|z, \tilde{\mathbf{z}}). \tag{5}$$

[14]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, 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, 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

$$w(z; \tilde{\mathbf{z}}) = V(z) - \beta \int_0^\infty V(z')\, dG(z'|z, \tilde{\mathbf{z}}), \tag{6}$$

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

$$w(z; \tilde{\mathbf{z}}) - w(z; \tilde{\mathbf{z}}') = -\beta \left[\int_0^\infty V(z')\, dG(z'|z, \tilde{\mathbf{z}}) - \int_0^\infty V(z')\, dG(z'|z, \tilde{\mathbf{z}}') \right]. \tag{7}$$

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 still attract the worker. In this sense, wages incorporate compensating differentials in learning.

3.3. *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,

$$B(z) = m \int_a N(\mathbf{z}(a), z)\, dA(a), \tag{8}$$

where m denotes the mass of firms in the economy.

Free entry requires that

$$\int_a [\pi(a) - c]\, dA(a) = 0. \tag{9}$$

3.4. *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 $\tilde{\mathbf{z}}$.[15] 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

[15]It is possible the equilibrium is such that many teams contain workers with knowledge z. In that case, it must be that multiple teams of coworkers maximize (5). In the special case in which the solution to the maximization in (5) is unique for each x, 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 (10) would be $\int_x G(z|x, \tilde{\mathbf{z}}(x))\, dB(x)$.

those whose interactions with coworkers leaves them with knowledge weakly less than z. Namely,

$$B(z) = \delta B_0(z) + (1-\delta)\int_x \int_{\tilde{\mathbf{z}}\in\tilde{\mathbf{Z}}} G(z|x,\tilde{\mathbf{z}})\,dO(\tilde{\mathbf{z}}|x)\,dB(x). \tag{10}$$

3.5. *Equilibrium*

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

1. V and w satisfy (5) and (6);
2. $\mathbf{z}(a)$ and $\pi(a)$ solve (4), namely, the choice of a team of workers maximizes the profit for a firm with technology a taking the wage schedule as given;
3. the labor market clears for each z, so (8) is satisfied;
4. the free entry condition (9) holds;
5. the law of motion for B in (10) is satisfied.

3.6. *Characterizing Equilibrium*

The methodology we use below in Section 4 requires the value function $V(z)$ to be strictly increasing. Here we provide one set of conditions that are sufficient to give rise to this property. In particular, we 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 $G(z'|z, \tilde{\mathbf{z}}_1) > G(z'|z, \tilde{\mathbf{z}}_2)$ for any z, z', and $z_1 < z_2$ implies that $G(z'|z_1, \tilde{\mathbf{z}}) > G(z'|z_2, \tilde{\mathbf{z}})$ for any $\tilde{\mathbf{z}}$, z'.*

ASSUMPTION 3: *There is free disposal of knowledge.*

The first assumption implies that 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 $(z_1, \tilde{\mathbf{z}}) = \mathbf{z}(a)$. Then for each $z_2 > z_1$, it must be that $w(z_2, \tilde{\mathbf{z}}) > w(z_1, \tilde{\mathbf{z}})$.*

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 $z_2 > z_1$ such that $w(z_2, \tilde{\mathbf{z}}) \le w(z_1, \tilde{\mathbf{z}})$. Then the firm should hire z_2 instead of z_1. 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. *Q.E.D.*

PROPOSITION 1: *$V(z)$ is 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 coworkers of z_1 in at least one such firm. Then this, along with Lemma 1, implies

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

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. *Q.E.D.*

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 (7). *Q.E.D.*

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

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

where the first inequality follows from Lemma 1 and the second inequality follows from Proposition 2. *Q.E.D.*

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

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

Optimality implies

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

The marginal cost to a firm of having its ith 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 (6) must hold for any $\tilde{\mathbf{z}}$, we can differentiate with respect to coworker i's knowledge to get

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

We can thus write the optimal condition for the firm as

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

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.

Our theory of learning in a competitive labor market provides a decomposition of the wage of a worker into two components, the competitive value they obtain in the market as a function of their knowledge and a compensating differential for learning. Hence, conditional on the team, a worker's wage is increasing in her knowledge. Furthermore, otherwise identical workers in teams with more knowledgeable coworkers earn less, since they learn more and so can expect larger future increases in wages. In the next section, we use the structure of the model to identify individual worker knowledge using data on wages and team composition which, in turn, allows for the estimation of the learning function, $G(z'|z, \tilde{\mathbf{z}})$.

4. STRUCTURAL ESTIMATION

We now turn to a structural estimation of the amount of learning within teams using the theory we developed in the previous section. One of the key problems interpreting the results in Section 2 is there is no one-to-one mapping between wages and 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 3 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 include teams' wages, and implement our strategy using the German data.

Heuristically, our method exploits two dimensions of the data. First, it uses the within-team distribution of wages observed in repeated cross-sections to back out the worker types z which are consistent with a particular $G(\cdot)$. The distribution of wages among team members at a point in time, together with a given learning function $G(\cdot)$, allows us to infer compensating differentials for learning and, therefore, the level of knowledge of those team members. Second, it uses the intertemporal dimension of the resulting panel of worker types to estimate $G(\cdot)$. That is, the distribution of z_i', given z_i and z_{-i}, identifies the learning function $G(\cdot)$.

4.1. *Identifying Learning Parameters*

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

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

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).[16] 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 which guarantee that $V(z)$ is increasing (e.g., Assumptions 1 to 3). The set of technologies and firms in the economy determine the set of teams we observe in equilibrium, but our strategy simply uses the set of observed teams.

We first note that z does not have a natural cardinality. We are therefore free to choose a convenient one: If $V(z)$ is the value function in the current equilibrium, we choose a cardinality of z so that $V(z) = z$. Then, (11) becomes

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

or

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

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

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

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

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

[16]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.

Formally, if a team has n workers, then given θ and w, (12) provides n equations for the n unknowns of $\{z_i\}$. Therefore, given the wages w_t and a vector of team assignments r_t, we can construct $Z(w_t, r_t, \theta)$ to be the $I \times 1$ vector of all workers' knowledge at t, where I is the number of individuals in the data. Given this, we can construct $M(w_t, r_t, \theta)$ to be the $I \times K$ matrix of moments so that the i, k entry of $M(w_t, r_t, \theta)$ is $m_k(z_i, \tilde{\mathbf{z}}_{-i})$, where z_i, $\tilde{\mathbf{z}}_{-i}$ are the knowledge of i and her coworkers implied by the wages, w_t, the assignment r_t, and parameters θ. Then the k moments conditions (14) can be stacked as

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

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).

1. We first guess parameters θ^{guess}.
2. Given this guess, we can back out the types z in a team solely from information on wages.[17] In other words, we invert (12) to solve for all workers' knowledge, $Z(w_t, r_t, \theta^{\text{guess}})$. We do this by finding a fixed point $\mathbf{z}$ of the operator

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

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

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

4. If our estimated $\hat{\theta} = \theta^{\text{guess}}$, then we have found a fixed point. This fixed point is a solution to (15). If not, we use $\hat{\theta}$ to update our guess and go back to step 1.

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 $G(\cdot)$:

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

where n is the worker's team size and $\Theta(\cdot)$ is a weakly increasing function. Below, we let $\Theta(\cdot)$ be piecewise 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'

[17] As discussed in the theory section above, the vector of types $\mathbf{z}$ is the solution to the firm problem in (4). Here, we simply use the composition of teams observed in the data.

knowledge. This functional form could be motivated by a variant of the model in Lucas (2009) in which a worker is equally likely to attain knowledge from any coworker, and the function Θ describes how the worker's learning depends on the gap between the worker and the coworker. In contrast to Lucas (2009), however, here agents only learn from coworkers, not from the whole population.

We begin by studying the parametric learning function

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

or

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

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 $n-1$.

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 VIII.[18] 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 by 100 euros raises that worker's expected present value of earnings over the next year by 7 to 9 euros times the share of more-highly-paid workers. In turn, doing so for the coworkers that are less well paid only

TABLE VIII

PARAMETRIC ESTIMATION RESULTS FOR THE LEARNING FUNCTION (17)[a]

	Team Definition	
	1	2
θ^+	0.0673 (0.0004)	0.0882 (0.0006)
θ^-	0.0111 (0.0003)	0.0370 (0.0004)
θ^0	0.0039 (0.00004)	0.0060 (0.00003)
Observations	4,763,089	4,590,120

[a]GMM standard errors in parentheses.

[18]The only other parameter we need to choose is β, which we set to 0.95 (annual) here. Our results are not particularly sensitive to this choice.

increases expected present value of earnings by 1 to 4 euros times the share of less-well-paid workers. This implies that learning is not very sensitive to the knowledge of less-well-paid coworkers. Furthermore, since both θ^+ and θ^- are positive, 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, and since $\theta^+ > \theta^-$, those at the bottom are more sensitive to the knowledge of their coworkers. Naturally, we find somewhat larger effects for the narrower team definition.[19] Clearly, these point 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 data set was very limited, as discussed in Appendix B.

Before we turn to richer specification, we briefly verify that the vector $\{z\}_t$ we construct indeed picks up the expected present value of income. To this end, we select workers that are in one of the teams we observe in 1999 and then are employed full-time in each of the following years until the end of our sample in 2010. We construct their realized present value of income using a 5% discount rate and assuming that their final wage in 2010 extends forever into the future. We then regress the realized present value of income on a constant and the $z_{i,1999}$ constructed in the course of estimating θ. The resulting regression coefficient is 1.05, which is reassuring.

A More Flexible Learning Function. 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

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

where $\mathbf{1}$ denotes the indicator function. This piecewise 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 $\{\theta^0, \theta^-, \theta^+, \theta^{++}\}$.[20] When implementing this learning function, we set $b = 10\%$.

[19] One important observation across all specifications we have worked with is that θ^0 is substantially larger for Team Definition 2.

[20] 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.

TABLE IX

ESTIMATES FOR THE LEARNING FUNCTION (18)[a]

	Team Definition	
	1	2
θ^{+}	0.0844 (0.0011)	0.1091 (0.0009)
θ^{++}	0.0668 (0.0004)	0.0853 (0.0006)
θ^{-}	0.0102 (0.0003)	0.0346 (0.0004)
θ^{0}	0.0038 (0.00004)	0.0058 (0.00003)
Observations	4,763,089	4,590,120

[a]GMM standard errors in parentheses.

The results are reported in Table IX. 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 knowledge of those team members with lower type affects an individual's expected learning 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 knowledge growth) to improving the knowledge of those above in the wage distribution appear to be somewhat larger for those in closer proximity in the distribution of knowledge. Just like in Table VIII, 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 both the basic learning function (17) and the piecewise linear learning function in (18) for both team definitions.

4.2.1. *Investment in Knowledge*

An individual receives compensation in two ways: with wages and with knowledge. We can use our estimated framework to gauge the quantitative importance of coworker learning in the economy by comparing the value of knowledge flows to the value of wage payments. Specifically, we compute the value of the annual flow of knowledge, $\beta(z_i' - z)$, where next period's knowledge z_i' is given by equation (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.

We present the results for both team definitions and both learning functions in Table X. Coworker knowledge flows account for roughly 4–9% of the average flow value workers receive, with the remainder 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 4–9% of their total

TABLE X

COWORKER LEARNING AS A FRACTION OF COMPENSATION[a]

Learning Function	Team Definition 1	Team Definition 2
Basic Learning	8.27%	3.64%
Piecewise Linear Learning	8.68%	4.21%

[a] *Notes*: First row refers to results for learning function (17) and second row to results for learning function (18).

compensation in learning from others at work.[21] While the results are fairly similar across the two different specifications, they are substantially larger for Team Definition 1. The reason is simply that there is more within-team knowledge dispersion with Team Definition 1: Workers do not only learn from their coworkers in the same occupation but also from everyone else in the establishment.

Naturally, there is substantial heterogeneity in this breakdown across the knowledge distribution. For the basic learning function and Team Definition 1, knowledge flows amount to 16.0% 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 −0.02%, reflecting the mild negative effect of having mostly less knowledgeable coworkers. Naturally, this is somewhat less pronounced for the second team definition where the bottom decile receives 6.2% of their flow compensation in terms of learning, dropping to −1.2% at the top. The numbers are very similar for the piecewise linear learning function.

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, 2009. Specifically, we fix the vector of estimated $z_{i,09}$ and then randomly reallocate the existing 2009 team IDs. This gives each worker a random set of coworkers from the population of workers while keeping the team size distribution unaltered. We then compute, for all workers, the resulting counterfactual conditional expectation $\mathcal{E}(z_i, \tilde{\mathbf{z}}^{\text{cf}}_{-i})$ for her type in 2010, $z_{i,10}$, where $\tilde{\mathbf{z}}^{\text{cf}}_{-i}$ is worker i's counterfactual peer group in 2009.

We then contrast the average of the counterfactual conditional expectation with the average of the factual conditional expectation, $\mathcal{E}(z_i, \tilde{\mathbf{z}}_{-i})$. We report the results in Table XI which shows that, under random sorting, the average growth in z rises between 64% and 91%. We highlight that the associated losses in the value of output must weakly exceed these knowledge gains in an equilibrium allocation. The results are naturally larger for Team Definition 2 because within-team knowledge dispersion is smaller in more narrowly defined teams.

[21] 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 substantially, to approximately 17% in all four cases.

TABLE XI

INCREASE IN KNOWLEDGE GROWTH FROM RANDOM ASSIGNMENT[a]

Learning Function	Team Definition 1	Team Definition 2
Basic Learning	64.91%	91.14%
Piecewise Linear Learning	64.48%	88.07%

[a]*Notes*: First row refers to results for learning function (17) and second row to results for learning function (18).

These findings suggest that 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.

That being said, there could be reasons for assortative matching aside from production complementarities that are outside of the model. It could be, for example, that geographic or other frictions in team formation lead to the assortative matching patterns.

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 the 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 unconditional variance of log wages in our data is approximately 0.13, while the unconditional variance in log compensation is approximately 0.10 or 0.09, depending on the team definition (and similar for both learning functions).

Table XII reports the difference between the variance of log wages and the variance in log compensation as a fraction of the variance of log wages. By this measure, inequality in

TABLE XII

DECLINE IN INEQUALITY WHEN TAKING LEARNING INTO ACCOUNT[a]

Learning Function	Team Definition 1	Team Definition 2
Basic Learning	−33.47%	−19.56%
Piecewise Linear Learning	−33.53%	−19.70%

[a]*Notes*: We first compute the variance of log wages and then the variance of log overall compensation. We report the percentage decline when moving from the first to the second measure of inequality. First row refers to results for learning function (17); second row for learning function (18).

compensation is one-fifth to one-third smaller than wage inequality. Note that the result that compensation inequality is smaller is not mechanical; it reflects the fact that those with less knowledge receive a larger share of their compensation in the form of learning.

4.2.4. *The Role of Firm Size*

Thus far, we have imposed that learning is independent of team size. To investigate the role of firm size, we now modify the basic learning function (17) to allow for size effects. We let

$$E\left[z_i' - z_i | z_i, \tilde{\mathbf{z}}_{-i}\right] = \theta^0 z_i + \frac{1}{(n-1)^k} \left\{ \theta^- \sum_{z_j < z_i} (z_j - z_i) + \theta^+ \sum_{z_j \geq z_i} (z_j - z_i) \right\}. \tag{19}$$

Specifically, we estimate $\theta = \{\theta^+, \theta^-, \theta^0\}$ separately for each $k \in \{0.6, 0.65, 0.7, 0.75, 0.8, 0.9, 1, 1.1, 1.2\}$. Note that $k < 1$ ($k > 1$) implies increasing (decreasing) returns to scale. To evaluate the relative performance of the different specifications we run, after we have found a fixed point, the regression

$$z_{i,t+1} = \beta^0 z_{i,t} + \beta^- \sum_{z_{j,t} < z_{i,t}} \frac{z_{j,t} - z_{i,t}}{(n-1)^k} + \beta^+ \sum_{z_{j,t} \geq z_{i,t}} \frac{z_{j,t} - z_{i,t}}{(n-1)^k} + \varepsilon_{i,t} \tag{20}$$

and compute the associated R^2.

We plot the resulting R^2 relative to the R^2 for the $k = 1$ case in Figure 2. The results suggest that, at the establishment level, coworker learning is indeed independent of size. At the same time, the specification with $k = 0.8$ obtains the best fit for the narrower team definition.[22] This suggests nontrivial returns to scale in learning from coworkers in

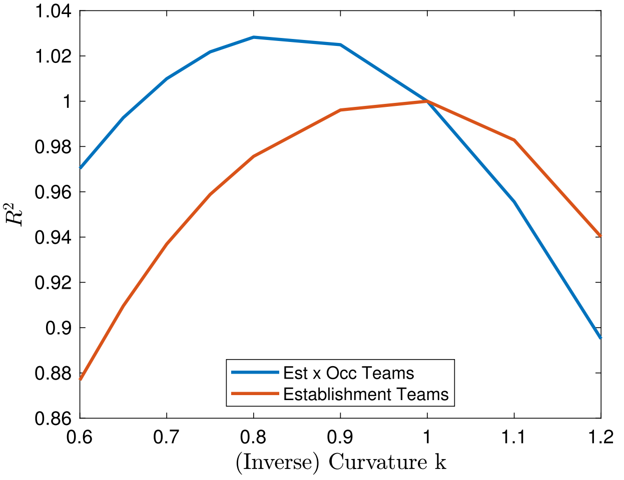

FIGURE 2.—Returns to team size. *Notes*: We plot the R^2 of the regression (20) after the estimation of the learning function (19) for each k in $\{0.6, 0.65, 0.7, 0.75, 0.8, 0.9, 1, 1.1, 1.2\}$. We express the R^2 relative to the resulting R^2 for $k = 1$ for both team definitions.

[22] For this specification, we obtain $\theta^+ = 0.030$ and $\theta^- = 0.012$, smaller than the results for $k = 1$ reported in Table VIII, column 2. However, once we multiply by $\frac{\bar{n}-1}{(\bar{n}-1)^k}$ (with $\bar{n}$ denoting average team size) to offset, in a simple way, for the mechanical effects of the different specification, we obtain $\theta^+ = 0.131$ and $\theta^- = 0.005$.

one's own occupation. We interpret this as reflecting gains from diversity in learning from coworkers in the same occupation.

4.3. *Extensions and Alternative Interpretations*

Our methodology can be extended to a variety of other settings. We can extend our framework to have the production function, the learning function, or the value placed on knowledge depend on observable worker characteristics aside from knowledge. For example, it may be that younger workers' knowledge growth is more sensitive to the knowledge of their peers than older workers. In fact, this extension allows us to relax the baseline assumptions of complete markets as well as perfect information about one's knowledge and the knowledge of coworkers. We can also relax the assumption of perfect competition in the labor market and incorporate search frictions and other adjustment costs. Finally, our methodology can be used in a setting in which knowledge is multidimensional. The common theme across these extensions is that, conditional on the composition of one's team and other observables, there is a bijective mapping from knowledge to values.

Appendix A of the Supplemental Material presents all these extensions of our methodology. As we make clear there, all of the extensions require additional data beyond team composition and wages. We do not currently have the data necessary to estimate learning functions in settings with incomplete markets, search frictions, or multidimensional knowledge. Nevertheless, these extensions should be useful to guide future researchers with access to richer data sets. In Appendix A, we do estimate an extension in which the learning function depends on a worker's age. Naturally, we find that the young learn more overall. Furthermore, we find that both the young and the old learn more from the young, but the discrepancy is starker for the young. That is, the young learn disproportionately from the young, closely in line with our reduced-form findings.

5. CONCLUSION

We presented evidence suggesting learning from coworkers is significant. Our results are intuitive and natural. Workers learn from those more knowledgeable than they are, particularly if they have the same occupation; knowledge growth is more sensitive to more-knowledgeable coworkers than to less-knowledgeable coworkers. Individuals—especially those that are younger and less knowledgeable—invest a substantial fraction of their compensation into knowledge growth. As a result, inequality in wages overstates inequality in total compensation.

We hope that these findings are useful in encouraging more empirical research with learning from coworkers at its core. Our baseline theory, although general in its specification of technology and existing complementarities in production, does assume that workers are simple income maximizers and that labor markets are competitive. We relax some of these assumptions in the final section, but do not have enough data to incorporate them in our estimation. It would be valuable to refine our estimates of the value of learning in the workplace with richer data sets that implement these extensions.

Finally, the importance of learning from coworkers implied by our findings suggests large aggregate 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 skill-biased technical change, as well as 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. Doing so would, however, require specifying and estimating a production function, which the rest of our analysis does not require. We therefore leave these exercises for future research.

REFERENCES

AKCIGIT, U., S. CAICEDO, E. MIGUELEZ, S. STANTCHEVA, AND V. STERZI (2018): "Dancing With the Stars: Innovation Through Interactions," Technical Report, National Bureau of Economic Research. [650]

ALVAREZ, F. E., F. J. BUERA, AND R. E. JR. LUCAS (2013): "Idea Flows, Economic Growth, and Trade," Technical Report, National Bureau of Economic Research. [649]

ANDERSON, A., AND L. SMITH (2010): "Dynamic Matching and Evolving Reputations," *The Review of Economic Studies*, 77 (1), 3–29. [650]

BUERA, F. J., AND E. OBERFIELD (2020): "The Global Diffusion of Ideas," *Econometrica*, 88 (1), 83–114. [649]

BURDETT, K., AND M. COLES (2003): "Equilibrium Wage-Tenure Contracts," *Econometrica*, 71 (5), 1377–1404. [660]

BURSTEIN, A. T., AND A. MONGE-NARANJO (2009): "Foreign Know-How, Firm Control, and the Income of Developing Countries," *The Quarterly Journal of Economics*, 124 (1), 149–195. [650]

CAICEDO, S., R. E. JR. LUCAS, AND E. ROSSI-HANSBERG (2019): "Learning, Career Paths, and the Distribution of Wages," *American Economic Journal: Macroeconomics*, 11 (1), 49–88. [650]

CORNELISSEN, T., C. DUSTMANN, AND U. SCHÖNBERG (2017): "Peer Effects in the Workplace," *American Economic Review*, 107 (2), 425–456. [650]

GARICANO, L., AND E. ROSSI-HANSBERG (2006): "Organization and Inequality in a Knowledge Economy," *The Quarterly Journal of Economics*, 121 (4), 1383–1435. [650]

HERKENHOFF, K., J. LISE, G. MENZIO, AND G. PHILLIPS (2018): "Worker Mobility and the Diffusion of Knowledge," Technical Report. [650]

JÄGER, S., AND J. HEINING (2019): "How Substitutable Are Workers? Evidence From Worker Deaths," Technical Report. [661]

JAROSCH, G., E. OBERFIELD E. ROSSI-HANSBERG (2021): "Supplement to 'Learning From Coworkers'," *Econometrica Supplemental Material*, 89, https://doi.org/10.3982/ECTA16915. [648,650]

JOVANOVIC, B. (2014): "Misallocation and Growth," *American Economic Review*, 104 (4), 1149–1171. [650]

JOVANOVIC, B., AND G. M. MACDONALD (1994): "Competitive Diffusion," *Journal of Political Economy*, 102 (1), 24–52. [649]

JOVANOVIC, B., AND R. ROB (1989): "The Growth and Diffusion of Knowledge," *Review of Economic Studies*, 56 (4), 569–582. [649]

KÖNIG, M. D., J. LORENZ, AND F. ZILIBOTTI (2016): "Innovation vs. Imitation and the Evolution of Productivity Distributions," *Theoretical Economics*, 11 (3), 1053–1102. [649]

LUCAS, R. E. (2009): "Ideas and Growth," *Economica*, 76 (301), 1–19. [649,669]

LUCAS, R. E., AND B. MOLL (2014): "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, 122 (1), 1–51. [649]

LUTTMER, E. G. J. (2014): "An Assignment Model of Knowledge Diffusion and Income Inequality," Federal Reserve Bank of Minneapolis Research Department Staff Report 715. [649]

MAS, A., AND E. MORETTI (2009): "Peers at Work," *American Economic Review*, 99 (1), 112–145. [650]

NIX, E. (2015): "Learning Spillovers in the Firm," Technical Report. [650]

PERLA, J., AND C. TONETTI (2014): "Equilibrium Imitation and Growth," *Journal of Political Economy*, 122 (1), 52–76. [649]

PERLA, J., C. TONETTI, AND M. E. WAUGH (2021): "Equilibrium Technology Diffusion, Trade, and Growth," *American Economic Review*, 111 (1), 73–128. [649]

POSTEL-VINAY, F., AND J.-M. ROBIN (2002): "Equilibrium Wage Dispersion With Worker and Employer Heterogeneity," *Econometrica*, 70 (6), 2295–2350. [660]

SAMPSON, T. (2015): "Dynamic Selection: An Idea Flows Theory of Entry, Trade and Growth," *The Quarterly Journal of Economics*. [649]

Co-editor Fabrizio Zilibotti handled this manuscript.

Manuscript received 19 December, 2018; final version accepted 23 September, 2020; available online 25 September, 2020.