Econometrica Supplementary Material

SUPPLEMENT TO "LEARNING FROM COWORKERS"

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

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

APPENDIX A: MODEL EXTENSIONS AND ADDITIONAL INTERPRETATIONS

IN THIS SECTION, we develop a number of extensions. We first let the production function, the learning function, or the value placed on knowledge depend on observable worker characteristics aside from knowledge. We then relax the baseline assumptions of complete markets, perfect information about one's knowledge and the knowledge of coworkers, perfect competition in the labor market, and incorporate search frictions and other adjustment costs. Finally, we show that the methodology can be used in a setting in which knowledge is multidimensional. We conclude the section by estimating the learning function in an environment in which a worker's learning depends on her age.

A.1. *Incorporating Other Observables*

We now study a setting in which either the production function or the learning function depends on worker characteristics aside from knowledge. These characteristics may or may not evolve endogenously. We require the characteristics to be observable.

An individual is described by knowledge z and a vector of observable characteristics x. These evolve according to a joint Markov process. For example, x could consist of an individual's age, schooling, occupation, location, etc. Denote the joint Markov process by $G(z', x'|z, x, \{\tilde{\mathbf{z}}, \tilde{\mathbf{x}}\})$. The value function for an individual with state z, x is $V(z, x)$ satisfying

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

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

$$v_{it} = w_{it} + \beta \int \int V\big(z', x'\big) G\big(dz', dx'|z_{it}, x_{it}, \{\tilde{\mathbf{z}}_{it}, \tilde{\mathbf{x}}_{it}\}\big). \tag{A.1}$$

The key step is to transform the learning function from the knowledge space to the value space. We require that V is strictly increasing in z for each x, and therefore has

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

 https://doi.org/10.3982/ECTA16915

a partial inverse $Z(v, x)$ that satisfies $v = V(Z(v, x), x)$ and $z = Z(V(z, x), x)$. Further, define the learning function in the value space as

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

where $\tilde{\mathbf{Z}}(\tilde{\mathbf{v}}; \tilde{\mathbf{x}})$ is the vector of coworkers' knowledge given their values and characteristics. With this, we can write (A.1) as

$$v_{it} = w_{it} + \beta \int \int v' \hat{G}(dv', dx'|v, x, \tilde{\mathbf{v}}, \tilde{\mathbf{x}}). \tag{A.2}$$

A.1.1. *Algorithm*

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

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

so that the realized Bellman equation (A.2) can be written as

$$v_{it} = w_{it} + \beta \mathcal{E}(v_{it}, x_{it}, \tilde{\mathbf{v}}_{-it}, \tilde{\mathbf{x}}_{-it}). \tag{A.3}$$

If we observed $\{v'_i, x'_i, v_i, x_i, \tilde{\mathbf{v}}_{-i}, \tilde{\mathbf{x}}_{-i}\}$ for each worker, we could directly estimate $\mathcal{E}$. Conversely, if we knew $\mathcal{E}$ and we observed $\{w_i, x_i, \tilde{\mathbf{w}}_{-i}, \tilde{\mathbf{x}}_{-i}\}$ for each worker, we could solve for $\{v_i\}$ for each team using the system of equations given by (A.3) for each team member.

We can again cast this in terms of a GMM estimation procedure. For example, suppose we assume $\mathcal{E}$ takes the form of a linear combination of moments $\{m_k(v, x, \tilde{\mathbf{v}}, \tilde{\mathbf{x}})\}_{k=1}^K$, so $\mathcal{E}(v, x, \tilde{\mathbf{v}}, \tilde{\mathbf{x}}) = \sum_{k=1}^K \theta_k m_k(v, x, \tilde{\mathbf{v}}, \tilde{\mathbf{x}})$. Next, define the expectational error term ε_{it+1} to be

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

We can then build moment conditions to estimate θ from $E[\varepsilon_{it+1}|v_{it}, x_{it}, \tilde{\mathbf{v}}_{it}, \tilde{\mathbf{x}}_{it}] = 0$. Our natural moment conditions would then be $E[\varepsilon_{it+1} m_k(v_{it}, x_{it}, \tilde{\mathbf{v}}_{it}, \tilde{\mathbf{x}}_{it})] = 0$. Formally, given θ, we can solve for $\{v_{it}\}$ using (A.3). Given the entire vector of wages w, observable characteristics x, and team assignments r, let $Y(w, x, r, \theta)$ be the corresponding values that have been solved for using (A.3) so that $\{v_{it}\} = Y(w_t, x_t, r_t, \theta)$. Given this, we can construct $M(w_t, x_t, r_t, \theta)$ to be the $I \times K$ matrix of moments so that the i, k entry of $M(w_t, x_t, r_t, \theta)$ is $m_k(v_i, x_i, \tilde{\mathbf{v}}_{-i}, \tilde{\mathbf{x}}_{-i})$, where v_i, $\tilde{\mathbf{v}}_{-i}$ are the values of i and her coworkers implied by the wages, w_t, the observable characteristics x_t, the assignment r_t, and parameters θ. Then the k moment conditions (14) can be stacked as

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

A.1.2. *Incomplete Markets*

In this section, we show that with additional information about workers' assets, one could relax the assumption of complete markets. Suppose an individual faces a consumption-savings problem in which she can only save in risk-free bonds, possibly including a constraint on her asset position. In such a model, an individual with assets b

and knowledge z chooses consumption and a team of coworkers to maximize the present value of utility given by

$$V(b, z) = \max_{c, b', \tilde{\mathbf{z}}} u(c) + \beta E\big[V\big(b', z'\big)|z, \tilde{\mathbf{z}}, b\big],$$

subject to the budget constraint $c = b + w(z, b, \tilde{\mathbf{z}}) - \frac{b'}{1+r}$, borrowing constraints, and the constraint that there is a firm willing to hire that worker along with those coworkers.[1] If, as is natural, the value is increasing in knowledge for any level of assets, the value function has a partial inverse, $Z(v, b)$. We can then express the learning function $G(z'|z, \tilde{\mathbf{z}})$ in value space as $\hat{G}(v'|v, b, \tilde{\mathbf{v}}, \tilde{\mathbf{b}})$.[2] Subsuming the optimal choices, the Bellman equations can be expressed as

$$v_i = u\left(b_i + w_i - \frac{b_i'}{1+r}\right) + \beta E\big[v_i'|v_i, b, \tilde{\mathbf{v}}, \tilde{\mathbf{b}}\big].$$

Thus, with data on worker assets (so that we observe b_i and b_i') and a choice of utility function, we could implement the methodology of Section A.1.1 to estimate $\hat{G}$.

A.1.3. *Imperfect Information*

In this section we discuss a setting in which individuals have imperfect information about their coworkers' knowledge. We discuss two extreme benchmark cases. First, we assume that information about each worker's knowledge is imperfect but all workers and firms share the same beliefs. Second, we assume each individual has perfect information about her own knowledge but no information about the knowledge of her coworkers.

We begin with a model in which beliefs about each worker's knowledge are imprecise but symmetric at all times. Each individual has a knowledge z^* that evolves stochastically over time according to the distribution $G^*(z^*|z^*, \tilde{\mathbf{z}}^*)$. Suppose further that, conditional on observables, beliefs about one's knowledge can be summarized by a single number, z_i. For example, suppose that beliefs were normal, with mean z_i and variance that depended only on age, experience, and tenure. The wage a firm had to pay a worker would depend on the beliefs about her knowledge as well as the compensating differential reflecting how much that worker expected to learn (which depends on beliefs about her coworkers' knowledge). Thus, the wage can be expressed as $w(z, x, \tilde{\mathbf{z}}, \tilde{\mathbf{x}})$, where x_i are the observables that are sufficient to characterize the precision of beliefs. A worker's Bellman equation can be expressed as

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

The true learning function G^*, along with the information generated in the production/learning process, give rise to a Markov process characterizing the evolution of beliefs, $G(z', x'|z, x, \tilde{\mathbf{z}}, \tilde{\mathbf{x}})$. This, along with the Bellman equation, are the same as those in Section A.1, and so we can implement the same approach.

We next turn to a model with asymmetric information. We consider an extreme case in which workers have perfect information about their own type and no information about their coworkers' types.[3] In such an environment, wages would not reflect compensating

[1]The wage schedule would depend on the worker's assets because workers with different levels of assets would place different values on new knowledge.

[2]Formally, $\hat{G}(v'|v, b, \tilde{\mathbf{v}}, \tilde{\mathbf{b}}) \equiv G(Z(v', b'(Z(v, b), b))|Z(v, b), \{Z(v_j, b_j)\}_{j\neq i})$, where $b'(z, b)$ is the continuation assets chosen by the worker.

[3]Here, we do not need to take a stand on firms' beliefs.

differentials for learning. Therefore, an individual's wage would perfectly reveal her type, that is, there would be a one-to-one mapping from wages to knowledge. In this type of environment, the reduced-form approach taken in Section 2 could be used to structurally identify the learning function.

A.1.4. *Search and Rent-Sharing*

In this section, we extend the model to relax the assumption of perfect competition for workers and incorporate search frictions and other forms of adjustment costs. Suppose that a firm's output depends on its productivity and its team of workers according to the production function $F(\mathbf{z}, a)$. Compensation of each worker is determined by multilateral bargaining. We impose two restrictions on the bargaining solution. First, surplus is split among the workers and the firm period by period. Second, the bargaining solution is such that each worker's compensation is a function of her type, the vector of coworkers, and the firm's productivity, $w(z, \tilde{\mathbf{z}}, a)$.[4] The firm solves a dynamic stochastic optimization problem that incorporates vacancy posting costs, worker attrition, and any other adjustment costs such as firing and training costs. The solution induces a Markov transition function that characterizes the (endogenous) law of motion for each worker's knowledge and the set of future coworkers, $G(a', \mathbf{z}'|a, \mathbf{z})$, which incorporates learning, hires, and separations.[5]

The present value of income for each worker and the present value of earnings for the firm would satisfy, respectively,

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

where $\pi(a, \mathbf{z})$ is the current flow profit, incorporating revenue, less the wage bill, less the cost of choosing the probability distribution over next period's team.

The key step in our methodology is to express the learning functions in terms of values. In particular, if we can write the learning functions for a worker and firm as $\hat{G}^w(v'|v, \tilde{\mathbf{v}}, y)$ and $\hat{G}^f(y'|y, \mathbf{v})$, respectively, then we can express the realized Bellman equations as

$$v_i = w_i + \beta E\big[v_i'|v_i, \tilde{\mathbf{v}}_{-i}, y_j\big], \tag{A.4}$$
$$y_j = \pi_j + \beta E\big[y_j'|y_j, \mathbf{v}\big]. \tag{A.5}$$

With these in hand, we can recover the learning functions as follows: If we know $\hat{G}^w$ and $\hat{G}^f$, we can recover the $n+1$ unknowns $\{v_i\}$ for each of the n workers and y_i for the firm with the $n+1$ equations, (A.4) for each of the n workers and (A.5) for the firm. Conversely, if we know v_i, v_i', $\tilde{\mathbf{v}}_{-i}$, and y_j for each worker as well as y_j, y_j', and $\mathbf{v}$ for each firm, we can directly estimate $\hat{G}^w$ and $\hat{G}^f$. We can thus follow the same iterative procedure outlined above to estimate the learning functions.

For this procedure to work, we must be able to invert the vector of value functions. Specifically, given a vector of values $\{\mathbf{v}, y\}$ for a team, we must be able to recover their knowledge $\{\mathbf{z}, a\}$. If static/dynamic peer effects are not too large, then the value functions

[4] Two standard examples are Stole and Zwiebel (1996) and multilateral Nash bargaining in which bargaining weights are either symmetric or depend only on workers knowledge and the firm's productivity.

[5] This specification nests models such as Hopenhayn (1992) in which productivity evolves exogenously as well as those that incorporate R&D.

$\{V(z, \tilde{\mathbf{z}}, a)\}$ and $Y(a, \mathbf{z})$ are jointly invertible because the Jacobian is diagonally dominant.[6]

What exactly would this recover? $E[v'|v, \tilde{\mathbf{v}}, a]$ incorporates information both about learning from others and how others affect the evolution of one's value in other ways (i.e., giving firms incentive to hire new workers that are complementary to you, or separate from them, etc.). Hence, we would obtain more of an estimate of dynamic peer effects that incorporates learning than of purely learning.

To reiterate, the key feature of this economy that makes our methodology feasible is the fact that we can invert value functions and recover knowledge of the team of coworkers. Future research could try to extend this methodology to situations in which a worker's compensation depends on more than the vector of knowledge in her team, for example, on a worker's past outside offers.

A.1.5. *Multidimensional Knowledge*

The equations characterizing an individual's value function use the idea that the change in one's knowledge depends on the distribution of knowledge among one's coworkers. We have proceeded under the presumption that knowledge can be described by a scalar, z. Suppose instead that z represents a vector of individual characteristics (e.g., different dimensions of knowledge). There are some restrictions on the environment and learning process under which our baseline procedure would still be appropriate to identify the learning.[7] Outside of these special cases, it may be possible, with additional data, to allow for more general patterns of learning. To be concrete, suppose that knowledge was two-dimensional, $z_i = \{z_{i1}, z_{i2}\}$, and the learning function took the form $G(z'_{i1}, z'_{i2}|z_{i1}, z_{i2}, \tilde{\mathbf{z}}_{-i1}, \tilde{\mathbf{z}}_{-i2})$. Suppose also that we observed an additional choice of each worker in addition to compensation. Label this choice $x(z_{i1}, z_{i2})$. Finally, suppose that the value function and choice function were jointly invertible, that is, we could express $z_{i1} = Z_1(v_i, x_i)$ and $z_{i2} = Z_2(v_i, x_i)$; then we could use a version of our methodology to recover the law of motion $\hat{G}(v', x'|v, x, \tilde{\mathbf{v}}, \tilde{\mathbf{x}})$. Clearly, this approach would not separately identify how each component of knowledge evolves, but it would recover the value of learning.[8]

[6]To be concrete, suppose the value functions are, in fact, invertible. Let $Z(v, \tilde{\mathbf{v}}, y)$ be the knowledge of a worker with value v and coworkers summarized by $\tilde{\mathbf{v}}$ working for a firm with value y, and let $A(y, \mathbf{v})$ be the productivity of a firm that has value y and a team summarized by the vector of values $\mathbf{v}$. Then $\hat{G}^w(v'|v, \tilde{\mathbf{v}}, y)$ and $\hat{G}^f(y'|y, \mathbf{v})$ are defined as follows. The law of motion G implies a conditional distribution for each worker's knowledge in the next period, $G^w(z'|z, \tilde{\mathbf{z}}, y)$, and a conditional distribution for the firm's productivity $G^f(a'|a, \mathbf{z})$. To derive the expressions for $\hat{G}^w$ and $\hat{G}^f$, we can simply plug $Z(v, \tilde{\mathbf{v}}, y)$ and $A(y, \mathbf{v})$ into these conditional distributions and take expectations, namely,

$$\hat{G}^w(v'_i|v_i, \tilde{\mathbf{v}}_{-i}, y_j) \equiv E_{\tilde{\mathbf{v}}'_{-i}, y'_j}\left[G^w\left(Z(v'_i, \tilde{\mathbf{v}}'_{-i}, y'_j) \,\middle|\, \begin{matrix} Z(v_i, \tilde{\mathbf{v}}_{-i}, y_j), \\ \{Z(v_k, \tilde{\mathbf{v}}_{-k}, y_j)\}_{k \neq i}, A(y_j, \{v_i, \tilde{\mathbf{v}}_{-i}\}) \end{matrix}\right) \middle| v_i, \tilde{\mathbf{v}}_{-i}, y_j, v'_i\right],$$

$$\hat{G}^f(y'|y, \mathbf{v}) \equiv E_{\mathbf{v}'}\left[G^f\left(A(y', \mathbf{v}')|A(y, \mathbf{v}), \{Z(v_i, \tilde{\mathbf{v}}_{-i}, y)\}_i\right)|y, \mathbf{v}, y'\right].$$

[7]For example, learning could be such that the change in one's value function depends on the composition of one's coworkers' value functions. In this case, it is natural to assume that the learning function depends on values, not on knowledge directly, since values provide a relevant summary of the vector of characteristics z. Then, the procedure outlined above to obtain an individual's knowledge from a panel of wages simply recovers the values of all agents, as in equation (11). Under this assumption, our methodology can be used exactly as described, and our quantitative results would be unchanged. We would simply interpret the estimated learning function as determining how the value of other agents determines the change in value of a given individual.

[8]In an environment with rent-sharing, a similar approach can be taken for firms. If firm technology were multidimensional, we could identify those dimensions separately if we observed several firm-level outcomes

A.2. *Age-Dependent Learning: Estimation Results*

We finish this section with an illustration of the methodology in Section A.1 that studies the differences in learning between young and old workers. Let $\{y, o\}$ indicate whether a worker is young or old. We implement the following parametric form for the conditional expectation:

$$E_y\left[v_i' - v_i | v_i, \tilde{\mathbf{v}}_{-i}\right] = \theta_y^0 v_i + \frac{1}{n-1}\left\{\theta_{yy}^+ \sum_{v_j > v_i, j \text{ young}} (v_j - v_i) + \theta_{yo}^+ \sum_{v_j > v_i, j \text{ old}} (v_j - v_i)\right\},$$
$$E_o\left[v_i' - v_i | v_i, \tilde{\mathbf{v}}_{-i}\right] = \theta_o^0 v_i + \frac{1}{n-1}\left\{\theta_{oy}^+ \sum_{v_j > v_i, j \text{ young}} (v_j - v_i) + \theta_{oo}^+ \sum_{v_j > v_i, j \text{ old}} (v_j - v_i)\right\}. \tag{A.6}$$

This specification allows knowledge flows to depend on both the age group of the worker and the age group of her coworkers. For instance, θ_{yo}^+ captures the strength of the knowledge flows from old to young coworkers. Furthermore, the specification allows for age group specific trend growth. For simplicity, we do not allow for any effects of coworkers j with $v_j < v_i$. The rest of the implementation follows exactly the same routine outlined in the previous section.

We present our results in Table A.I. In line with our previous findings, we find little trend growth for either age group. Second, the young learn more overall. Furthermore, we find that both the young and the old learn more from the young. Regardless of the team definition, this discrepancy is starker for the young. That is, the young learn disproportionately from the young, closely in line with our reduced-form findings.

TABLE A.I

ESTIMATES FOR THE LEARNING FUNCTION (A.6)[a]

	Team Definition	
	1	2
Trend growth young: θ_y^0	0.0023 (0.0001)	0.0041 (0.0001)
Learning of young from young: θ_{yy}^+	0.1190 (0.0034)	0.1473 (0.0045)
Learning of young from old: θ_{yo}^+	0.0489 (0.0016)	0.0649 (0.0024)
Trend growth old: θ_o^0	0.0008 (0.0001)	0.0020 (0.0001)
Learning of old from young: θ_{oy}^+	0.0530 (0.0014)	0.0692 (0.0012)
Learning of old from old: θ_{oo}^+	0.0385 (0.0011)	0.0533 (0.0015)
Observations	4,763,089	4,590,120

[a]*Notes*: Old defined as 40 and older. GMM standard errors in parentheses.

---

such as revenue, investment, and R&D spending, and if we could invert the value function to express each dimension of productivity in terms of those outcomes.

APPENDIX B: DATA APPENDIX

We present below summary statistics of the variables of interest in our data set. Supplemental Appendix D describes the exact construction of our data set.

B.1. *Summary Statistics*

We begin by briefly describing the data set along key dimensions. The longitudinal version of the *Linked-Employer-Employee-Data of the IAB (LIAB LM 9310)* contains information on the complete workforce of a subset of German establishments. The sample establishments are the ones selected—at least once—in an annually conducted survey between 2000 and 2008. The employee part of the data set then contains the employment biographies from 1993 to 2010 of all individuals which were, for at least one day, employed at one of the sample establishments between 1999 and 2009.[9] As a consequence, we observe the complete peer groups at the sample establishments from 1999 to 2009.

The employment biographies come in spell format and contain information, among other things, on a worker's establishment, occupation, and average daily earnings along with a rich set of observables (age, gender, job and employment tenure, education, location, among others). We organize the resulting data set as an annual panel. Specifically, the annual observation recorded (employer, average daily wage, etc.) for each individual pertains to the spell which overlaps a particular *reference date* (January 31st).

To construct a baseline sample, we then proceed as follows. We select the panel case establishments for which we obtain information on the full workforce. We then include those individuals who were employed at one of those establishments at the reference date during at least one year between 1999 and 2009. This leaves us with the employment biographies (between 1993 and 2010) of the full workforce (at a reference date) of a large number of establishments. The following subsections contain more detailed information on the construction of the baseline sample.

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

Team Size Distribution. Figure B.1 plots the unweighted size distribution for both team definitions for the year 2000. We restrict attention to teams that have size ≥ 2. The team size distribution is naturally more compressed under the second, narrower, team definition, but for both definitions a sizable fraction of teams are fairly large. The sample contains 4478 establishments with average size 116. When working with the second team definition, we have a total of 28,524 teams with an average size of 18.[10]

Wage Distribution. Figure B.2 plots the histogram of the average daily earnings during the year-2000 reference spell in euros. The "mass-points" reflect top-coding of the earnings data at the social security contribution ceiling (which is lower in Eastern Germany). As a consequence, 7.7% of the wage observations in the regression sample are are top-coded.[11] A simple variance decomposition implies that the within-team component

[9] For more detail, see Klosterhuber, Heining, and Seth (2014).

[10] It follows that, on average, an establishment has 6–7 different occupations with at least two individuals. The occupational classification follows the 1988 classification of occupations (KldB_88) published by the German Federal Employment Agency which has three digits and comprises about 340 values.

[11] While there exist imputation methods to address the truncation, we instead treat the top-coded observations as actual wage observations and do not correct for the top-coding. We have experimented with various

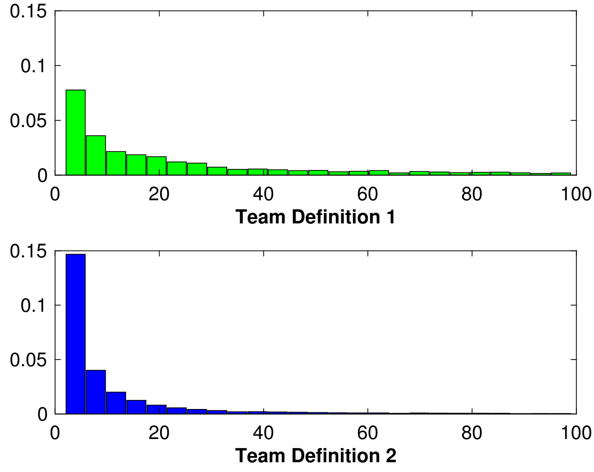

FIGURE B.1.—Team size distribution. *Notes*: Top panel plots the unweighted team size distribution in the year 2000 for teams of size 2–99 for Team Definition 1 (so it corresponds to the establishment size distribution). The bottom panel plots the unweighted team size distribution in the year 2000 for teams of size 299 for Team Definition 2.

accounts for 47.9% of the overall variance in wages under Team Definition 1 and 22.7% under Team Definition 2. Finally, there is fairly little wage growth in the decade covered by our data set. The cohort whose wages are depicted in Figure B.2 experienced average annual wage growth of 0.97% over the next 5 years. The annual growth rate drops to 0.43% in the second half of the decade.

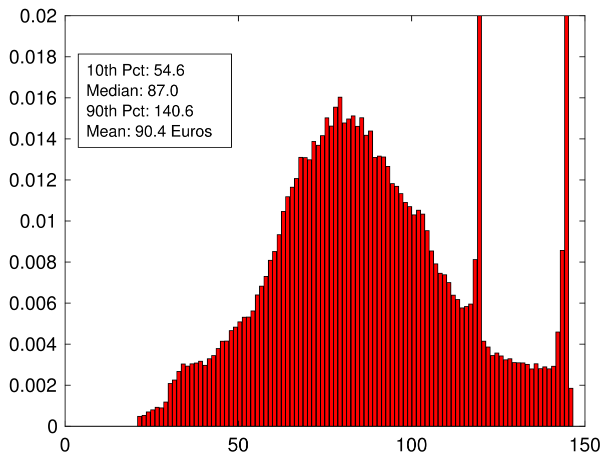

FIGURE B.2.—Wage distribution. *Notes*: Distribution of mean daily wages during spell overlapping 01/31/2000 for full-time employees working subject to social security.

ways of treating the top-coded observations and found our main empirical results to be robust. We show in Appendix C.3 how our reduced-form results change when omitting all teams with top-coded wage observations. For more detail regarding the construction of wages, see Supplemental Appendix D. There, we also describe how we eliminate some extreme wage observations.

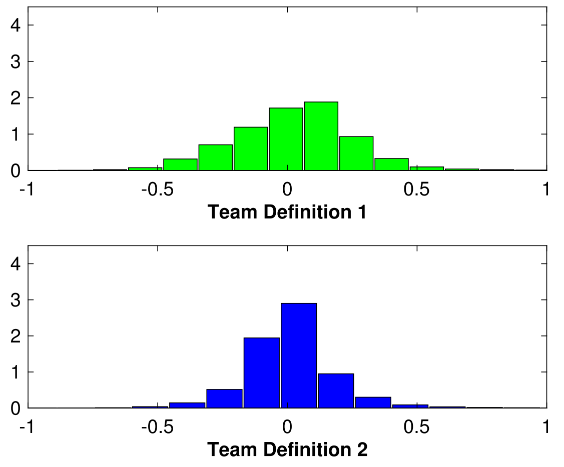

FIGURE B.3.—Wage gap distribution. *Notes*: Top panel plots the distribution of wage gaps as defined in the main text in the year 2000 for Team Definition 1. Bottom panel: Team Definition 2.

Wage Gap to Coworkers. We are interested in how a worker's future wage growth relates to her coworker's (relative) wages. To gauge the extent of wage differences across peers, Figure B.3 plots the histogram of wage gaps, defined as the log difference between an individual's wage and the mean wage of her peers, for each team definition. Under the first team definition, the gap has mean 0.026 and amounts to −0.27, 0.04, and 0.29 at the 10th, 50th, and 90th percentiles in the year 2000. Under the second team definition, the gap has mean 0.012 and is −0.17, 0.00, and 0.21 at the 10th, 50th, and 90th percentiles. Naturally, within-team wage dispersion is smaller under the narrower team definition.

Correlations

We compute a set of correlations of various wage moments at the team level. Specifically, Table B.I reports the correlation matrix of team average pay, team pay dispersion, team mean-median ratio (skewness), team size, and max wage at the team. All entries of the matrix are positive except the correlation between team average pay and the mean-median ratio.[12]

TABLE B.I

PAIRWISE CORRELATIONS AT THE TEAM LEVEL—TEAM DEFINITION 2[a]

	Mean Wage	SD Wage	Mean/Median	Team Size	Max Wage
Mean Wage	1				
Wage SD	0.25	1			
Mean/Median	−0.15	0.09	1		
Team Size	0.09	0.23	0.05	1	
Max Wage	0.92	0.48	0.02	0.28	1

[a] All variables *in logs*.

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

APPENDIX C: ADDITIONAL REDUCED-FORM EMPIRICAL RESULTS

C.1. *An Identification Strategy*

This section develops an instrumental variable (IV) strategy in the spirit of Jäger and Heining (2019), who used arguably random, unexpected worker deaths to study the substitutability of workers. In particular, we identify (pseudo-)random exit events where a full-time employed, prime-age (26–54) worker permanently leaves employment subject to social security. Recall that we observe peer groups from 1999 to 2009 and have individual-level information until 2010. We therefore only study exits that occurred before January 2009 so that we know that the respective worker did not return into employment (including part-time) for a minimum of 24 months (and longer for most). A second, stricter version of the exit instrument applies the additional criterion that the worker was continuously full-time employed during the 24 months before the exit occurred.

We construct the instrument as follows: We identify all of i's peers that exit between year $t-1$ and t, then compute the average wage bill of i's peers in $t-1$ with and without the workers that exit. The log difference is our instrument.

Because we have only one instrument, we modify the specification (1) so that we regress the cumulative growth in one's wages between t and $t+n$, $\Delta w_{i,t}^n$, on the log wage gap to one's peers, $\hat{w}_{i,t} \equiv \log(\frac{\bar{w}_{-i,t}}{w_{i,t}})$, which has only one potentially endogenous regressor:

$$\Delta w_{i,t}^n = \alpha + \beta \hat{w}_{i,t} + \omega_{\text{age}} + \omega_{\text{tenure}} + \omega_{\text{gender}} + \omega_{\text{educ}} + \omega_{\text{occ}} + \omega_t + \varepsilon_{i,t}. \tag{C.1}$$

Further, since our instrument is a shifter of the *change* in one's peers rather than the level, we take first differences of (C.1):

$$D.\Delta w_{i,t}^n = \alpha + \beta D.\hat{w}_{i,t} + \epsilon_{i,t}, \tag{C.2}$$

where $D.$ denotes variables in first differences and $\epsilon_{i,t} \equiv \varepsilon_{i,t} - \varepsilon_{i,t-1}$. $D.\hat{w}_{i,t}$ is the variable we instrument.[13]

The number of exit events is not very large. Thus, while the instrument is strong, the two-stage least squares estimates are less precise than our baseline. Table C.I contains the results for Team Definition 2 for both versions of the instrument along with the estimates of (C.1) and (C.2) using OLS.[14] All point estimates are positive, and rather large, consistent with the view that increases in the wage gap lead to faster wage growth. However, the results are significant at the 5% significance level only for 2- and 3-year horizons. Longer horizons have too few observations.

C.2. *Figure 1*

Table C.II offers the regression output underlying Figure 1. Table C.III then offers the same table for the alternative team definition followed by the results for various restricted samples at horizon $h = 3$.

[13]We drop all fixed effects when differencing. We have experimented with different strategies and found it makes little difference.

[14]The estimates from (C.2) in the second panel are biased because the $\varepsilon_{i,t-1}$ determines $w_{i,t}$, which means that $\epsilon_{i,t}$ is correlated with $\hat{w}_{i,t}$ and hence $D.\hat{w}_{i,t}$. Note that this is not an issue for the IV specifications because the instrument does not contain $w_{i,t}$.

TABLE C.I

IV RESULTS—TEAM DEFINITION 2[a]

	Specification (C.1)			
Horizon in Years	1	2	3	5
$\hat{w}$	0.10 (0.0037)	0.15 (0.0056)	0.19 (0.0076)	0.25 (0.011)

	Specification (C.2)—OLS			
Horizon in Years	1	2	3	5
D.$\hat{w}$	0.97 (0.017)	0.81 (0.015)	0.83 (0.015)	0.86 (0.016)

	Specification (C.2)—First IV			
Horizon in Years	1	2	3	5
D.$\hat{w}$	0.14 (0.26)	0.46 (0.22)	0.50 (0.26)	0.29 (0.26)
C-D F-Stat	1058.8	1052.8	976.6	819.3
K-P F-Stat	42.9	50.2	62.3	64.9
Observations	1,243,851	1,052,910	894,950	621,950

	Specification (C.2)—Second IV			
Horizon in Years	1	2	3	5
D.$\hat{w}$	0.41 (0.32)	0.81 (0.31)	0.96 (0.41)	0.46 (0.46)
C-D F-Stat	828.1	713.4	549.7	345.3
K-P F-Stat	46.5	46.8	43.1	32.7
Observations	1,059,259	899,139	766,043	529,574

[a]*Notes*: Standard errors clustered at the establishment level. Cragg–Donald and Kleibergen–Paap F-statistics. Standard errors in parentheses.

C.3. *Robustness*

This section evaluates the robustness of the main reduced-form empirical results reported in Section 2.1. We do so for Team Definition 2 at the horizon $h = 3$ years and report the corresponding tables for Team Definition 1 below.

To do so, we begin by contrasting our baseline results for specification (2) when omitting teams that have any apprentices.[15] We then restrict the sample exclusively to teams without any top-coded wage observations.[16] We report the corresponding results, contrasted with our benchmark results, for $h = 3$ in Table C.IV and, for the alternative team

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

[16]The ceiling varies from year to year and differs between former Eastern and Western Germany. Furthermore, the data display a certain amount of bunching in a small interval around the officially reported ceiling

TABLE C.II

RESULTS FOR FIGURE 1 USING SPECIFICATION (3)[a]

	Horizon in Years				
	1	2	3	5	10
Bin 2	0.000088 (0.000039)	0.00024 (0.000063)	0.00036 (0.000085)	0.00047 (0.00011)	0.00100 (0.00026)
Bin 3	0.000023 (0.000033)	0.000091 (0.000049)	0.00018 (0.000067)	0.00025 (0.00011)	0.00082 (0.00021)
Bin 4	−0.000062 (0.000030)	−0.00000048 (0.000049)	0.000075 (0.000064)	0.00013 (0.000098)	0.00048 (0.00020)
Bin 5	−0.000079 (0.000029)	−0.000013 (0.000049)	0.000088 (0.000066)	0.00019 (0.00010)	0.00075 (0.00022)
Bin 6	0.000049 (0.000023)	0.00015 (0.000039)	0.00026 (0.000054)	0.00039 (0.000090)	0.00079 (0.00019)
Bin 7	0.00025 (0.000026)	0.00036 (0.000045)	0.00049 (0.000066)	0.00065 (0.00010)	0.0011 (0.00022)
Bin 8	0.00037 (0.000030)	0.00051 (0.000047)	0.00068 (0.000068)	0.00085 (0.00011)	0.0014 (0.00023)
Bin 9	0.00042 (0.000040)	0.00057 (0.000061)	0.00073 (0.000084)	0.00092 (0.00012)	0.0014 (0.00025)
Bin 10	0.00040 (0.000042)	0.00051 (0.000069)	0.00066 (0.000098)	0.00080 (0.00015)	0.0011 (0.00025)
Bin 11	0.00081 (0.000047)	0.0012 (0.000076)	0.0016 (0.00010)	0.0023 (0.00016)	0.0033 (0.00029)
Within R^2	0.88	0.81	0.76	0.66	0.46
Observations	3,354,925	2,942,967	2,537,838	1,846,999	436,728

[a]*Notes*: Each row reports the coefficient on the weight of bins 2 through 11 where the weight on the bottom bin is the omitted category. Each column corresponds to one line in Figure 1. 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.

definition, in Supplementary Appendix E. While the results vary across the samples, the main takeaway from our reduced-form exercises is robust.

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

levels. To identify workers with top-coded wages, we thus simply group workers into 50 euro-cent wide bins in each year and flag the two bins with the most mass.

TABLE C.III

RESULTS FROM SPECIFICATION (3) UNDER TEAM DEFINITION 2 FOR VARIOUS RESTRICTED SAMPLES[a]

	All	Above Team-Median	Below Team-Median	2nd Pct.	4th Pct.	7th Pct.	9th Pct.
Bin 2	0.00036	0.0010	0.00028	0.00062	−0.00016	0.0016	0.00016
	(0.000085)	(0.00036)	(0.000071)	(0.00030)	(0.00044)	(0.00036)	(0.00013)
Bin 3	0.00018	0.0013	0.00015	0.00012	0.00054	0.00038	0.000060
	(0.000067)	(0.00049)	(0.000058)	(0.00030)	(0.00035)	(0.00018)	(0.00016)
Bin 4	0.000075	0.0011	0.00011	0.00029	0.00050	0.00067	−0.000041
	(0.000064)	(0.00040)	(0.000062)	(0.00025)	(0.00034)	(0.00023)	(0.00011)
Bin 5	0.000088	0.00057	0.00021	0.00055	0.00028	0.00051	0.00017
	(0.000066)	(0.00039)	(0.000060)	(0.00025)	(0.00032)	(0.00023)	(0.00010)
Bin 6	0.00026	0.00087	0.00044	0.00063	0.00041	0.00088	0.00028
	(0.000054)	(0.00037)	(0.000058)	(0.00023)	(0.00034)	(0.00022)	(0.00011)
Bin 7	0.00049	0.0011	0.00058	0.00093	0.00078	0.0010	0.00068
	(0.000066)	(0.00038)	(0.000074)	(0.00024)	(0.00033)	(0.00020)	(0.000098)
Bin 8	0.00068	0.0013	0.00075	0.00090	0.00060	0.0012	0.00098
	(0.000068)	(0.00037)	(0.00012)	(0.00024)	(0.00034)	(0.00021)	(0.00010)
Bin 9	0.00073	0.0014	0.00044	0.0011	0.0010	0.0012	0.0013
	(0.000084)	(0.00038)	(0.00011)	(0.00025)	(0.00045)	(0.00021)	(0.00011)
Bin 10	0.00066	0.0014	0.00033	0.00060	0.0013	0.0016	0.0079
	(0.000098)	(0.00038)	(0.00014)	(0.00029)	(0.00038)	(0.00020)	(0.0037)
Bin 11	0.0016	0.0024	0.00020	0.0017	0.0015	0.0017	0
	(0.00010)	(0.00040)	(0.00016)	(0.00029)	(0.00035)	(0.00025)	(0)
Within R^2	0.76	0.71	0.78	0.091	0.048	0.061	0.17
Observations	2,537,838	1,307,463	1,230,371	235,351	245,260	244,998	244,236

[a]*Notes*: Each row reports the coefficient on the weight of bins 2 through 11 where the weight on the bottom bin is the omitted category. Columns 2 and 3 report the results when the sample is restricted to workers above (below) the team median wage. The remaining columns restrict the sample to workers from particular parts of the wage distribution. Team Definition 2. Column titles indicate horizon 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.

TABLE C.IV

SUBSAMPLES—TEAM DEFINITION 2[a]

	Baseline	Teams w/o Apprentices	Teams w/o Top-Coded Wages	Before 2005	After 2004
$\bar{w}^+$	0.16	0.16	0.12	0.17	0.16
	(0.015)	(0.015)	(0.0084)	(0.017)	(0.015)
$\bar{w}^-$	0.041	0.054	0.056	0.051	0.028
	(0.0081)	(0.0069)	(0.0069)	(0.0096)	(0.0093)
Within R^2	0.76	0.76	0.73	0.76	0.77
Observations	2,617,097	1,667,028	1,319,836	1,825,318	791,774

[a]*Notes*: $\hat{\beta}^+$ and $\hat{\beta}^-$ as estimated from specification (2). Team Definition 2. Column (1): Baseline. Column (2): Sample restricted to teams without workers in apprenticeship. Column (3): Sample restricted to teams without top-coded wages. Columns (4) and (5) restrict split sample by years. 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 C.V

COLLECTIVE BARGAINING—TEAM DEFINITION 2[a]

	All 2000	>10% CB	No CB, No Benchmarking
$\bar{w}^+$	0.14 (0.021)	0.14 (0.024)	0.11 (0.051)
$\bar{w}^-$	0.044 (0.013)	0.047 (0.014)	0.027 (0.027)
Within R^2	0.75	0.75	0.67
Observations	336,073	275,890	10,591

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

Finally, we offer results for various specifications where we include higher-dimensional fixed effects. In particular, we extend the baseline specification (2) in three different ways. First, we include additional fixed effects for establishment (establishment $\times$ occupation in the second team definition). In a second specification, we include additional team fixed effects. In a third specification, we include occupation $\times$ year fixed effects.

In doing so, we replace the separate right-hand-side variables $\bar{w}^+$ and $\bar{w}^-$ in specification (2) with the gap $\bar{w}^+ - \bar{w}^-$. The reason is the following: One's own wage, those of higher paid teammates, and those of lower paid teammates are approximately (but not exactly) collinear with a team fixed effect. If we include a team fixed effect, we cannot recover both of β^+ and β^-, but we can recover their difference. The results are reported in Table C.VI.

We complement this with an additional set of results that add, to the baseline specification (2), five lags of an individual's log wage, w_i. We view this as a way of informing a worker fixed effect in a backward-looking way that is not affected by her contemporaneous peers. The results, albeit smaller, line up closely with our baseline findings. We present the results in Table C.VII.

TABLE C.VI

FIXED EFFECTS[a]

	Baseline	Est × Occ FE	Team	Occ × Yr
$\bar{w}^+ - \bar{w}^-$	0.057 (0.0079)	0.087 (0.010)	0.12 (0.012)	0.060 (0.0079)
Within R^2	0.76	0.47	0.51	0.76
Observations	2,617,097	2,614,229	2,585,060	2,617,037

[a]*Notes*: We replace the separate right-hand-side variables $\bar{w}^+$ and $\bar{w}^-$ in specification (2) with the gap $\bar{w}^+ - \bar{w}^-$. Team Definition 2, horizon $h = 3$. Column (1): Baseline. Column (2): Baseline plus establishment $\times$ occupation fixed effects. Column (3): Baseline plus establishment $\times$ occupation $\times$ year fixed effects. Column (4): Baseline plus occupation x year fixed effects. The regressions include current wage and fixed effects for age decile, tenure decile, gender, education, occupation, and year. Standard errors in parentheses.

TABLE C.VII

ADDITIONAL LAGS[a]

	Horizon in years			
	1	2	3	5
$\bar{w}^+$	0.045 (0.0062)	0.064 (0.0096)	0.093 (0.013)	0.10 (0.018)
$\bar{w}^-$	0.016 (0.0043)	0.025 (0.0069)	0.038 (0.0091)	0.057 (0.013)
Within R^2	0.90	0.85	0.81	0.73
Observations	1,075,048	872,091	663,704	291,095

[a]*Notes*: Baseline specification (2) with five additional lags of log wage. Standard errors in parentheses.

Table C.VIII contains results for specification (2) when we control for establishment-level growth in two different ways. Our first measure of establishment growth is the annual growth rate of the number of full-time employees in our data set. The second is the annual growth rate of the wage bill of the full-time employed. We then add to the otherwise unchanged specification (2) three linear controls, namely, the respective growth measure between $t-2$ and $t-1$, $t-1$ and t, and t and $t+1$. The main takeaway is that the results hardly change when we add these controls.

TABLE C.VIII

COUNTERPART TO TABLE II WITH CONTROLS FOR ESTABLISHMENT-LEVEL WAGE BILL (OR EMPLOYMENT) GROWTH—TEAM DEFINITION 2[a]

	Wage Bill Growth Controls			
Horizon in Years	1	2	3	5
$\bar{w}^+$	0.090 (0.0076)	0.12 (0.011)	0.17 (0.015)	0.24 (0.021)
$\bar{w}^-$	0.028 (0.0056)	0.033 (0.0069)	0.042 (0.0082)	0.060 (0.011)
Within R^2	0.88	0.82	0.77	0.67

	Employment Growth Controls			
Horizon in Years	1	2	3	5
$\bar{w}^+$	0.090 (0.0077)	0.12 (0.011)	0.17 (0.015)	0.24 (0.021)
$\bar{w}^-$	0.028 (0.0057)	0.033 (0.0070)	0.042 (0.0083)	0.060 (0.011)
Within R^2	0.88	0.82	0.77	0.67
Observations	2313,583	2,186,058	1,865,434	1,244,737

[a]*Notes*: Additional controls: Growth in total wage bill (or number) of the full-time employed at the establishment between $t-2$ and $t-1$, $t-1$ and t, and t and $t+1$. Standard errors in parentheses.

REFERENCES

HOPENHAYN, H. A. (1992): "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica: Journal of the Econometric Society*, 1127–1150. [4]

JÄGER, S., AND J. HEINING (2019): "How Substitutable Are Workers? Evidence From Worker Deaths," Technical Report. [10]

KLOSTERHUBER, W., J. HEINING, AND S. SETH (2014): "Linked-Employer-Employee-Data From the IAB: LIAB Longitudinal Model 1993–2010 (LIAB LM 9310)," FDZ Datenreport, Documentation on Labour Market Data 201308_en, Institut fuer Arbeitsmarkt- und Berufsforschung (IAB) February 2014. [7]

STOLE, L. A., AND J. ZWIEBEL (1996): "Intra-Firm Bargaining Under Non-Binding Contracts," *The Review of Economic Studies*, 63 (3), 375–410. [4]

Co-editor Fabrizio Zilibotti handled this manuscript.

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