PEER EFFECTS IN THE WORKPLACE

Thomas Cornelissen, Christian Dustmann, Uta Schönberg

ONLINE APPENDIX

Table of Contents

Appendix A: Model Details

A.1 Assumptions on *m*

We impose two bounds on m in the peer pressure function P(.), which can be thought of as the "pain" from working in a high pressure environment.

First, like Barron and Gjerde (1997), we assume that m is large enough so that the total cost from peer pressure is increasing in peer quality on average in the peer group. This assumption captures workers' dislike of working in a high-pressure environment and is a sufficient, albeit not necessary, condition to ensure that peer effects in productivity lead to peer effects in wages. Inspection of b^* in equation (A.6) reveals that $b^* \leq 1$if $\frac{1}{N}\sum_i \frac{\partial P(e_i,\bar{y}_{\sim i})}{\partial \bar{y}_{\sim i}}\Big|_{\text{optimal}} \frac{\partial \bar{e}^*_{\sim i}}{\partial b} \geq 0, \quad \Leftrightarrow \frac{1}{N}\sum_i \lambda^P (m - e_i^*) \frac{\partial \bar{e}^*_{\sim i}}{\partial b} \geq 0.$ Expressed verbally, the derivative of the cost of peer pressure should be non-decreasing in peer quality on average when the average is weighted by $\frac{\partial \bar{e}^*_{\sim i}}{\partial b}$. If this condition does not hold, then the firm can lower its wage cost by increasing b^* higher than one, because then workers on average will like the additional peer pressure created by their peers' higher effort and be willing to forgo wages to enjoy it. Our assumption rules this case out. The lower bound for m is thus implicitly defined by

$$\frac{1}{N}\sum_i \lambda^P (m^{lower} - e_i^*) \frac{\partial \bar{e}^*_{\sim i}}{\partial b} = 0.$$

Second, we require an upper bound for m to ensure that the combined disutility from the direct cost of effort $C(e_i)$ and peer pressure $P\big(e_i, \bar{f}_{\sim i}\big)$ increases on average in the effort of individual workers in the peer group, i.e., $\frac{1}{N}\sum_i \frac{\partial [C(e_i)+P(e_i,\bar{y}_{\sim i})]}{\partial e_i} = \frac{1}{N}\sum_i [2ke_i^* - \lambda^P \bar{y}_{\sim i}] > 0$ or equivalently, $(2k - \lambda^P)\bar{e}^* - \lambda^P \bar{a} > 0$. Substituting $\bar{e}^* = \frac{1}{N}\sum_i e_i^* = \frac{b^* + b^*\lambda^K \bar{a} + \lambda^P \, \bar{a}}{2k - \lambda^P}$, obtained from the optimal effort levels e_i^* derived in Section A.2 below, gives

$$b^* + b^*\lambda^K \bar{a} + \lambda^P \, \bar{a} - \lambda^P \bar{a} > 0$$

$$b^* + b^*\lambda^K \bar{a} > 0$$

$$b^* > 0,$$

implying that only values of *m* that lead to a positive b^* can satisfy this condition. Using b^* derived in equation (A.6) below, the upper bound for *m* is implicitly defined by

$$\frac{\sum_i \frac{\partial e_i^*}{\partial b}(1+\lambda^K \bar{a}_{\sim i}) - \sum_i \lambda^P (m^{upper} - e_i^*)\frac{\partial \bar{e}_{\sim i}^*}{\partial b}}{\sum_i \frac{\partial e_i^*}{\partial b}(1+\lambda^K \bar{a}_{\sim i})} = 0.$$

A.2 The Worker's Maximization Problem

We model the wage contract as $w_i = \alpha_i + bf_i$, where the individual-specific intercept allows the wage contract to match heterogeneous outside options of different workers. Because of risk neutrality, workers maximize their expected wage minus the combined cost of effort:[1]

$$\begin{aligned} EU_i &= E\left[w_i - C(e_i) - P\left(e_i, \bar{f}_{\sim i}\right)\right] = E[w_i] - C(e_i) - P(e_i, \bar{y}_{\sim i}) \\ &= \alpha_i + b[a_i + e_i(1+\lambda^K \bar{a}_{\sim i})] - ke_i^2 - \lambda^P (m - e_i)\bar{y}_{\sim i}. \end{aligned} \tag{A.1}$$

This maximization problem leads to a linear system of *N* reaction functions in which each worker in the peer group equates the expected marginal benefit of exerting effort, $b(1+\lambda^K \bar{a}_{\sim i})$, with its expected marginal cost $\frac{\partial C(e_i)}{\partial e_i} + \frac{\partial P(e_i, \bar{y}_{\sim i})}{\partial e_i}$, resulting in the following first order condition:

$$b(1+\lambda^K \bar{a}_{\sim i}) - \frac{\partial[C(e_i) + P(e_i, \bar{y}_{\sim i})]}{\partial e_i} = 0 \qquad for \;\; i = 1, \dots, N, \text{ or}$$

$$b(1+\lambda^K \bar{a}_{\sim i}) - (2ke_i - \lambda^P \bar{e}_{\sim i} - \lambda^P \bar{a}_{\sim i}) = 0 \qquad for \;\; i = 1, \dots, N, \text{ or}$$

$$e_i = \frac{\lambda^P}{2k}\bar{e}_{\sim i} + \frac{b}{2k} + \frac{\lambda^P + b\lambda^K}{2k}\bar{a}_{\sim i} \qquad for \;\; i = 1, \dots, N \tag{A.2}$$

We assume $k > \lambda^P$, which ensures that the firm's maximization problem has an interior solution (see Section A.3). This implies $2k > \lambda^P$ from which it follows that there exists a unique solution to the reaction function system. Note that $\bar{e}_{\sim i} = \frac{\mathrm{N\bar{e}} - e_i}{N-1}$, meaning that equation (A.2) can be rewritten as

$$e_i = \frac{\lambda^P}{2k}\frac{1}{N-1}[\mathrm{N\bar{e}} - e_i] + \frac{b}{2k} + \frac{\lambda^P + b\lambda^K}{2k}\bar{a}_{\sim i}$$

[1] Here, we use the fact that $E\left[P\left(e_i, \bar{f}_{\sim i}\right)\right] = P(e_i, \bar{y}_{\sim i})$ because P(.) is linear in $\bar{f}_{\sim i}$, $\bar{f}_{\sim i}$ is linear in $\bar{\varepsilon}_{\sim i}$, and $E[\bar{\varepsilon}_{\sim i}] = 0$. In the subsequent discussion, we simplify the notation by using $P(e_i, \bar{y}_{\sim i})$ in place of $E\left[P\left(e_i, \bar{f}_{\sim i}\right)\right]$.

Solving for e_i then gives

$$e_i\left(1+\frac{\lambda^P}{2k}\frac{1}{N-1}\right)=\frac{\lambda^P}{2k}\frac{\mathrm{N}}{N-1}\bar{\mathrm{e}}+\frac{b}{2k}+\frac{\lambda^P+b\lambda^K}{2k}\bar{a}_{\sim i}$$

$$e_i\left(\frac{2k(N-1)+\lambda^P}{2k(N-1)}\right)=\frac{\lambda^P}{2k}\frac{\mathrm{N}}{N-1}\bar{\mathrm{e}}+\frac{b}{2k}+\frac{\lambda^P+b\lambda^K}{2k}\bar{a}_{\sim i}$$

$$e_i=\frac{\lambda^P\mathrm{N}}{2k(N-1)+\lambda^P}\bar{\mathrm{e}}+\frac{(N-1)b}{2k(N-1)+\lambda^P}+\frac{(\lambda^P+b\lambda^K)(N-1)}{2k(N-1)+\lambda^P}\bar{a}_{\sim i} \qquad \text{(A.3)}$$

Taking averages on both sides of this equation yields

$$\bar{\mathrm{e}}=\frac{\lambda^P\mathrm{N}}{2k(N-1)+\lambda^P}\bar{\mathrm{e}}+\frac{(N-1)b}{2k(N-1)+\lambda^P}+\frac{(\lambda^P+\beta\lambda^K)(N-1)}{2k(N-1)+\lambda^P}\bar{a}$$

after which solving for $\bar{\mathrm{e}}$ gives

$$\bar{\mathrm{e}}\left(\frac{(2k-\lambda^P)(N-1)}{2k(N-1)+\lambda^P}\right)=\frac{(N-1)b}{2k(N-1)+\lambda^P}+\frac{(\lambda^P+\beta\lambda^K)(N-1)}{2k(N-1)+\lambda^P}\bar{a}$$

$$\bar{\mathrm{e}}=\frac{b}{(2k-\lambda^P)}+\frac{(\lambda^P+b\lambda^K)}{(2k-\lambda^P)}\bar{a}$$

$$\bar{\mathrm{e}}=\frac{b}{(2k-\lambda^P)}+\frac{(b\lambda^K+\lambda^P)(N-1)}{(2k-\lambda^P)N}\bar{a}_{\sim i}+\frac{(b\lambda^K+\lambda^P)}{(2k-\lambda^P)N}a_i$$

Substituting this expression into (A.3) yields

$$\begin{aligned}e_i^*=&\frac{\lambda^P\mathrm{N}}{2k(N-1)+\lambda^P}\left[\frac{b}{(2k-\lambda^P)}+\frac{(b\lambda^K+\lambda^P)(N-1)}{(2k-\lambda^P)N}\bar{a}_{\sim i}+\frac{(b\lambda^K+\lambda^P)}{(2k-\lambda^P)N}a_i\right]\\&+\frac{(N-1)b}{2k(N-1)+\lambda^P}+\frac{(\lambda^P+b\lambda^K)(N-1)}{2k(N-1)+\lambda^P}\bar{a}_{\sim i}\\=&\frac{b[2k(N-1)+\lambda^P]}{[2k(N-1)+\lambda^P](2k-\lambda^P)}+\frac{\lambda^P(b\lambda^K+\lambda^P)(N-1)}{[2k(N-1)+\lambda^P](2k-\lambda^P)}\bar{a}_{\sim i}\\&+\frac{\lambda^P(b\lambda^K+\lambda^P)}{[2k(N-1)+\lambda^P](2k-\lambda^P)}a_i+\frac{(\lambda^P+b\lambda^K)(N-1)}{2k(N-1)+\lambda^P}\bar{a}_{\sim i},\end{aligned}$$

or

$$e_i^* = \frac{b[2k(N-1)+\lambda^P] + \lambda^P(b\lambda^K+\lambda^P)a_i + 2k(b\lambda^K+\lambda^P)(N-1)\bar{a}_{\sim i}}{[2(N-1)k+\lambda^P](2k-\lambda^P)} \qquad \text{(A.4)}$$

A.3 The Firm's Optimization Problem

Substituting equation (A.1) evaluated at optimal effort levels into the participation constraint $EU_i = v(a_i)$ gives $Ew_i - C(e_i^*) - P(e_i^*, \bar{y}_{\sim i}) = v(a_i)$. Solving this expression for Ew_i yields equation (2) in Section 1.3 of the main text (i.e., $Ew_i = v(a_i) + C(e_i^*) + P(e_i^*, \bar{y}_{\sim i})$). Substituting this into the profit function $EP = \sum_i E[f_i - w_i]$ produces the following optimization problem for the firm's choice of b:

$$\max_b EP = \sum_i [a_i + e_i^*(1+\lambda^K\bar{a}_{\sim i}) - v(a_i) - C(e_i^*) - P(e_i^*, \bar{a}_{\sim i} + \bar{e}_{\sim i}^*)]$$

with first order condition

$$\sum_i \frac{\partial e_i^*}{\partial b}(1+\lambda^K\bar{a}_{\sim i}) - \sum_i \left(\frac{\partial C_i}{\partial e_i} + \frac{\partial P_i}{\partial e_i}\right)\frac{\partial e_i^*}{\partial b} - \sum_i \frac{\partial P_i}{\partial \bar{e}_{\sim i}^*}\frac{\partial \bar{e}_{\sim i}^*}{\partial b} = 0. \qquad \text{(A.5)}$$

Note that because of the workers' first order condition of maximizing marginal cost and marginal benefit, we have $\frac{\partial C_i}{\partial e_i} + \frac{\partial P_i}{\partial e_i} = b(1+\lambda^K\bar{a}_{\sim i})$ and hence we can rewrite (A.5) as

$$\sum_i \frac{\partial e_i^*}{\partial b}(1+\lambda^K\bar{a}_{\sim i}) - b\sum_i \frac{\partial e_i^*}{\partial b}(1+\lambda^K\bar{a}_{\sim i}) - \sum_i \frac{\partial P_i}{\partial \bar{e}_{\sim i}^*}\frac{\partial \bar{e}_{\sim i}^*}{\partial b} = 0.$$

Rearranging these elements gives

$$b^* = \frac{\sum_i \frac{\partial e_i^*}{\partial b}(1+\lambda^K\bar{a}_{\sim i}) - \sum_i \frac{\partial P_i}{\partial \bar{e}_{\sim i}^*}\frac{\partial \bar{e}_{\sim i}^*}{\partial b}}{\sum_i \frac{\partial e_i^*}{\partial b}(1+\lambda^K\bar{a}_{\sim i})}$$

$$= \frac{\sum_i \frac{\partial e_i^*}{\partial b}(1+\lambda^K\bar{a}_{\sim i}) - \sum_i \lambda^P(m-e_i^*)\frac{\partial \bar{e}_{\sim i}^*}{\partial b}}{\sum_i \frac{\partial e_i^*}{\partial b}(1+\lambda^K\bar{a}_{\sim i})}. \qquad \text{(A.6)}$$

Since peer pressure causes no extra utility to workers on average (because of our assumptions on *m*, see Section A.1 above), we have $\frac{1}{N}\sum_i \lambda^P(m-e_i^*)\frac{\partial \bar{e}_{\sim i}^*}{\partial b} \geq 0.$

Additionally, from both the expression for optimal effort given in Equation (A.4) and $\bar{e}^*_{\sim i} = \frac{b[2k(N-1)+\lambda^P]+2k(b\lambda^K+\lambda^P)a_i+(b\lambda^K+\lambda^P)[2k(N-2)+\lambda^P]\bar{a}_{\sim i}}{[2(N-1)k+\lambda^P](2k-\lambda^P)}$, it follows that $\frac{\partial e_i^*}{\partial b} > 0$ and $\frac{\partial \bar{e}^*_{\sim i}}{\partial b} > 0$. As a result, $b^* \leq 1$ for positive values of λ^P: Interestingly, in the presence of peer pressure, b^* is hence smaller than 1, and peer pressure constitutes a further reason for the firm to reduce incentives in addition to the well-known trade-off between risk and insurance, which is often emphasized in the principal agent model as important for risk-averse workers.[2] As m reaches its upper bound (very high pain from peer pressure), we even get $b^* = 0$, see Section A.1 above. In the absence of peer pressure (i.e., $\lambda^P = 0$), we obtain the standard result of an optimal incentive parameter for risk neutral workers that is equal to 1.

In the general case, there is no analytical closed-form solution for b^*, but for simplifying cases we can calculate a closed-form solution. Consider the case in which all workers have equal ability $a_i = \bar{a}$ and hence exert equal optimal effort $e_i^* = \bar{e}^* = \frac{b+b\lambda^K\bar{a}+\lambda^P\bar{a}}{2k-\lambda^P}$. The first order condition (A.5) simplifies to $\sum_i \frac{\partial \bar{e}^*}{\partial b}(1+\lambda^K\bar{a}) - b\sum_i \frac{\partial \bar{e}^*}{\partial b}(1+\lambda^K\bar{a}) - \sum_i \lambda^P(m-\bar{e}^*)\frac{\partial \bar{e}^*}{\partial b} = 0 \;\; \Leftrightarrow (1+\lambda^K\bar{a}) - b(1+\lambda^K\bar{a}) - \lambda^P(m-\bar{e}^*) = 0$, yielding the solution $b = \frac{(\lambda^P)^2\,\bar{a}+(2k-\lambda^P)\lambda^K\bar{a}-(\lambda^P m-1)(2k-\lambda^P)}{2(k-\lambda^P)(1+\lambda^K\bar{a})}$, which under the second order condition $-(1+\lambda^K\bar{a}) + \frac{\lambda^P}{2k-\lambda^P}(1+\lambda^K\bar{a}) < 0 \Leftrightarrow k > \lambda^P$ maximizes firm profits.

A.4 Productivity versus wage spillover effects

How does the spillover effect in wages, given by equation (3), compare with that in productivity, given by $\frac{1}{N}\sum_i \frac{dEf_i}{d\bar{a}_{\sim i}}$? The latter consists of two parts, the marginal effect of peer ability on productivity holding effort constant, $\frac{1}{N}\sum_i \lambda^K e_i^*$, plus an additional effect arising from the endogenous response of effort, $\frac{1}{N}\sum_i \frac{\partial f_i}{\partial e_i}\frac{de_i^*}{d\bar{a}_{\sim i}}$. There are two opposing

[2] This outcome results from an externality: the failure of individual workers to internalize in their effort choices the fact that peer pressure causes their peers additional "pain" for which the firm must compensate. The firm mitigates this externality by setting $b^* < 1$.

effects. On the one hand, the first term in equation (3) (i.e., $\frac{1}{N}\sum_i \beta^* \frac{\partial f_i}{\partial e_i}\frac{de_i^*}{d\bar{a}_{\sim i}}$) is smaller than the productivity spillover effect, for two reasons. First, firms do not compensate workers for an increase in productivity that is not induced by an increase in effort (i.e., $\frac{1}{N}\sum_i \lambda^K e_i^*$ is missing from the peer effect in wages—an effect that arises only under knowledge spillover); and second, increases in productivity induced by an increase in effort (i.e., $\frac{1}{N}\sum_i \frac{\partial f_i}{\partial e_i}\frac{de_i^*}{d\bar{a}_{\sim i}}$) translate into wages at a rate of smaller than 1, given by b^*. (Note that this effect arises only under peer pressure; if $\lambda^P = 0, b^* = 1$.) On the other hand, the wage spillover effect includes an additional term that is absent from the productivity spillover effect (term 2 in equation (3)—an effect that arises once again only under peer pressure. This term captures that workers dislike working in high pressure environments, forcing firms to compensate workers for this extra disutility. The effect in equation (3) contains both direct and indirect (social multiplier) effects of $\bar{a}_{\sim i}$ on the wage. For example, $\frac{de_i^*}{d\bar{a}_{\sim i}}$ not only contains the direct effect of $\bar{a}_{\sim i}$ on e_i, i.e., $\frac{\lambda^P + b\lambda^K}{2k}$ in equation (1), but also additional multiplier effects as own effort and peer effort reinforce each other (see A.2).

References:

Barron, John M., and Kathy Paulson Gjerde. "Peer pressure in an agency relationship." *Journal of Labor Economics* (1997): 234-254.

Appendix B: Variation used in the within-peer group estimator

Denoting peer group size by N_{ojt}, since $\frac{1}{N_{ojt}}\sum_i \bar{a}_{\sim i,ojt} = \frac{1}{N_{ojt}}\sum_i a_i = \bar{a}_{ojt}$, the within-peer group transformation of equation (6) that eliminates the peer group fixed effect is

$$\ln(w_{iojt}) - \overline{\ln(w)}_{ojt} = (x_{iojt}^T - \bar{x}_{ojt}^T)\beta + (a_i - \bar{a}_{ojt}) + \gamma(\bar{a}_{\sim i,ojt} - \bar{a}_{ojt}) +$$
$$(\varepsilon_{iojt} - \bar{\varepsilon}_{ojt}),$$

which can in turn be transformed into[3]

$$\ln(w_{iojt}) - \overline{\ln(w)}_{ojt} = (x_{iojt}^T - \bar{x}_{ojt}^T)\beta + (a_i - \bar{a}_{ojt}) + \gamma \frac{-1}{(N_{ojt}-1)}(a_i - \bar{a}_{ojt}) + (\varepsilon_{iojt} - \bar{\varepsilon}_{ojt})$$

This calculation shows a close association in the within-peer group transformed model between individual ability and average peer ability: for a one-unit change in individual ability relative to the average peer ability $a_i - \bar{a}_{ojt}$, peer quality relative to the average $\bar{a}_{\sim i,ojt} - \bar{a}_{ojt} = \frac{-1}{(N_{ojt}-1)}(a_i - \bar{a}_{ojt})$ changes by a factor of $\frac{-1}{(N_{ojt}-1)}$. This outcome not only reflects the fact that better individuals within a peer group have worse peers but also shows that the magnitude of the drop in peer quality for each additional unit of individual ability declines with peer group size. Thus, in the within-peer group transformed model, individual ability $a_i - \bar{a}_{ojt}$ and peer quality $\bar{a}_{\sim i,ojt} - \bar{a}_{ojt}$ only vary independently if there is heterogeneity in the peer group size N_{ojt}. The parameter γ is thus identified by an interaction of a term involving N_{ojt} and within-transformed individual ability $(a_i - \bar{a}_{ojt})$.

Appendix C: Estimation method

The solution to estimating equation (4) by nonlinear least squares minimizes the following objective function:

$$\min_{\beta,\gamma,a_i,\omega_{ot},\delta_{jt}} M = \sum_i \sum_t \left[\ln(w_{iojt}) - x_{iojt}\beta - a_i - \gamma\bar{a}_{\sim i,ojt} - \omega_{ot} - \delta_{jt} - \theta_{oj}\right]^2 \qquad \text{(A.7)}$$

The algorithm proposed by Arcidiacono et al. (2012) first fixes a_i at starting values and then iterates the following steps:

[3] Here, we use $\bar{a}_{\sim i,ojt} - \bar{a}_{ojt} = \frac{N_{ojt}\bar{a}_{ojt} - a_i}{(N_{ojt}-1)} - \frac{(N_{ojt}-1)\bar{a}_{ojt}}{(N_{ojt}-1)} = \frac{-1}{(N_{ojt}-1)}(a_i - \bar{a}_{ojt})$.

1. Hold a_i and $\bar{a}_{\sim i,ojt}$ at the values from the previous step and obtain the least square estimates of the now linear model.
2. Update the a_is based on the nonlinear least squares objective function M given in (A.7), where all other coefficients are set to their estimated values from Step 2. Solving $\frac{\partial M}{\partial a_i} = 0$ for a_i yields functions $a_i = f(a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_N)$, which are applied to all a_i repeatedly until convergence, which is ensured under the condition that feedback effects are not too strong (i.e., $\gamma < 0.4$).
3. With the newly updated a_i go back to Step 2 until the parameter estimates converge.

Because the linear model to be solved in Step 2 still includes the high dimensional fixed effects δ_{jt}, ω_{ot}, and θ_{oj}, we employ a variant of the preconditioned conjugate gradient algorithm to solve this step (see Abowd, Kramarz and Margolis, 1999; Abowd, Creecy, and Kramarz, 2002, for details) that is efficient for very large data matrices.[4]

Because the algorithm does not deliver standard errors and the data matrix is too large to be inverted without hitting computer memory restrictions, we compute the standard errors by implementing a wild bootstrapping with clustering on firms (Cameron, Gelbach, and Miller, 2008).[5] For the baseline model, we verify that when using 100 bootstraps, standard errors are very stable after the 30th bootstrap. Because the estimation is time consuming, therefore, we generally use 30 bootstraps for each model.

References:

Abowd, John M., Francis Kramarz, and David N. Margolis. "High wage workers and high wage firms." *Econometrica* 67, no. 2 (1999): 251-333.

Abowd, John M., Robert H. Creecy, and Francis Kramarz. "Computing person and firm effects using linked longitudinal employer-employee data." Technical Paper No. TP-2002-06, U.S. Census Bureau, 2002.

Arcidiacono, Peter, Gigi Foster, Natalie Goodpaster, and Josh Kinsler. "Estimating spillovers using panel data, with an application to the classroom." *Quantitative Economics* 3, no. 3 (2012): 421-470.

[4] We implement the estimation in Matlab based on sparse matrix algebra for efficient data manipulation of the large dummy variable matrices.

[5] Rather than using different observations across bootstraps, this method draws a new residual vector at each iteration, which has the advantage of leaving the structure of worker mobility between firms unchanged across the bootstraps, thereby allowing identification of the same set of worker and firm fixed effects in each bootstrap. Another advantage is that this bootstrap is applicable to clusters of different sizes.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. "Bootstrap-based improvements for inference with clustered errors." *Review of Economics and Statistics* 90, no. 3 (2008): 414-427.

Appendix D: Bias from wrong peer group definitions

Defining the peer group at the firm-occupation level leads to two possible error types: excluding relevant peers from outside the occupational group or including irrelevant peers inside the occupational group. In this section, we discuss the possible bias resulting from a wrong peer group definition. This demonstration assumes that the individual and average peer abilities are known. If (as in practice) they need to be estimated, an additional bias may arise from a false definition of peer group.

We denote the average quality of individual i's true and observed peer group by $\bar{a}^{\text{true}}_{\sim i,ojt}$ and $\bar{a}^{\text{obs}}_{\sim i,ojt}$ and suppose that the true model of peer effects is

$$\ln w_{iojt} = \mu_{iojt} + \gamma \bar{a}^{\text{true}}_{\sim i,ojt} + u_{iojt},$$

where μ_{iojt} summarizes the control variables and multiple fixed effects included in the baseline or the within-peer group specification. Because the worker's true peer group is unobserved, we instead run the regression

$$\ln w_{iojt} = \mu_{iojt} + \gamma \bar{a}^{\text{obs}}_{\sim i,ojt} + e_{iojt}$$

with $e_{iojt} = \gamma \bar{a}^{\text{true}}_{\sim i,ojt} - \gamma \bar{a}^{\text{obs}}_{\sim i,ojt} + u_{iojt}$. The coefficient on $\bar{a}^{\text{obs}}_{\sim i,ojt}$ then identifies $E[\hat{\gamma}] = \gamma\rho$ with ρ equal to the coefficient from a regression of true average peer quality on observed average peer quality $\bar{a}^{\text{true}}_{\sim i,ojt} = r_{iojt} + \rho \bar{a}^{\text{obs}}_{\sim i,ojt} + \varepsilon_{iojt}$, where r_{iojt} includes the same control variables as μ_{iojt}.[6] The factor ρ that characterizes the bias is thus determined by the extent to which observed peer quality shifts true peer quality.

Consider the following special case. The observed peer group is defined at a given level of aggregation indexed by l (say, at the level of the firm), $\bar{a}^{\text{obs}}_{\sim i,ojt} = E[a_i|l]$, while the true peer group is defined at a lower level of aggregation indexed by k (say, three digit occupations within firms). If k is nested within l, then the mean of true average peer

[6] This can be seen by noting that e_{iojt} includes two omitted variables, $\bar{a}^{\text{true}}_{\sim i,ojt}$ and $\bar{a}^{\text{obs}}_{\sim i,ojt}$. The omitted variable bias due to these two terms is equal to their respective effect on $\ln w_{it}$ (which is γ for $\bar{a}^{\text{true}}_{\sim i,ojt}$ and $-\gamma$ for $\bar{a}^{\text{obs}}_{\sim i,ojt}$) multiplied by how much each of them is shifted by the included regressor $\bar{a}^{\text{obs}}_{\sim i,ojt}$ (which is ρ for $\bar{a}^{\text{true}}_{\sim i,ojt}$ and 1 for $\bar{a}^{\text{obs}}_{\sim i,ojt}$). Thus, the bias is $\gamma\rho - \gamma$, and thus $E[\hat{\gamma}] = \gamma + \gamma\rho - \gamma = \gamma\rho$.

quality at level k can be decomposed into its mean at the wider level l and its deviation from that mean, i.e., $\bar{a}^{\text{true}}_{\sim i,ojt} = E[a_i|k] = E[a_i|l] + (E[a_i|k] - E[a_i|l]) = \bar{a}^{\text{obs}}_{\sim i,ojt} + (E[a_i|k] - E[a_i|l])$. Note that the second part of this expression, $E[a_i|k] - E[a_i|l]$ (the deviation from the mean), is not correlated with $\bar{a}^{\text{obs}}_{\sim i,ojt}$ (the mean itself), thus $\text{Cov}\left(\bar{a}^{\text{true}}_{\sim i,ojt}, \bar{a}^{\text{obs}}_{\sim i,ojt}\right) = \text{Var}\left(\bar{a}^{\text{obs}}_{\sim i,ojt}\right)$. Therefore, in the case of no further control variables (i.e., $\mu_{iojt} = \mu$ and $r_{iojt} = r$), $\rho = \frac{\text{Cov}\left(\bar{a}^{\text{true}}_{\sim i,ojt}, \bar{a}^{\text{obs}}_{\sim i,ojt}\right)}{\text{Var}\left(\bar{a}^{\text{obs}}_{\sim i,ojt}\right)} = 1$. Thus, in this special case of nested peer groups and no control variables, there is no bias ($\rho = 1$) from defining the peer group as too large.

The opposite case of defining the peer group as too small can be considered by simply switching true and observed peer group, i.e., the true peer group is now at the wider level of aggregation l, $\bar{a}^{\text{true}}_{\sim i,ojt} = E[a_i|l]$, while the observed peer group is now at the narrower level k which can again be decomposed into the mean at level l and its deviation from the mean, leading to $\bar{a}^{\text{obs}}_{\sim i,ojt} = \bar{a}^{\text{true}}_{\sim i,ojt} + (E[a_i|k] - E[a_i|l])$. This then leads to $\rho = \frac{\text{Cov}\left(\bar{a}^{\text{true}}_{\sim i,ojt}, \bar{a}^{\text{obs}}_{\sim i,ojt}\right)}{\text{Var}\left(\bar{a}^{\text{obs}}_{\sim i,ojt}\right)} = \frac{\text{Var}\left(\bar{a}^{\text{true}}_{\sim i,ojt}\right)}{\text{Var}\left(\bar{a}^{\text{true}}_{\sim i,ojt} + (E[a_i|k] - E[a_i|l])\right)} < 1$. In this configuration there is excess variance or noise in the observed average peer ability which leads to attenuation bias similar to classical measurement error.

Thus, in the simple case of different levels of aggregation with nested peer group definitions and without control variables, defining the peer group as too large leads to no bias, whereas defining the peer group as too small leads to attenuation bias.

The result of no bias when the peer group is defined as too large does, however, not in general hold when adding control variables. Suppose $\rho = 1$ holds in a bivariate regression of $\bar{a}^{\text{true}}_{\sim i,ojt}$ on $\bar{a}^{\text{obs}}_{\sim i,ojt}$. When augmenting this regression by additional control variables that are positively (or negatively) related to both $\bar{a}^{\text{true}}_{\sim i,ojt}$ and $\bar{a}^{\text{obs}}_{\sim i,ojt}$, ρ will be reduced and become smaller than one. Obvious examples in our context are worker and firm fixed effects, which are both positively correlated with peer quality measured at different levels of aggregation within the firm. We thus expect $\rho < 1$ (attenuation bias) both when defining the peer group as too large or defining it as too small.

Some evidence consistent with attenuation bias is provided in row (x) of Table 7, which shows a substantial drop in the peer effect when defining the peer group at a smaller than the 3-digit occupational level. Our results from Panel A of Table 6 further show that peers outside the own 3-digit occupation in the same firm do not seem to affect wages. This leads us to believe that the 3-digit occupational level within the firm is the most appropriate peer group definition in our context.

Appendix E: Imputation of censored wage observations

To impute the top-coded wages, we first define age-education cells based on five age groups (with 10-year intervals) and three education groups (no post-secondary education, vocational degree, college or university degree). Within each of these cells, following Dustmann et al. (2009) and Card et al. (2013), we estimate Tobit wage equations separately by year while controlling for age; firm size (quadratic, and a dummy for firm size greater than 10); occupation dummies; the focal worker's mean wage and mean censoring indicator (each computed over time but excluding observations from the current time period); and the firm's mean wage, mean censoring indicator, mean years of schooling, and mean university degree indicator (each computed at the current time period by excluding the focal worker observations). For workers observed in only one time period, the mean wage and mean censoring indicator are set to sample means, and a dummy variable is included. A wage observation censored at value c is then imputed by the value $X\hat{\beta} + \hat{\sigma}\Phi^{-1}[k + u(1-k)]$, where Φ is the standard normal CDF, u is drawn from a uniform distribution, $k = \Phi[(c - X\hat{\beta})/\hat{\sigma}]$, and $\hat{\beta}$ and $\hat{\sigma}$ are estimates for the coefficients and standard deviation of the error term from the tobit regression.

References:

Card, David, Jörg Heining, and Patrick Kline. "Workplace heterogeneity and the rise of West German wage inequality." *Quarterly Journal of Economics* 128, no. 3 (2013): 967-1015.

Dustmann, Christian, Johannes Ludsteck, and Uta Schönberg. "Revisiting the German Wage Structure." *Quarterly Journal of Economics* 124, no. 2 (2009): 843–881.

Appendix F: Additional Results

F.1: Short T Bias – Monte Carlo Simulations

The peer effects estimator we use is consistent for large N and fixed T under the assumption that error terms are uncorrelated across observations (Theorem 1 in Arcidiacono et al., 2012). Correlated random shocks in the error terms of peers in the same peer group would violate this assumption. Positively correlated shocks would partly be absorbed in the peers' estimated fixed effects, causing an upward bias due to a spurious positive correlation between estimated peer quality and wages. This bias is likely to disappear as T gets large.

In Table F.1 below we report results from a Monte Carlo study to explore this type of bias. We show that adding a peer-group level random shock to the error term indeed induces an upward bias in our baseline estimator, and that this bias increases with the size of the shock (as measured by its share in the total error variance), and decreases with higher T. We also show that this bias is absent in the within-peer group estimator, as this estimator absorbs common peer group level shocks. Finally, we show that serial correlation of a plausible magnitude in the individual error term does not seem to bias our estimates in any important way.

The dependent variable is simulated in the following way. We first predict the log wage in our original estimation sample, setting coefficients of control variables and fixed effects to their estimates from the baseline model. For the simulations of peer-group specific shocks, we then add a normally distributed error term with a variance equal to the estimated error variance from the baseline model, composed of two components, an idiosyncratic shock and a peer-group-by-time-level shock. For the simulations of serial correlation, on the other hand, we add a normally distributed error term with variance equal to the estimated error variance from the baseline model, and with first-order serial correlation at individual level. Across the rows of the table, we vary the true coefficient on the average peer fixed effect in repetitive occupations, using values 0, 0.03 and 0.045. Across columns of Panels A through C of the table, we vary the variance of the group-level shock as a share of the error total variance using the values 0, 0.03 and 0.06. A share of 0.06 is equal to the R-squared from a regression of the predicted error of our baseline model onto peer-group-by-time fixed effect and therefore seems to be an appropriate choice as an upper bound. In Panel D, we vary the first-order autocorrelation coefficient, using values 0, 0.1, and 0.2. The value of 0.2 is equal to the autocorrelation that we detect empirically when regressing the residual from our baseline specification on its lagged values, and we thus choose this as an upper bound for the simulations. In column (4) of the respective panels, we report the difference between the simulation result with an error share of 0.06 and an error share of 0 (in Panels A-C) and between an autocorrelation coefficient of 0.2 and 0 (in Panel D). We interpret these differences as upper bounds for the bias.

In Panel A, we estimate the model using our baseline estimator and exploiting the full number of time periods. In this case, our estimate for the upper bound of the bias in column (4) does not exceed 0.03. Thus, it is unlikely that our baseline peer effect estimate of 0.064 is purely a result of statistical bias.

To assess the importance of the number of time periods T, we omit in Panel B every second time period of our sample, reducing the maximum number of time periods from 17 years to 9 years. This increases the upward bias to about 0.045, confirming that the bias can indeed be thought of as a short T bias.

In Panel C, we use the within-peer group estimator outlined in Section 2.3. This estimator conditions on the full set of time-variant peer group fixed effects p_{ojt} and thus on shocks to the peer group. The results reveal essentially no upward bias for statistical reasons for the within-peer group estimator. Hence, this estimator does not only eliminate a possible bias in the peer effect due to economic reasons, but also due to statistical reasons and the similarity of results from our baseline specification and the within peer group specification suggests that any possible upward bias due to peer-group level shocks because of either economic or statistical reasons is small.

The results in Panel D further suggest that serial correlation of a plausible magnitude does not seem to bias our estimates in any important way.

References:

Arcidiacono, Peter, Gigi Foster, Natalie Goodpaster, and Josh Kinsler. "Estimating spillovers using panel data, with an application to the classroom." *Quantitative Economics* 3, no. 3 (2012): 421-470.

Table F.1: Monte Carclo Study to assess bias from correlated shocks

Panel A: Baseline specification, maximum number of time periods T_{max} = 17

	(1)	(2)	(3)	(4)
Share of group level error in total error	0	0.03	0.06	diff. (3)-(1)
True coefficient = 0	0.003	0.013	0.03	0.027
True coefficient = 0.03	0.028	0.043	0.058	0.030
True coefficient = 0.045	0.044	0.061	0.073	0.029

Panel B: Baseline specification, maximum number time periods reduced to T_{max} = 9

	(1)	(2)	(3)	(4)
Share of group level error in total error	0	0.03	0.06	diff. (3)-(1)
True coefficient = 0	0.001	0.022	0.045	0.044
True coefficient = 0.03	0.028	0.054	0.075	0.047
True coefficient = 0.045	0.048	0.075	0.094	0.046

Panel C: Within-peer group estimator, maximum number of time periods T_{max} = 17

	(1)	(2)	(3)	(4)
Share of group level error in total error	0	0.03	0.06	diff. (3)-(1)
True coefficient = 0	0.003	0.001	0.002	-0.001
True coefficient = 0.03	0.033	0.031	0.028	-0.005
True coefficient = 0.045	0.046	0.048	0.041	-0.005

Panel D: Baseline specification, maximum number of time periods T_{max} = 17, serial correlation

	(1)	(2)	(3)	(4)
Serial correlation coefficient	0	0.1	0.2	diff. (3)-(1)
True coefficient = 0.045	0.045	0.044	0.045	-0.0001

Note: The table assesses the bias from correlated wage shocks when the number of time periods T is short using a Monte Carlo Study. Throughout the table, the dependent variable is simulated in the following way. We first predict the log wage in our original estimation sample, setting coefficients of control variables and fixed effects to their estimates from the baseline model. We then add a normally distributed error term with a variance equal to the estimated error variance from the baseline model, composed of two components, an idiosyncratic shock and a peer-group-by-time-level shock. Across the rows of the table we vary the true coefficient on the average peer fixed effect in repetitive occupations, using values 0, 0.03 and 0.045. Across columns of the table, we vary the variance of the group-level shock as a share of the error total variance using the values 0, 0.03 and 0.06. A share of 0.06 is equal to the R-squared from a regression of the predicted error of our baseline model onto peer-group-by-time fixed effect and therefore seems to be an appropriate choice as an upper bound. In column (4), we report the difference between the simulation result with an error share of 0.06 and an error share of 0, which we interpret as an upper bound for the bias.

In Panel A, we estimate the model using our baseline estimator and exploiting the full number of time periods. In Panel B, we drop every second year of our sample reducing the maximum number of time periods from 17 years to 9 years. In Panel C, we use the within-peer group estimator. In Panel D, we model serial correlation in the individual error term instead of a common peer-group level shock, and estimate the model by our baseline estimator. Each simulation is based on 10 repetitions for the baseline estimator and 15 repetitions for the within-peer group estimator. N=12,832,842 in Panels A and C, N=6,787,474 in Panel B.

Data Source: Social Security Data, One Large Local Labor Market, 1989-2005.

F.2: Sample Selection: Munich vs West Germany

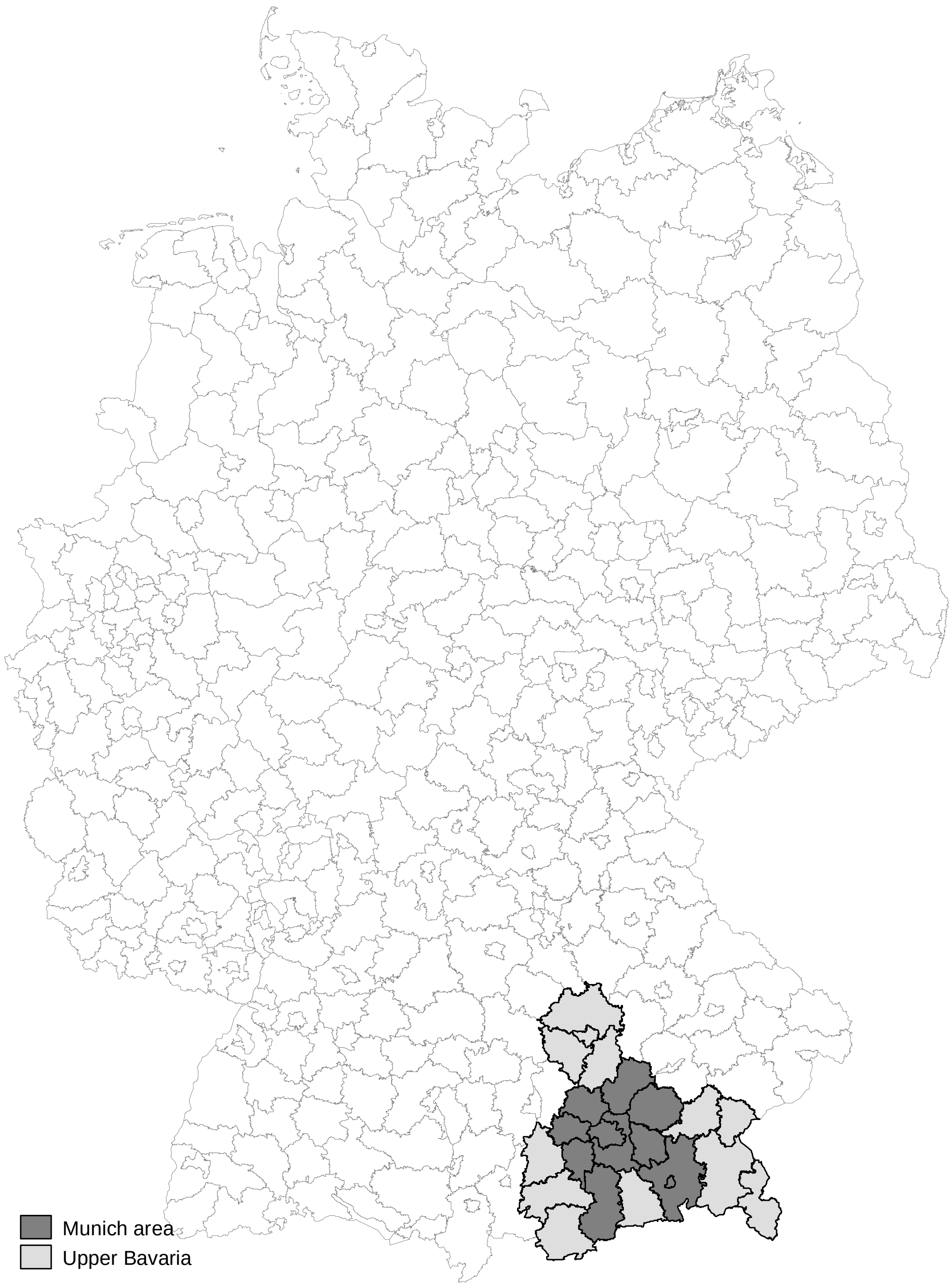
Our baseline estimation sample covers the metropolitan area of Munich, while in the robustness checks section we extend this by including surrounding rural areas ("upper Bavaria"). The geographical location is indicated on the map in Figure F.1 below. In Table F.2 below we compare the Munich and the Upper Bavaria samples with other metropolitan areas (a joint sample of Hamburg, Frankfurt and Cologne) and with the whole of Germany in terms of socio-economic and labor market characteristics. The table entries show that – in terms of observable characteristics – our main sample is very similar to other metropolitan areas, with perhaps a slightly higher share of college degree holders. The sample of Upper Bavaria is similar to that of Germany overall, with a higher share of foreign nationals.

Table F.2: Comparison of estimation sample with other metropolitan areas and Germany as a whole (1997)

	Munich	Upper Bavaria	Other Metropolitan areas	Germany
Log wage	4.49	4.43	4.43	4.29
Job tenure	5.83	5.97	6.10	5.60
College degree	0.20	0.16	0.15	0.12
Vocational and/or school degree	0.71	0.74	0.74	0.78
Years of schooling	13.95	13.75	13.68	13.52
Female	0.37	0.36	0.34	0.35
Age <= 35	0.44	0.45	0.43	0.43
Age	39.01	38.64	39.25	38.97
Foreign	0.16	0.14	0.11	0.08
Occupations with >=10% college graduates	0.51	0.45	0.45	0.38
Occupations with <=2.5 repetitiveness index	0.15	0.18	0.19	0.23
Firm size	1873.04	1836.27	1236.13	992.99
Censoring indicator	0.18	0.15	0.13	0.08
No. of observations	814,179	1,123,570	3,021,463	20,706,154

Note: The table presents mean values of socio-economic and labor market indicators to compare our estimation samples of Munich and Upper Bavaria (which extends the metropolitan area of Munich mainly by adding rural surrounding districts) with other metropolitan areas (Hamburg, Cologne and Frankfurt) and with West Germany as a whole. Entries refer to 1997, the middle of our estimation period.
Data Source: Social Security Data, West Germany, 1997.

Figure F.1: Map of metropolitan area of Munich (dark grey) and the additional districts of Upper Bavaria (light grey)



Note: The map shows the districts of the metropolitan area of Munich (our baseline sample) and the additional districts of the region Upper Bavaria (the sample used in robustness check (vi) in Table 7).

F.3: List of Occupations in Sub-Samples

Table F.3 below lists the occupations used in the different sub-samples of Table 4 in the main text.

Table F.3: List of Occupations in Sub-Samples

(1)

5% most repetitive occupations	**Share in %**
Unskilled laborer, helper (no further specification)	15.12
Packagers, goods receivers, despatchers	11.58
Metal workers (no further specification)	10.66
Postal deliverers	7.58
Assemblers (no further specification)	5.47
Street cleaners, refuse disposers	4.70
Assemblers of electrical parts or appliances	4.68
Cashiers	**4.00**
Railway controllers and conductors	3.96
Laundry workers, pressers	3.69
Machinery or container cleaners and related occupations	2.87
Railway engine drivers	2.80
Milk and fat processing operatives	2.62
Vehicle cleaners, servicers	2.57
Clothing sewers	2.02
Wood preparers	1.96
Metal grinders	1.92
Ceramics workers	1.20
Brick or concrete block makers	1.06
Tobacco goods makers	0.97
Sheet metal pressers, drawers, stampers	0.86
Solderers	0.86
Agricutlural helpers	**0.75**
Model or form carpenters	0.68
Sewers	0.66
Meat and sausage makers	0.61
Stoneware and earthenware makers	0.49
Enamellers, zinc platers and other metal surface finishers	0.42
Leather clothing makers and other leather processing operative	0.33
Metal moulders (non-cutting deformation)	0.27
Rubber makers and processors	0.27
Other wood and sports equipment makers	0.24
Earth, gravel, sand quarriers	0.22
Machined goods makers	0.20
Moulders, coremakers	0.19
Vulcanisers	0.18
Textile finishers	0.16
Footwear makers	0.15
Other textile processing operatives	0.15
Ready-meal, fruit and vegetable preservers and preparers	0.13
Weavers	0.13
Spinners, fibre preparers	0.12
Textile dyers	0.09
Planers	0.07
Spoolers, twisters, ropemakers	0.05
Post masters	0.05
Radio operators	0.04
Hat and cap makers	0.04
Ship deckhand	0.04
Cartwrights, wheelwrights, coopers	0.03
Rollers	0.03
Wood moulders and related occupations	0.02
Fine leather goods makers	0.02
Fish processing operatives	0.01
Metal drawers	0.01
Jewel preparers	0.01

(2)

Hand-picked occupations with low learning content	**Share in %**
Salespersons	24.0
Motor vehicle drivers	19.4
Store and warehouse workers	10.9
Household cleaners	8.9
Waiters, stewards	8.0
Unskilled laborer, helper	5.9
Packagers, goods receivers, despatchers	4.5
Gardeners, garden workers	3.7
Goods examiners, sorters, n.e.c.	3.3
Street cleaners, refuse disposers	1.8
Cashiers	1.6
Glass, buildings cleaners	1.5
Laundry workers, pressers	1.4
Transportation equipment drivers	1.4
Vehicle cleaners, servicers	1.0
Earthmoving plant drivers	0.8
Construction machine attendants	0.7
Crane drivers	0.4
Stowers, furniture packers	0.3
Agricutlural helpers	0.3

(3)

10% most skilled occupations	**Share in %**
Electrical engineers	24.3
Mechanical, motor engineers	13.1
Management consultants, organisors	10.3
Other engineers	7.6
Architects, civil engineers	7.3
Physicians	5.2
Economic and social scientists, statisticians	4.7
Scientists	3.8
Ministers of religion	3.2
Other manufacturing engineers	2.9
Senior government officials	2.9
Physicists, physics engineers, mathematicians	2.5
Technical, vocational, factory instructors	2.0
Legal representatives, advisors	1.9
Primary, secondary (basic), special school teachers	1.7
Chemists, chemical engineers	1.5
University teachers, lecturers	1.3
Gymnasium teachers	1.3
Pharmacists	0.8
Academics / Researchers in the Humanities	0.8
Garden architects, garden managers	0.4
Survey engineers	0.3
Veterinary surgeons	0.1
Mining, metallurgy, foundry engineers	0.1
Dentists	0.1

(4)

10% most innovative occupations	**Share in %**
Data processing specialists	38.7
Electrical engineers	24.7
Mechanical, motor engineers	13.3
Architects, civil engineers	7.5
Scientists	3.8
Other manufacturing engineers	2.9
Physicists, physics engineers, mathematicians	2.5
Chemists, chemical engineers	1.5
University teachers, lecturers	1.3
Musicians	0.9
Interior, exhibition designers, window dresser	0.8
Packaging makers	0.5
Garden architects, garden managers	0.4
Brokers, property managers	0.3
Survey engineers	0.3
Scenery, sign painters	0.2
Veterinary surgeons	0.2
Mining, metallurgy, foundry engineers	0.1
Forestry managers, foresters, hunters	0.1
Coachmen	0.0

(5)

Hand-picked occupations with high learning content	**Share in %**
Electrical engineers	20.6
Entrepreneurs, managing directors, divisional manage	18.1
Mechanical, motor engineers	11.1
Management consultants, organisors	8.7
Other engineers	6.5
Architects, civil engineers	6.2
Chartered accountants, tax advisers	5.5
Physicians	4.4
Economic and social scientists, statisticians	4.0
Scientists	3.2
Other manufacturing engineers	2.5
Senior government officials	2.4
Physicists, physics engineers, mathematicians	2.1
Legal representatives, advisors	1.6
Chemists, chemical engineers	1.3
Humanities specialists	0.7
Association leaders, officials	0.6
Survey engineers	0.3
Veterinary surgeons	0.1
Mining, metallurgy, foundry engineers	0.1
Dentists	0.1

Note: The table presents the lists of occupations in for the different sub-samples of occupations used in table 4.
Data Source: German Social Security Data, One Large Local Labor Market, 1989-2005. N=12,832,842.

F.4: Results by Peer Group and Firm Size

Table F.4 below reports the peer effect by peer group size and firm size categories for the samples of repetitive and skilled occupations. The estimates are all very similar across the different categories of peer group and firm sizes for both most repetitive and high skilled occupations, and similar to the average estimates.

Table F.4: Additional heterogeneity by peer group size and firm size

	5% most repetitive occupations	10% most skilled occupations
Panel A: Peer effect by peer group size		
Group size 2-10	0.068	0.014
	(0.0020)	(0.0013)
Group size 11-20	0.079	0.016
	(0.0037)	(0.0026)
Group size 21-50	0.078	0.016
	(0.0034)	(0.0029)
Group size 51-100	0.081	0.014
	(0.0049)	(0.0029)
Group size >100	0.081	0.014
	(0.0042)	(0.0027)
Panel B: Peer effect by firm size		
Firm size 2-20	0.072	0.015
	(0.0022)	(0.0015)
Firm size 21-50	0.073	0.015
	(0.0023)	(0.0015)
Firm size 51-100	0.072	0.015
	(0.0023)	(0.0016)
Firm size 100-500	0.072	0.015
	(0.0023)	(0.0016)
Firm size >500	0.073	0.014
	(0.0024)	(0.0017)

Note: The table reports estimates for peer effects in the 5% most repetitive occupations (N=681,391) and the 10% most skilled occupations (N=1,309,070) separately by peer group size and firm size. Estimates are based on pre-estimated fixed effects from our baseline specification, and include the same set of controls and fixed effects as our baseline specification in equation (4).

Data Source: Social Security Data, One Large Local Labor Market, 1989-2005.

F.5: IV Estimates

In Table F.5 we present IV estimates from regressions for peer group stayers of the wage change on the change in peer quality based on pre-estimated fixed effects from the baseline model. We instrument the change in peer quality by the average quality of leavers from the peer group who in t-1 were close to retirement age; the rationale being that leaving into retirement may be more exogenous than other reasons for the turnover of peers. Specifically, the instrument is the average wage fixed effect of peer group leavers who were close to retirement age (aged 63 or above), multiplied by the share of these leavers relative to the peer group size. This gives us a strong first stage (F-value 151.1) with expected negative sign: a higher quality of the peers leaving into retirement reduces the change in peer quality. The IV peer effect coefficient is 0.041, not too far off our baseline estimate of 0.064, although imprecisely estimated and not statistically significant.

Table F.5: Instrumental variables (IV) estimates of the peer effect (5% most repetitive occupations, based on pre-estimated fixed effects)

	IV: Average wage fixed effect of leavers close to retirement
IV estimate of peer effect	0.0409
	(0.0679)
First stage effect	-0.3487
	(0.0284)
First stage F statistic	151.1
N	79,316

Note: The table presents IV estimates of regressions for stayers of the wage change on the change in peer quality based on pre-estimated fixed effects from the baseline model. The instrument is the average wage fixed effect of peer group leavers (workers who were in the peer group in t-1 but not in t) who were close to retirement age (aged 63 or above), multiplied by the share of these leavers relative to the peer group size.
Data Source: German Social Security Data, One Large Local Labor Market, 1989-2005.