Firm Productivity, Wages, and Sorting
Online Appendix

Benjamin Lochner and Bastian Schulz

A Model Details

A.1 Matching Technology

Due to random search, firms cannot target their vacancies to specific worker types. They post vacancies v subject to a productivity-dependent cost $c(\Omega)$. Meetings are generated by a Cobb–Douglas matching function with constant returns to scale (Pissarides, 2000; Petrongolo and Pissarides, 2001). Without loss of generality, let worker ability x and firm productivity Ω be distributed uniformly over $[0, 1]$. Meeting rates are functions of aggregate labor market tightness, $\theta = {}^{V}/_{U}$, where $V = \int g_v(\Omega)\mathrm{d}\Omega$ and $U = \int g_u(x)\mathrm{d}x$ are the aggregate numbers of vacancies and unemployed workers, respectively. $q_v(\theta)$ is the rate at which firms meet workers, and $q_u(\theta)$ is the rate at which unemployed workers meet vacancies.

$g_v(\Omega)$ is the PDF of vacancies at type Ω firms, and $g_u(x)$ is the PDF of unemployed workers of type x. In this environment, a match is not guaranteed conditional on a meeting. Suppose that a type x worker meets a productivity Ω firm. Both parties may prefer to continue searching if the match surplus $S(x, L, \Omega)$, defined below, is negative. Note that the surplus depends on the worker type, the firm type, and the total composite labor input. In the steady state, existing matches can end only at the exogenous rate δ. Endogenous separations may happen outside of the steady state in the case that a shock to firm productivity reduces the surplus with specific worker types below zero.

A.2 The Firm's Problem

In the outlined environment, the profit flow of a firm with productivity Ω is the solution to the following Bellman equation. The firm's problem is to maximize output given the current composite labor input and productivity less the total wage bill and hiring costs. Current employment is a state variable. The firm controls its future discounted profits by posting costly vacancies given the expected evolution of its productivity:

$$\Pi(L, \Omega) = \max_{v} \left\{ F(L, \Omega) - \sum_{x=1}^{n} w(x, L, \Omega)L_x - vc(\Omega) + \beta \int \Pi(L', \Omega')\mathrm{d}G(\Omega'|\Omega) \right\}. \quad \text{(A.1)}$$

This profit flow is maximized subject to n constraints that capture the evolution of employment for every worker type x at the firm:

$$L'_x = (1-\delta)L_x + vq_v(\theta)\frac{g_u(x)}{U}\mu(x, L, \Omega) \ \forall x. \tag{A.2}$$

${}^{g_u(x)}/_U$ is the probability that conditional on meeting, the worker is of type x. The indicator function $\mu(x, L, \Omega)$ returns the value one if a match of a type x worker and productivity Ω firm with composite labor input L has a strictly positive surplus and zero otherwise. In the case that $\mu(x, L, \Omega) = 0$, no additional type x workers are hired.

The match surplus is defined as

$$S(x, L, \Omega) = J_x(L, \Omega) + E(x, \Omega) - U(x), \tag{A.3}$$

and depends on the three option value equations defined below. Thus, the indicator $\mu(x, L, \Omega)$ is defined as

$$\mu(x, \Omega) = \begin{cases} 1 & \text{if} \quad S(x, L, \Omega) > 0 \\ 0 & \text{if} \quad S(x, L, \Omega) \leq 0. \end{cases} \tag{A.4}$$

Below, we denote $\mu(x, L, \Omega) = 1$ ($\mu(x, L, \Omega_j) = 0$) by writing $\mu^+(x, L, \Omega)$ ($\mu^-(x, L, \Omega)$).

Optimality Conditions

We closely follow Cahuc et al. (2008) and define the marginal value of an additional worker of type x at a firm with productivity Ω and workforce L as

$$J_x(L, \Omega) = \frac{\partial \Pi(L, \Omega)}{\partial L_x}. \tag{A.5}$$

The marginal product of labor (MPL) for a type x worker at a productivity Ω firm is

$$F_x(L, \Omega) = \frac{\partial F(L, \Omega)}{\partial L_x}. \tag{A.6}$$

The FOC for the maximization problem (A.1) with respect to v is

$$0 = -c(\Omega) + q_v(\theta)\frac{g_u(x)}{U}\mu(x, L, \Omega)J_x(L', \Omega'). \tag{A.7}$$

The envelope theorem implies that

$$J_x(L, \Omega) = \frac{\partial F(L, \Omega)}{\partial L_x} - \sum_{k=1}^{n} L_k \frac{\partial w_k(L, \Omega)}{\partial L_x} - w(x, L, \Omega) + \beta(1-\delta)J_x(L', \Omega'). \tag{A.8}$$

Assuming a steady state where $L' = L$ and $\Omega' = \Omega$, equation (A.7) can be rewritten as

$$J_x(L,\Omega) = \frac{c(\Omega)}{q_v(\theta)\frac{g_u(x)}{U}\mu^{+}(x,L,\Omega)}. \tag{A.9}$$

Therefore, for every worker type within the firm's matching set ($\mu(x,L,\Omega)=1$), marginal profit is equal to the expected recruitment cost at the optimal level of employment. In the case where a type x worker is not part of the firm's matching set, $\mu(x,L,\Omega)=0$, marginal profits are undefined. Integrating the worker type out of (A.9) yields the firm's expected marginal profit from posting a vacancy:

$$J(L,\Omega) = \frac{c(\Omega)}{q_v(\theta)\int \frac{g_u(x)}{U}\mu^{+}(x,L,\Omega)\mathrm{d}x}. \tag{A.10}$$

Applying the steady state assumption to equation (A.8) yields

$$J_x(L,\Omega) = \frac{F_x(L,\Omega) - w(x,L,\Omega) - \sum_{k=1}^{n} L_k \frac{\partial w(k,L,\Omega)}{\partial L_x}}{1-\beta(1-\delta)}, \tag{A.11}$$

so the marginal profit can also be expressed as the discounted marginal product net of the individual wage and net of the effect of the marginal hire on the total wage bill.

Equating (A.9) and (A.11), we obtain

$$F_x(L,\Omega) = w(x,L,\Omega) + \frac{c(\Omega)(1-\beta(1-\delta))}{q_v(\theta)\frac{g_u(x)}{U}\mu(x,L,\Omega)} + \sum_{k=1}^{n} L_k \frac{\partial w(k,L,\Omega)}{\partial L_x}. \tag{A.12}$$

Therefore, the MPL for worker type x at an (L,Ω) firm equals the wage plus the expected turnover costs and the marginal effect of the additional worker on the total wage bill.

A.3 Wage Determination

To derive the wage equation, we rely on the Nash sharing rule

$$\frac{\alpha}{1-\alpha} J_x(L,\Omega) = E(x,L,\Omega) - U(x), \tag{A.13}$$

where $\alpha \in (0,1)$ is the workers' common bargaining parameter. The RHS captures the worker's surplus from working at a firm with productivity Ω and workforce L relative to the worker's outside option, the value of unemployment, $U(x)$. The firm's surplus consists of the marginal profits from hiring an additional worker of type x, $J_x(L,\Omega)$, as defined above. Its threat point is to fire the worker and renegotiate wages with all other employees (Stole and Zwiebel, 1996). Following Cahuc et al. (2008), we assume that wages are continuously and instantaneously (re)negotiated, so L is fixed during (re)negotiations.

In the steady state, the value of employment for the worker is

$$E(x,L,\Omega) = w(x,L,\Omega) + \underbrace{\beta\delta U(x)}_{\text{separation}} + \underbrace{\beta(1-\delta)E(x,L,\Omega)}_{\text{continued employment}}. \tag{A.14}$$

The value of unemployment is

$$\begin{aligned} U(x) = b(x) + \underbrace{\beta(1-q_u(\theta))U(x)}_{\text{no meeting}} + \underbrace{\beta q_u(\theta)\int \frac{g_v(\Omega)}{V}\mu^+(x,L,\Omega)E(x,L,\Omega)\mathrm{d}\Omega}_{\text{successful match}} \\ + \underbrace{\beta q_u(\theta)U(x)\int \frac{g_v(\Omega)}{V}\mu^-(x,L,\Omega)\mathrm{d}\Omega}_{\text{meet unacceptable firm}}, \end{aligned} \tag{A.15}$$

where $b(x)$ is the flow value of unemployment, e.g., the value of increased leisure, home production or unemployment insurance benefits.

Next, we compute the difference $E(x,L,\Omega)-U(x)$ to be plugged into equation (A.13):

$$E(x,L,\Omega) - U(x) = w(x,L,\Omega) + \beta\delta U(x) + \beta(1-\delta)E(x,L,\Omega) - U(x). \tag{A.16}$$

After adding and subtracting $\beta U(x)$, this can be rearranged as

$$E(x,L,\Omega) - U(x) = \frac{w(x,L,\Omega) - (1-\beta)U(x)}{1-\beta(1-\delta)}, \tag{A.17}$$

which can be combined with equation (A.13) to obtain

$$\frac{\alpha}{1-\alpha}J_x(L,\Omega) = \frac{w(x,L,\Omega) - (1-\beta)U(x)}{1-\beta(1-\delta)}. \tag{A.18}$$

Finally, substituting the marginal profits from equation (A.11) and rearranging yields the wage bargaining outcome:

$$w(x,L,\Omega) = \alpha\left(F_x(L,\Omega) - \sum_{k=1}^{n} L_k \frac{\partial w(k,L,\Omega)}{\partial L_x}\right) + (1-\alpha)(1-\beta)U(x). \tag{A.19}$$

Due to our assumption of perfect substitutability of worker ability units at the firm level, the inframarginal adjustment term $\sum_{k=1}^{n} L_k \frac{\partial w(k,L,\Omega)}{\partial L_x}$ reflects solely decreasing returns and is unambiguously negative. Moreover, it does not vary with x. This yields the following simplified differential equation:

$$w(x,L,\Omega) = \alpha\left(F_x(L,\Omega) - L\frac{\partial w(x,L,\Omega)}{\partial L}\right) + (1-\alpha)(1-\beta)U(x), \tag{A.20}$$

which we can solve following the steps for the "single labor case" described in the Appendix to Cahuc et al. (2008). For details, see their equations (B.1)–(B.6), p 961–962. The solution is

$$w(x, L, \Omega) = (1 - \alpha)(1 - \beta)U(x) + \int_0^1 z^{\frac{1-\alpha}{\alpha}} F_x(Lz, \Omega)\mathrm{d}z. \tag{A.21}$$

A.4 Wages, Labor Demand, and Productivity

It is worth considering how wages vary with firm productivity. Partially differentiating the wage equation (A.21) twice with respect to Ω yields

$$\frac{\partial w(x, L, \Omega)}{\partial \Omega} = \int_0^1 z^{\frac{1-\alpha}{\alpha}} \beta_l \frac{F_x(Lz, \Omega)}{\Omega}\mathrm{d}z > 0, \tag{A.22}$$

$$\frac{\partial^2 w(x, L, \Omega)}{\partial \Omega^2} = (\beta_l - 1) \int_0^1 z^{\frac{1-\alpha}{\alpha}} \beta_l \frac{F_x(Lz, \Omega)}{\Omega^2}\mathrm{d}z < 0, \tag{A.23}$$

which implies that wages increase in firm productivity but at a decreasing rate. This model property is a direct consequence of the assumed production function, which exhibits complementarities at the match level; see equation (2). Worker ability and firm productivity jointly determine how much one unit of labor of a given type contributes to output, but the total composite labor input is subject to decreasing returns. In other words, the positive effect of higher productivity on output is dampened by decreasing returns to labor, and this effect carries over into wages. For this reason, the second derivative of the wage with respect to productivity is negative.

Based on the solution to the wage equation in equation (A.21), we can also compute the derivative of the last term of (A.12) for the single labor case:

$$F_x(L, \Omega) = w(x, L, \Omega) + \frac{c(\Omega)(1 - \beta(1 - \delta))}{q_v(\theta)\frac{g_u(x)}{U}\mu(x, L, \Omega)} + (\beta_l - 1) \int_0^1 z^{\frac{1-\alpha}{\alpha} + \beta_l - 1} F_x(Lz, \Omega)\mathrm{d}z. \tag{A.24}$$

This expression pins down labor demand because it equalizes the marginal product and the labor cost, which consists of the wage, turnover costs, and the effect of employment on the wage. The factor that we write in front of the integral is negative when $\beta_l < 1$. This implies that firms can reduce their labor costs by expanding employment in the presence of decreasing returns (overemployment, as in Stole and Zwiebel, 1996; Smith, 1999; Cahuc et al., 2008).

A.5 Linearity of the Wage Equation

Plugging in the worker–firm-specific MPL (3) into equation (A.21) yields

$$w(x, L, \Omega) = (1 - \alpha)(1 - \beta)U(x) + x \int_0^1 z^{\frac{1-\alpha}{\alpha}} \beta_l \frac{F(Lz, \Omega)}{Lz}\mathrm{d}z. \tag{A.25}$$

Therefore, the integral expression is scaled only by, and hence linear in, worker ability x.

To establish that the full wage equation is linear in x, the outside option $U(x)$ must also be linear in x. Consider the outside option according to equation (A.15). A straightforward assumption that ensures that $U(x)$ is indeed linear in x is that all worker types' matching sets cover the whole type space. In other words, conditional on meeting, there are no unacceptable firms. In the model, this implies that $\mu(x, L, \Omega) = 1$ holds for all potential matches, and thus, the last term of equation (A.15) is zero. Now rearrange equation (A.14) such that

$$E(x, L, \Omega) = \frac{w(x, L, \Omega) + \beta\delta U(x)}{1 - \beta(1 - \delta)}. \tag{A.26}$$

Under our assumption, this can be plugged into equation (A.15) to yield an expression in terms of the wage and the outside option only.

$$U(x) = b(x) + \beta(1 - q_u(\theta))U(x) + \beta q_u(\theta) \int \frac{g_v(\Omega)}{V} \frac{w(x, L, \Omega) + \beta\delta U(x)}{1 - \beta(1 - \delta)} \mathrm{d}\Omega. \tag{A.27}$$

Plugging in our solution for the wage, equation (A.25), into this expression and collecting the $U(x)$ terms in front of the integral yields

$$\begin{aligned} U(x) = b(x) + \beta(1 - q_u(\theta))U(x) + \beta q_u(\theta)\frac{1 - \alpha + \beta(\alpha + \delta - 1)}{1 - \beta(1 - \delta)} U(x) \int \frac{g_v(\Omega)}{V} \mathrm{d}\Omega \\ + \frac{\beta q_u(\theta)}{1 - \beta(1 - \delta)} \int \frac{g_v(\Omega)}{V} x \int_0^1 z^{\frac{1-\alpha}{\alpha}} \beta_l \frac{F(Lz, \Omega)}{Lz} \mathrm{d}z \mathrm{d}\Omega, \end{aligned} \tag{A.28}$$

where $\int \frac{g_v(\Omega)}{V} \mathrm{d}\Omega = 1$. After collecting all $U(x)$ terms on the LHS and dividing, we obtain the following expression for $U(x)$:

$$U(x) = \frac{(1 - \beta(1 - \delta))b(x) + x\frac{\beta q_u(\theta)}{1-\beta} \int \frac{g_v(\Omega)}{V} \int_0^1 z^{\frac{1-\alpha}{\alpha}} \beta_l \frac{F(Lz,\Omega)}{Lz} \mathrm{d}z\mathrm{d}\Omega}{1 - \beta(1 - \delta - \alpha q_u(\theta))}, \tag{A.29}$$

where worker ability x can be written in front of both integral signs. Thus, for $U(x)$ to be linear in x, we must additionally assume that the workers' flow value of unemployment, $b(x)$, is proportional to x (i.e., $b(x) \propto x$), which is a standard assumption used in, e.g., Postel-Vinay and Robin (2002). Note that $U(x) = x\bar{U}$ is log-additive in x by the product rule: $\ln U(x) = \ln x + \ln \bar{U}$. This also applies to the integral in the wage equation (A.25).

B Details of Data Preparation

B.1 Wage Imputation

In the BeH data, earnings are right-censored at the contribution assessment ceiling ("Beitragsbemessungsgrenze"). This earnings limit is given by the statutory pension fund and is adjusted annually due to changes in earnings. In each year, we identify censored wage observations by comparing wages to the contribution assessment ceiling. We define a wage observation as censored whenever the reported wage is higher than 99% of the censoring thresholds. Following CHK and Dustmann et al. (2009), we fit a series of tobit regressions to impute the right tail of the wage distribution. We estimate the tobit regressions by year, sex, education and age group. In all these regressions, we also control for exact age, the mean log wage in other years, the fraction of censored wages in other years, the number of full-time employees at the current establishment and its square, an indicator for large firms, mean years of schooling and the fraction of university graduates at the current establishment, the mean log wage of coworkers and the fraction of coworkers with censored wages, an indicator for individuals observed in only one year, an indicator for employees in one-worker establishments, and an indicator for region. We assume that the error term is normally distributed, but each education and age category can have a different variance. For each year, we impute the censored wages as the sum of the predicted wage and a random component that is computed based on the standard error of the forecast. This component is drawn from separate normal distributions with mean zero and the different variances for each education and age category.

B.2 Education Imputation

Employee education information is reported by employers each year and whenever a job ends. Employers do not face consequences for nonreporting or misreporting education. The existence of a reporting rule allows for corrections. It prescribes that only the highest educational degree of an employee needs to be reported. Therefore, individual educational attainment should not decline over consecutive periods. We follow the procedure suggested by Fitzenberger et al. (2006), which exploits this reporting rule by assuming that there is some overreporting in the data. The original education variable in the BeH defines six categories, plus an additional category for missing information. Following CHK, we convert these into the following categories: (0) missing; (1) primary/lower secondary or intermediate school leaving certificate, or equivalent, with no vocational training; (2) primary/lower secondary or intermediate school leaving certificate, or equivalent, with vocational training; (3) upper secondary school certificate with or without vocational training; and (4) some university degree. Within each job, we assign the modal education category observed for an individual during the years that he/she is at the same job.

C Production Function Estimation Details

We start with the production function to be estimated, which corresponds to equation (9). Lower-case letters indicate logs, and here, we ignore firm-level controls.

$$y_{jt} = \beta_0 + \beta_l l_{jt} + \beta_k k_{jt} + \omega_{jt} + \epsilon_{jt}, \tag{C.1}$$

Following ACF, the control function that we use is the demand for intermediate inputs *conditional on labor*[36] and capital:

$$m_{jt} = f_t(l_{jt}, k_{jt}, \omega_{jt}). \tag{C.2}$$

A natural interpretation of this demand function is that conditional on both labor and capital, more productive firms use more intermediate goods in production. Intermediate inputs m_{jt} are chosen either simultaneously with the labor input l_{jt} or afterwards, reflecting the fact that labor is a dynamic input.

Under the assumption of strict monotonicity in ω_{jt}, we can invert equation (C.2) to obtain $\omega_{jt} = f_t^{-1}(l_{jt}, k_{jt}, m_{jt})$, and substitute ω_{jt} out of (C.1):

$$y_{jt} = \beta_0 + \beta_l l_{jt} + \beta_k k_{jt} + f_t^{-1}(l_{jt}, k_{jt}, m_{jt}) + \epsilon_{jt} = \Phi_t(l_{jt}, k_{jt}, m_{jt}) + \epsilon_{jt}. \tag{C.3}$$

We adopt the two-stage procedure proposed by ACF. First, value added is regressed on a polynomial approximation of $\Phi_t(l_{jt}, k_{jt}, m_{jt})$. This does not identify any of the parameters but delivers the estimate $\hat{\Phi}_t(l_{jt}, k_{jt}, m_{jt})$ and separates productivity ω_{jt} from the transitory shocks absorbed by ϵ_{jt}.

To derive the second-stage estimation equation, we use the assumption that firm productivity follows an AR(1) process (equation 11). Thus, the conditional expectation of ω_{jt} formed at $t-1$ is $E\left[\omega_{jt}|\omega_{j,t-1}\right] = \rho\omega_{j,t-1}$, and the productivity innovation is denoted ξ_{jt}. Additionally, the following hold: $\omega_{j,t-1} = y_{j,t-1} - \beta_0 - \beta_l l_{j,t-1} - \beta_k k_{j,t-1} - \epsilon_{j,t-1}$ (from equation C.1) and $\epsilon_{j,t-1} = y_{j,t-1} - \hat{\Phi}_{t-1}(l_{j,t-1}, k_{j,t-1}, m_{j,t-1})$ (from equation C.3 with the first-stage estimate plugged in). Using these expressions along with the conditional expectation of firm productivity in equation (C.1) yields:

$$y_{jt} = \beta_0 + \beta_l l_{jt} + \beta_k k_{jt} + \rho(\hat{\Phi}_{t-1}(.) - \beta_0 - \beta_l l_{jt-1} - \beta_k k_{jt-1}) + \xi_{jt} + \epsilon_{jt}, \tag{C.4}$$

where $\xi_{jt} + \epsilon_{jt}$ is a composite error term. Following ACF, we use GMM and four (uncon-

[36] ACF use the conditional (on labor) input demand function to address the problem of functional dependence and improve identification of the labor input parameter relative to Olley and Pakes (1996) and Levinsohn and Petrin (2003).

ditional) second-stage moment conditions to identify the four parameters $(\beta_0, \beta_l, \beta_k, \rho)$:

$$E\left[\left(\underbrace{y_{jt} - \beta_0 - \beta_l l_{jt} - \beta_k k_{jt} - \rho(\hat{\Phi}_{t-1}(.) - \beta_0 - \beta_l l_{jt-1} - \beta_k k_{jt-1})}_{\xi_{jt}+\epsilon_{jt}}\right) \otimes \begin{pmatrix} 1 \\ k_{jt} \\ \hat{\Phi}_{t-1}(.) \\ l_{j,t-1} \end{pmatrix}\right] = 0 \tag{C.5}$$

The first three moment conditions are standard in the literature and identical to those in Levinsohn and Petrin (2003): (i) The composite error term $\xi_{jt} + \epsilon_{jt}$ is zero in expectation. (ii) The capital stock is predetermined—recall equation (10)—and therefore uncorrelated with the error term. (iii) The polynomial representation of the control function, estimated in the first stage and evaluated at $t-1$, $\hat{\Phi}_{t-1}$, is also uncorrelated with the composite error term because the intermediate input choice during the last period should be uncorrelated with the current productivity innovation and the transitory shock.

The fourth moment condition is specific to the ACF approach and reflects the fact that labor is a dynamic input. Only $l_{j,t-1}$ needs to be uncorrelated with the error term, while l_{jt} is allowed to covary freely with ξ_{jt}. In other words, labor is chosen in response to the current firm productivity realization, subject to across-firm differences in adjustment costs and wage setting (firm-specific, serially correlated, unobserved shocks to the price of labor). The conditional intermediate input demand (that is, conditional on l_{jt}) does not depend on these costs/shocks. The ACF approach does not allow for shocks to the price of intermediate inputs (due to the unobservable scalar assumption, which is required for the inversion of $f_t(l_{jt}, k_{jt}, \omega_{jt})$).[37] Olley and Pakes (1996) cannot allow for any input price shocks, Levinsohn and Petrin (2003) can allow for shocks related to k_{jt} but not to l_{jt} and m_{jt}, and ACF can allow for shocks related to k_{jt} and l_{jt} but not to m_{jt}.

Finally, note that the exclusion of m_{jt} from the production function implies that we estimate a value added production function (value added is simply defined as revenue minus the cost of intermediate inputs). One structural interpretation of this fact is that the production function is Leontief in the intermediate input, and this intermediate input needs to be proportional to output. As explained by ACF, their approach is designed for value added production functions and is not suitable for identifying the parameters of gross output production functions without further assumptions (see also Bond and Söderbom (2005) and Gandhi et al. (2020)).

[37] Arguably, intermediate inputs are often commodities, implying only little price variation across firms.

D Additional Results

D.1 Wage Regressions

We estimate a log-linear wage equation for worker i who works at firm $j(i,t)$ in year t:

$$w_{it} = \alpha_i + \psi_{j(i,t)} + x_{it}'\beta + \varepsilon_{it}. \tag{C.6}$$

w_{it} represents log real daily wages, α_i represents the worker fixed effects, $\psi_{j(i,t)}$ represents the firm fixed effects, and x_{it}' contains time-varying controls (an unrestricted set of year dummies, quadratic/cubic terms for age interacted with education). ε_{it} is the residual.

Table D.1: Wage Variance Decompositions

	(a) Regression (C.6) BeH, full		(b) Regression (C.6) BeH, women		(c) Regression (C.6) BeH, men		(d) Regression (C.6) BeH, men, West	
Var(w_{it})	0.276	(100%)	0.277	(100%)	0.245	(100%)	0.226	(100%)
Var($\hat{\alpha}_i$)	0.126	(46%)	0.138	(50%)	0.105	(43%)	0.106	(47%)
Var($\hat{\psi}_{j(i,t)}$)	0.068	(25%)	0.076	(27%)	0.061	(25%)	0.049	(22%)
Var($x_{it}'\hat{\beta}$)	0.005	(2%)	0.006	(2%)	0.005	(2%)	0.005	(2%)
$2 \times$ Cov($\hat{\alpha}_i, \hat{\psi}_{(i,t)}$)	0.049	(18%)	0.032	(12%)	0.047	(19%)	0.037	(16%)
$2 \times$ Cov($\hat{\alpha}_i, x_{it}'\hat{\beta}$)	0.004	(0%)	0.001	(0%)	0.005	(2%)	0.006	(3%)
$2 \times$ Cov($\hat{\psi}_{j(i,t)}, x_{it}'\hat{\beta}$)	0.003	(0%)	0.001	(0%)	0.004	(2%)	0.004	(2%)
Var($\hat{\varepsilon}_{it}$)	0.021	(8%)	0.025	(9%)	0.018	(7%)	0.019	(8%)
Sample mean wage	4.450		4.261		4.553		4.621	
R^2	0.92		0.93		0.91		0.92	
# Observations	233,117,492		82,267,794		150,849,698		123,087,610	

Source: Authors' calculations based on the BeH.
Note: Variance decompositions for log real daily wages based on regression model (C.6) for various BeH samples. Mean wages, variances, and covariances are rounded to three decimal places.

Table D.1 shows the variance decompositions for multiple groups of workers based on the estimated AKM wage components. Column (a) includes all person-years, (b) all women, (c) all men, and (d) all men in West Germany (the CHK sample). We replicate the well-known finding that the majority of the wage variance is explained by unobserved worker heterogeneity; i.e., the worker fixed effect explains almost half of the observed variation in wages (slightly more for women and slightly less for men). The second-most important source of variation is the firm fixed effects, which explain roughly one-quarter of the wage variation across the four groups. The third largest determinant is the covariance between the worker and firm effects, which explains between 12% and 19% of the wage variance.[38] At only 2%, the share of the wage variance explained by the time-varying observable characteristics is almost negligible. The same is true for the covariances of the observable characteristics with the worker and firm effects. Note that time-invariant covariates such as education are absorbed by the worker effect. The residual explains between 7% and 9% of the wage variance across the four BeH samples. Estimated worker and firm fixed effects, as well as their covariance, could be biased due to limited mobility in the connected set.[39] We apply the parametric correction suggested by Andrews et al. (2008) for two subperiods in our data and find that the limited mobility bias is small.[40]

D.2 Comparison with Alternative Rankings

We compare our productivity ranking with the results of other ranking techniques used in the literature on wage dispersion and labor market sorting. We create two alternative firm rankings. The first is based on the AKM firm fixed effects and the second on the poaching index used in Bagger and Lentz (2019) and Taber and Vejlin (2020).[41] Correlations with the productivity ranking are positive, i.e., 0.280 and 0.119. To graphically analyze how the rankings are related, we create two sets of 15-firm bins based on both the firm fixed effects ranking and the poaching ranking. This approach allows us to compute the empirical distribution of firm-years across the different firm bin combinations. We observe how firms with a given productivity rank are distributed across the firm fixed effect and poaching rank bins. Figure D.1 plots the contours of these empirical distributions.

In Panel (a), the AKM firm fixed effects bins are depicted on the vertical axis, and the ω ranking bins are depicted on the horizontal axis. The mass of observations is concentrated along the diagonal, which is in line with the positive correlation reported above. In the upper-right quadrant of the plot, the observations are highly concentrated somewhat above the diagonal, thereby reaffirming our observation that the highest-paying firms (here, in terms of the AKM wage premia) are located below the top of the productivity

[38]Interestingly, women are less positively sorted in terms of wages than men. This is in line with what Card et al. (2016) and Bruns (2019) find using Portuguese and German data, respectively.

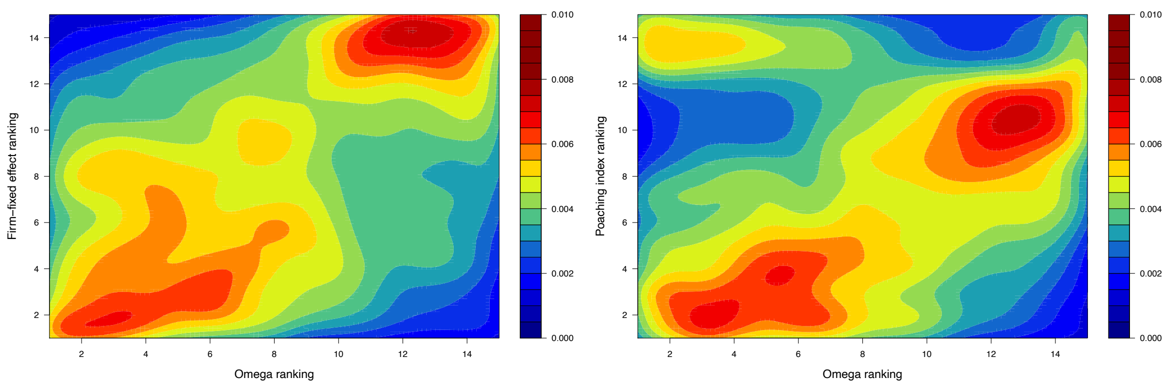
[39]The estimated variances of worker and firm fixed effects could be biased upward, and their covariance biased downward (Andrews et al., 2008, 2012; Borovičková and Shimer, 2020; Kline et al., 2020).

[40]The fact that we estimate the AKM model on the universe of German register data over a ten-year timeframe and include both men and women mitigates concerns regarding severe limited mobility bias due to the large number of observed transitions. Lochner et al. (2020) and Lachowska et al. (Forthcoming) confirm that this bias is limited in data sets with a large number of transitions. The Andrews et al. (2008) correction implies that in our setting, the variance in worker effects decreases by 5% (4%) and the variance in firm effects decreases by 4% (3%) in the 1998–2002 (2003–2008) subperiod. The covariance between the worker and firm effects increases by 7% (5%).

[41]The poaching index is based on the idea that high-paying firms poach workers from other firms rather than hiring unemployed workers. We compute this index by comparing the annual number of workers hired directly from other firms to the number of all hires at the firm level and then rank firms based on the firm-level mean of their time-varying poaching index. We use the Administrative Wage and Labor Market Flow Panel (AWFP); see Stüber and Seth (2019). In this data set, the aggregated establishment-level worker flows needed to compute the poaching index are readily available.

Figure D.1: Comparison with Alternative Firm Rankings

(a) Firm fixed effects (b) Poaching index



Source: Authors' calculations based on the BHP, EP, BeH.
Note: The two plots depict the contours of the joint empirical distribution of firm-years across combinations of the ω bins with the AKM firm-effects bins (Panel (a)) and the poaching index bins (Panel (b)).

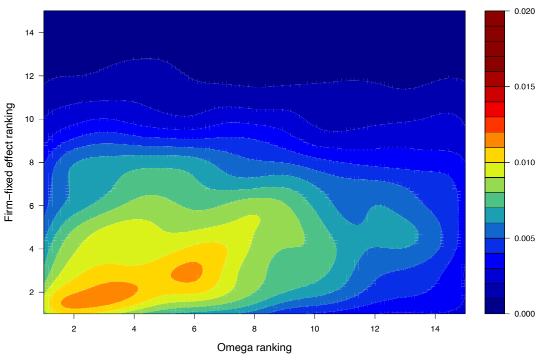
distribution. In the lower-left quadrant, the observations are more dispersed. It is not uncommon to observe firm-years in which the estimated AKM wage premium is around the median but estimated productivity is very low and vice versa. Here, the disagreement between the two rankings is large. Figure D.2 shows that the observations in the lower-left quadrant are mainly young and small firms. The high-wage firms in the upper-right quadrant are older and larger firms. Panels (a) and (b) in Figure D.2 show that low-wage (high-wage) firms have, on average, a low (high) AKM rank; however, there is a sizable overlap between the two groups in terms of productivity.

In Panel (b), the mass of observations lies below the diagonal; that is, a firm's poaching rank tends to be lower than its productivity rank. Many high-productivity firms have high poaching ranks, but they are almost never located at the top. Additionally, it is not uncommon even for medium-productivity firms to hire mainly OON, as evidenced by the high density of low-poaching rank firms that extends far to the right. Interestingly, the firm years with the highest poaching index are clustered in the upper-left quadrant. Apparently, some low-productivity firms very actively poach workers from other firms. Figure D.3 shows that many of the firms at the top of the poaching index distribution are small and young. One possible explanation is that these firms attempt to grow quickly by poaching workers from other firms. The larger and older firms, which also pay the highest wages, are concentrated in the upper-middle portion of the poaching rank distribution. They hire a nonnegligible number of employees OON.

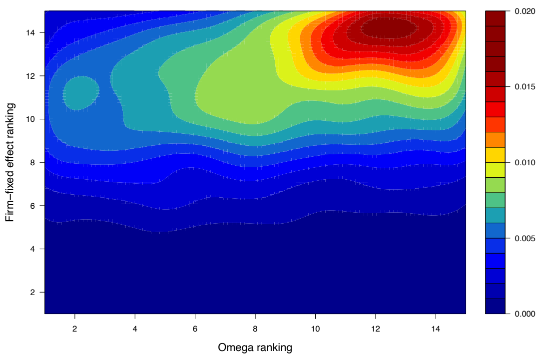
In summary, the comparison of the different rankings shows that firm ranks based on firm wage premia and observed worker mobility are systematically different from our productivity-based firm ranking.

Figure D.2: Comparison of Productivity-Based and Fixed Effect-Based Firm Rankings by Wages, Age, and Size

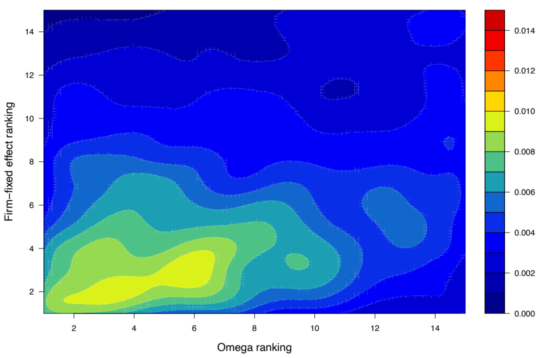
(a) Low-wage firms



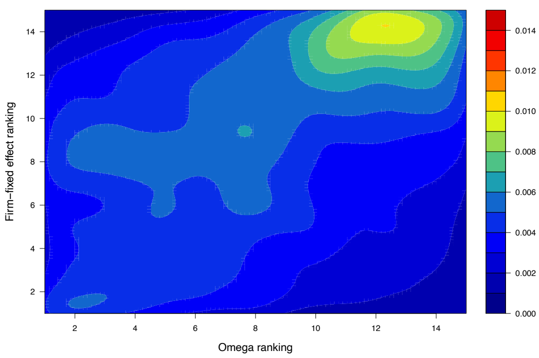
(b) High-wage firms



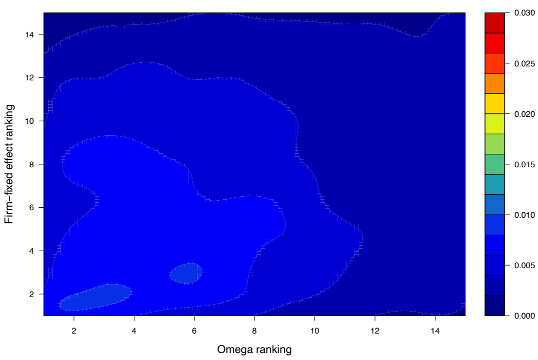
(c) Young firms



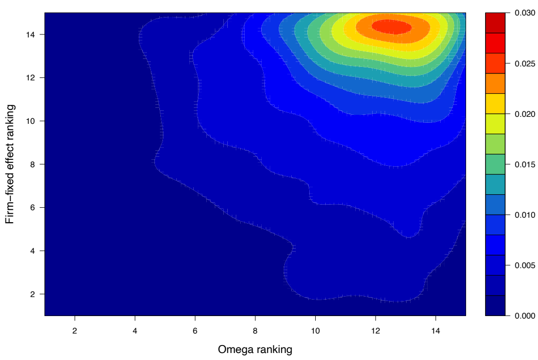
(d) Old firms



(e) Small firms



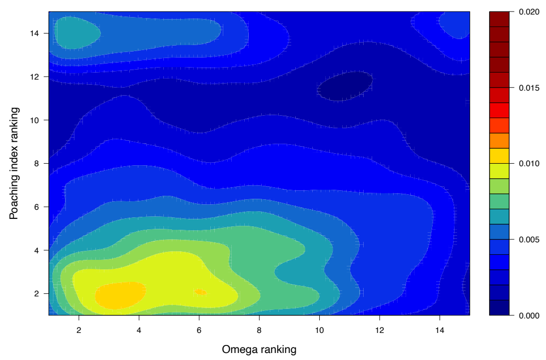
(f) Large firms



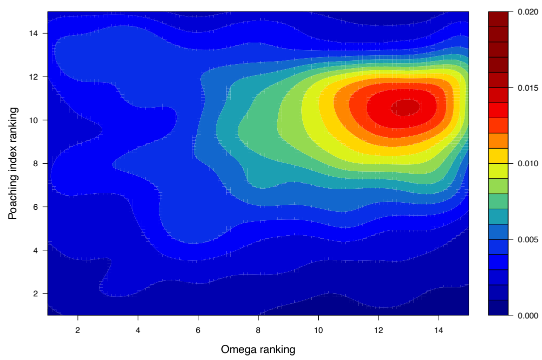
Source: Authors' calculations based on the BHP, EP, BeH.
Note: These six plots depict the contours of the joint empirical distributions of firm-years across combinations of the Ω ranking (15 bins) and AKM firm fixed effect ranks (15 bins). In Panels (a) and (b), high-wage firms are those that pay more than the grand mean of all firm-level mean wages, and low-wage firms are those that pay less. In Panels (c) and (d), young firms are those that are less than 15 years old, and old firms are those that are 15 years or older. In Panels (e) and (f), small firms are those with fewer than 100 employees, and large firms are those with more.

Figure D.3: Comparison of Productivity-Based Firm Ranking and Poaching Index-Based Firm Ranking by Wages, Age, and Size

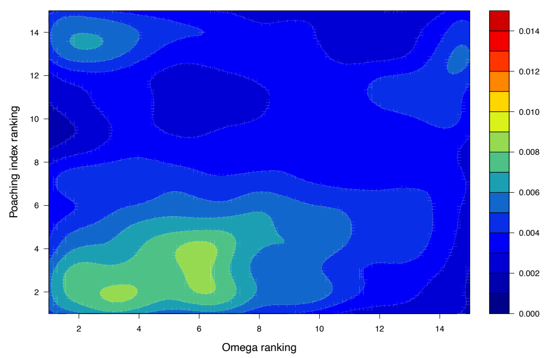
(a) Low-wage firms



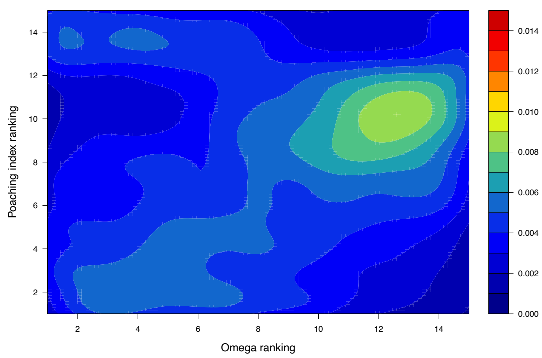
(b) High-wage firms



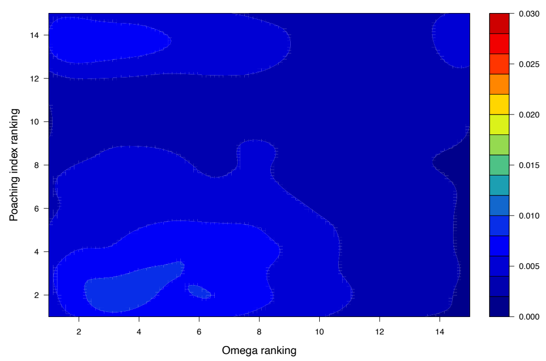
(c) Young firms



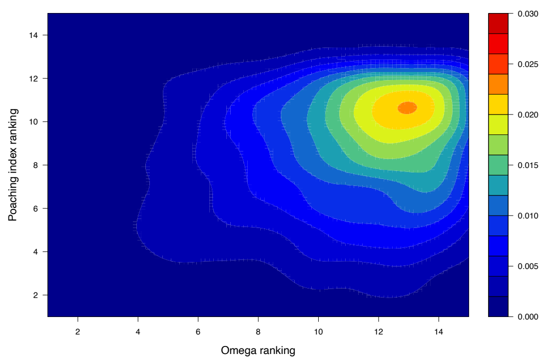
(d) Old firms



(e) Small firms



(f) Large firms



Source: Authors' calculations based on the BHP, EP, BeH.
Note: These six plots depict contours of the joint empirical distributions of firm-years across combinations of the Ω ranking (15 bins) and poaching index ranks (15 bins). In Panels (a) and (b), high-wage firms are those that pay more than the grand mean of all firm-level mean wages, and low-wage firms pay less. In Panels (c) and (d), young firms are those that are less than 15 years old, and old firms are those that are 15 years or older. In Panels (e) and (f), small firms are those with fewer than 100 employees, and large firms are those with more.

D.3 Rank Correlations

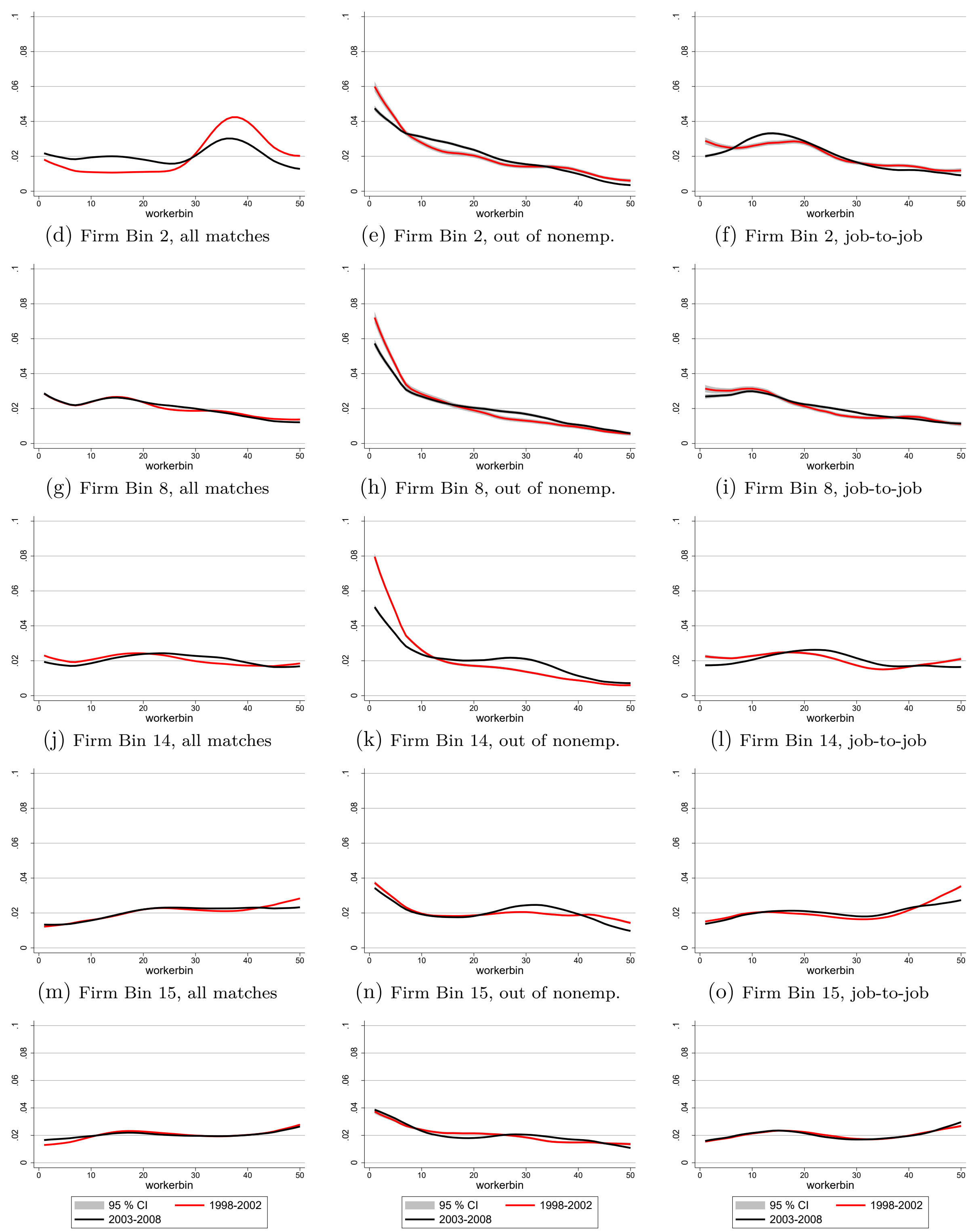
Table D.2: Spearman Rank Correlation Coefficients and Numbers of Observations for Different Time Intervals and Samples

	All Matches	New Matches	Out of Nonemp.	Job-to-Job
1998-2008	0.065 (4,695,108)	0.124 (1,656,280)	0.132 (601,954)	0.110 (1,082,460)
1998-2002	0.055 (2,182,011)	0.139 (474,341)	0.141 (174,310)	0.120 (305,339)
2003-2008	0.074 (2,513,097)	0.118 (1,181,939)	0.129 (427,644)	0.107 (777,121)
1998	0.013 (311,861)	–	–	–
1999	0.046 (338,125)	0.140 (35,865)	0.133 (15,094)	0.129 (20,771)
2000	0.048 (493,323)	0.107 (107,740)	0.108 (41,731)	0.090 (66,009)
2001	0.073 (536,559)	0.148 (158,351)	0.142 (56,627)	0.137 (101,724)
2002	0.077 (502,143)	0.152 (172,385)	0.165 (60,777)	0.134 (111,608)
2003	0.080 (470,279)	0.146 (180,623)	0.156 (64,926)	0.131 (115,697)
2004	0.065 (458,467)	0.114 (191,207)	0.129 (67,932)	0.100 (123,275)
2005	0.081 (428,122)	0.129 (197,755)	0.136 (69,890)	0.118 (127,865)
2006	0.101 (415,153)	0.146 (206,984)	0.139 (74,340)	0.142 (132,644)
2007	0.051 (391,535)	0.084 (209,567)	0.096 (77,811)	0.074 (131,756)
2008	0.063 (349,541)	0.094 (195,803)	0.119 (72,001)	0.076 (123,802)

Source: Authors' calculations based on the BHP, EP, BeH.
Note: In all cells, we test the null hypothesis that the worker and firm bins are statistically independent. All rank correlation coefficients are different from 0 at the 1% level of significance. Results are rounded to 3 decimal places. Numbers of observations (matches according to the respective definition) are reported in brackets.

D.4 Distributional Dynamics

Figure D.4: Changes in the Worker Type Distributions within Different Firm Bins



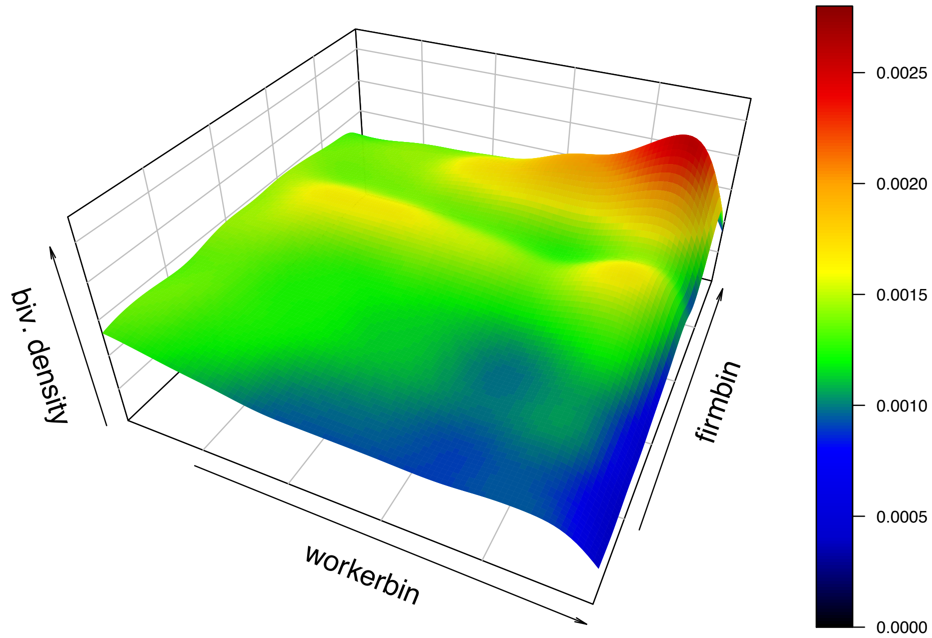
(a) Firm Bin 1, all matches
(b) Firm Bin 1, out of nonemp.
(c) Firm Bin 1, job-to-job
(d) Firm Bin 2, all matches
(e) Firm Bin 2, out of nonemp.
(f) Firm Bin 2, job-to-job
(g) Firm Bin 8, all matches
(h) Firm Bin 8, out of nonemp.
(i) Firm Bin 8, job-to-job
(j) Firm Bin 14, all matches
(k) Firm Bin 14, out of nonemp.
(l) Firm Bin 14, job-to-job
(m) Firm Bin 15, all matches
(n) Firm Bin 15, out of nonemp.
(o) Firm Bin 15, job-to-job

Source: Authors' calculations based on the BHP, EP, BeH.
Note: Univariate kernel densities conditional on worker bin, time, and match type. Kernel: Epanechnikov. Bandwidth: Silverman's rule. Pointwise confidence intervals are calculated using quantiles of the standard normal distribution.

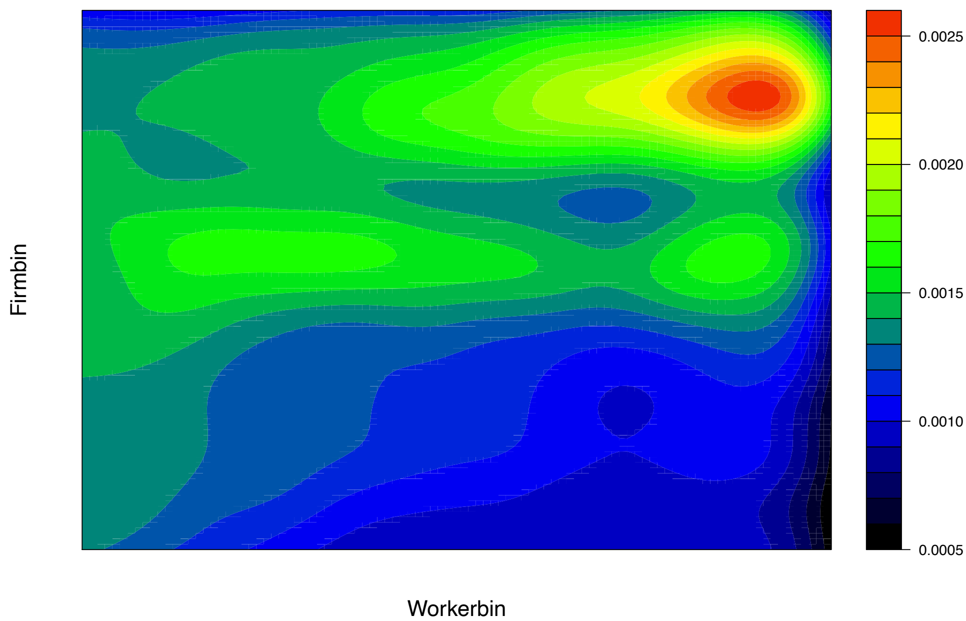
D.5 Joint Distribution of Matches

Figure D.5: Joint Distribution of All Worker-Firm Type Combinations (1998–2008)

(a) Joint Density



(b) Projection on the ($\bar{x}$, $\bar{\omega}$) plane

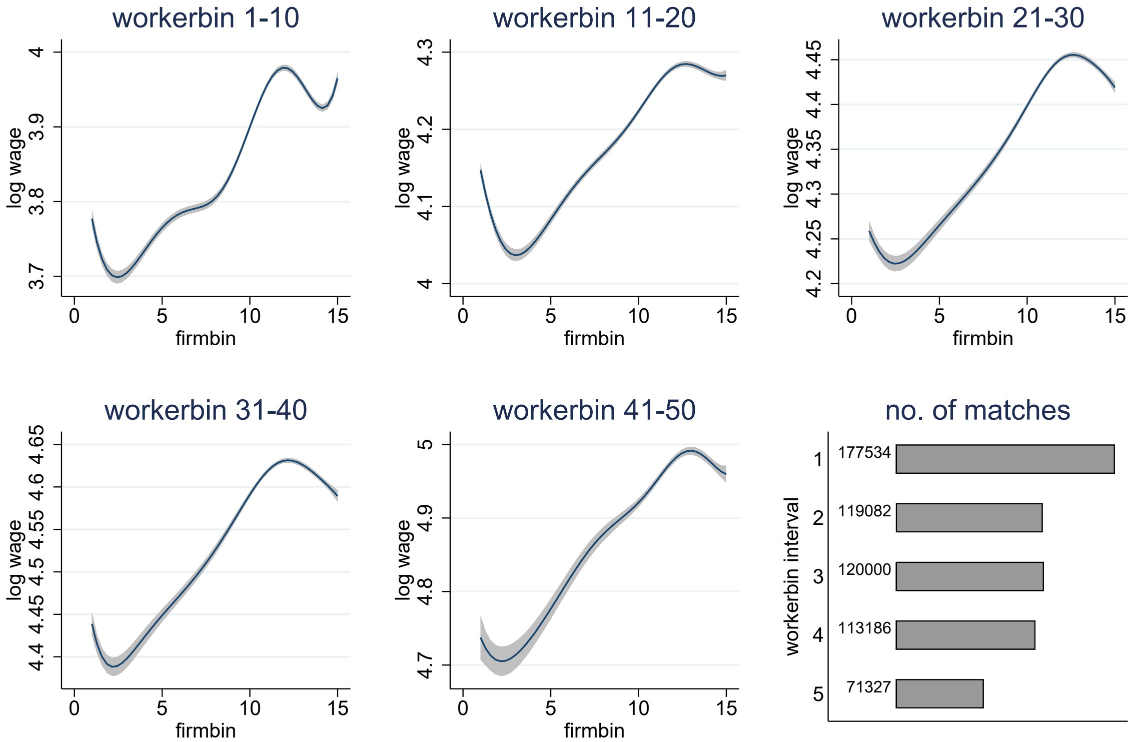


Source: Authors' calculations based on the BHP, EP, BeH.
Note: The plots show the estimated joint kernel density of matches and its projection onto the ($\bar{x}$, $\bar{\omega}$ plane for all combinations of worker and firm types in the sample of all matches on a grid with dimensions 50×15 (#worker types $\times$ #firm types).

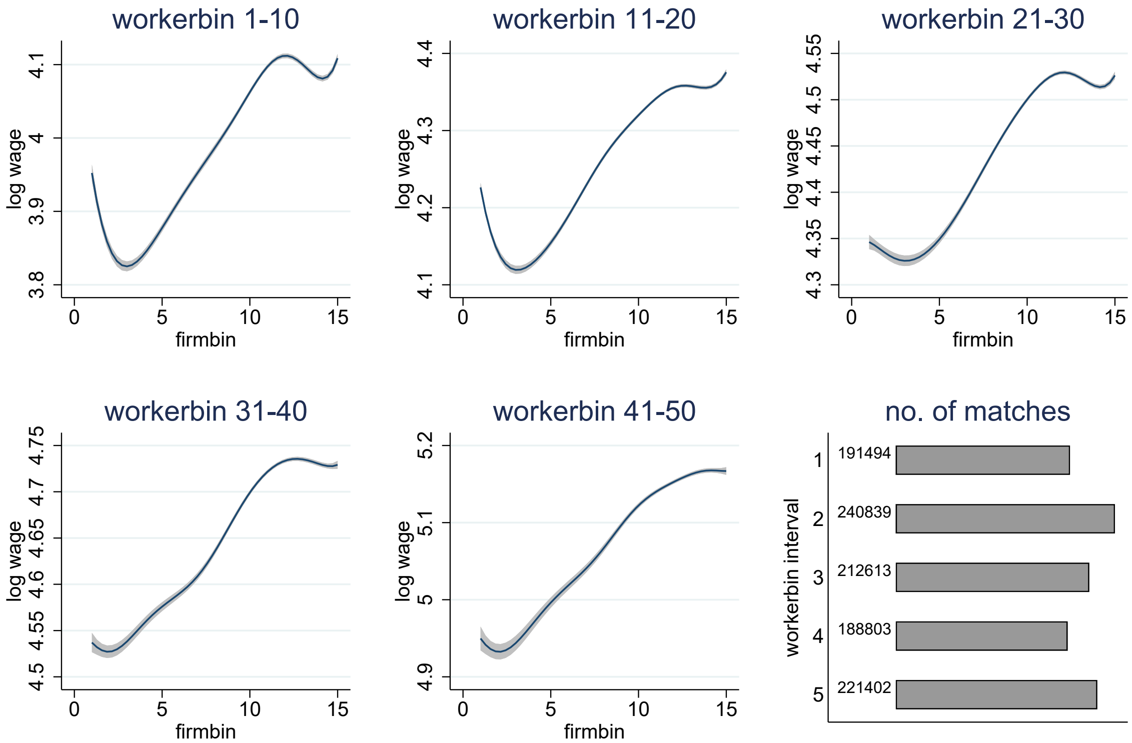
D.6 Wages and Transitions

Figure D.6: Wage–Productivity Profiles, New Matches

(a) Out of Nonemployment



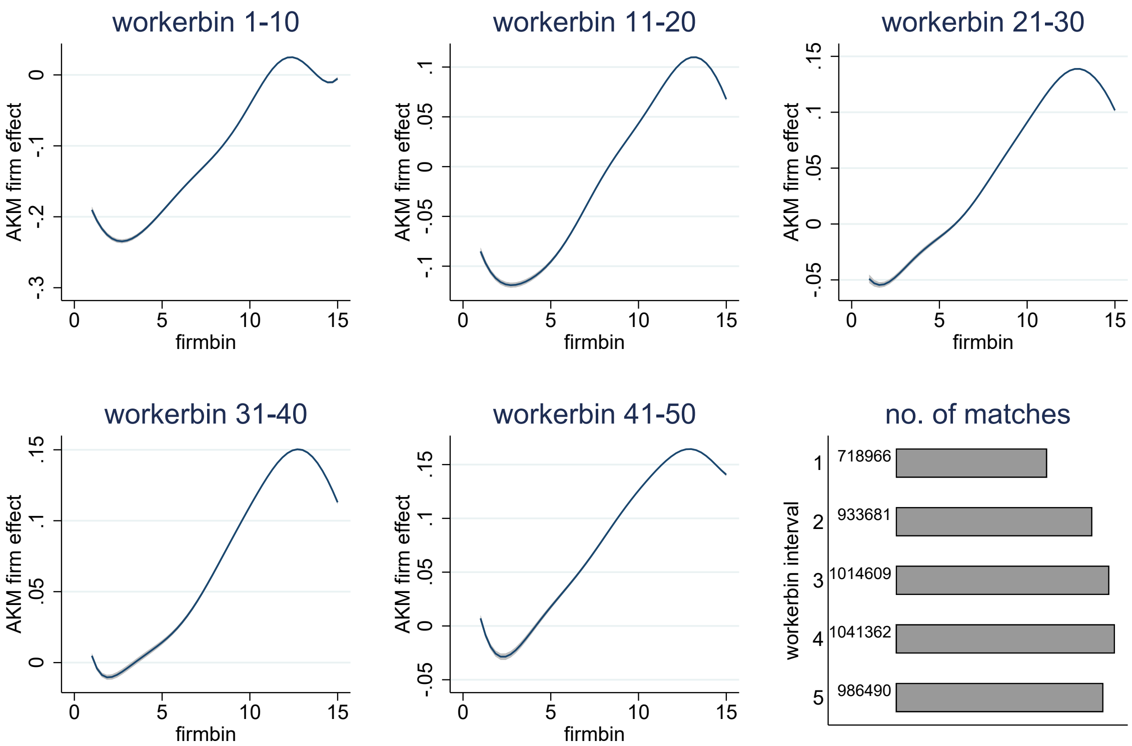
(b) Job-to-Job



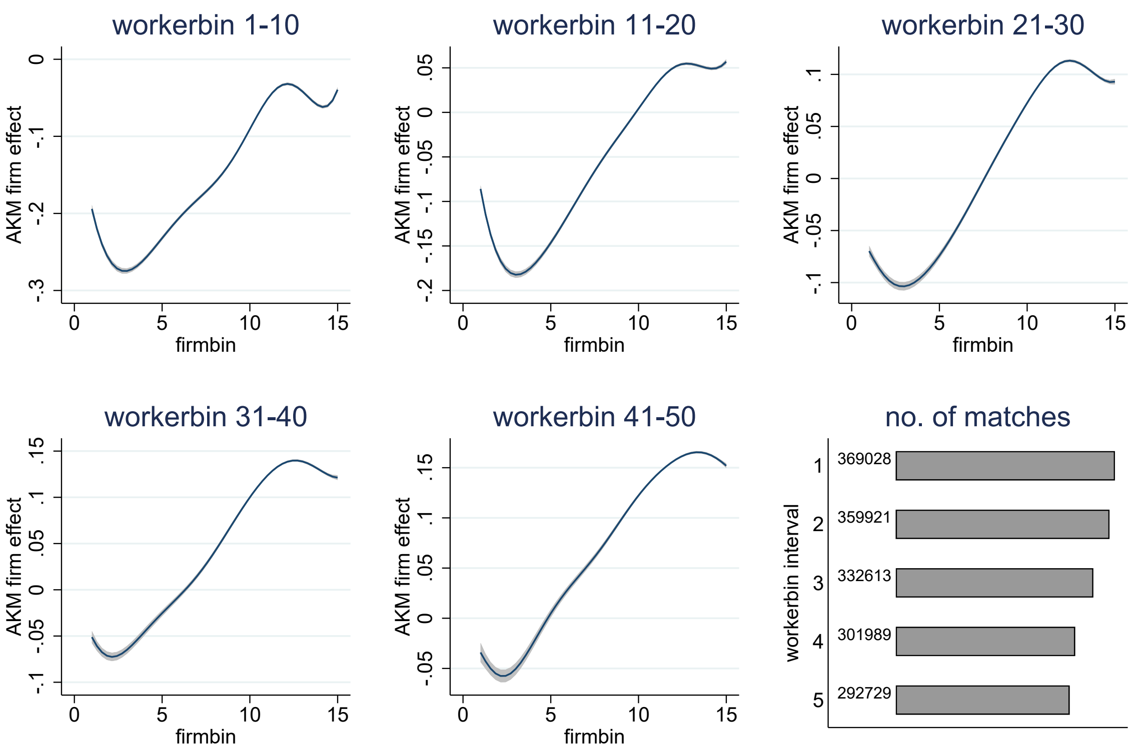
Source: Authors' calculations based on the BHP, EP, BeH.
Note: Plots show the estimated wage–productivity profiles for new matches out of nonemployment and from job-to-job. Based on kernel-weighted local polynomial regressions. Kernel: Gaussian. Bandwidth: 2. 95% confidence bands in gray.

Figure D.7: Wage–Productivity Profiles, Firm Fixed Effects Only

(a) All Matches



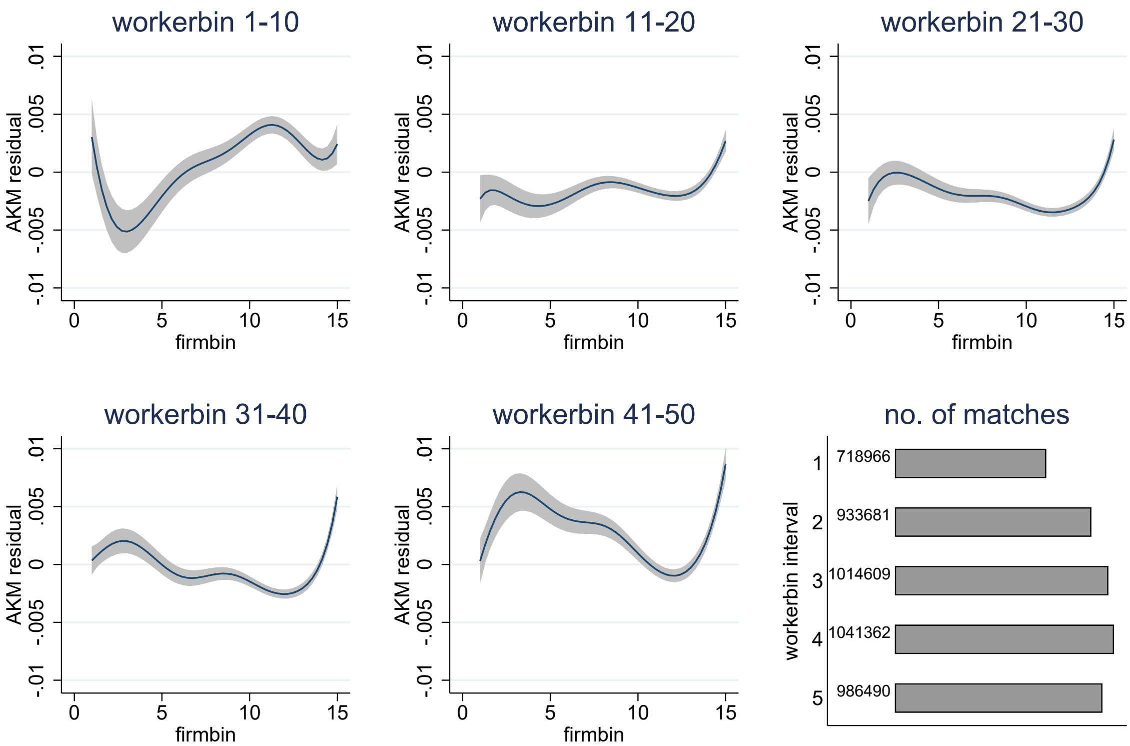
(b) New Matches



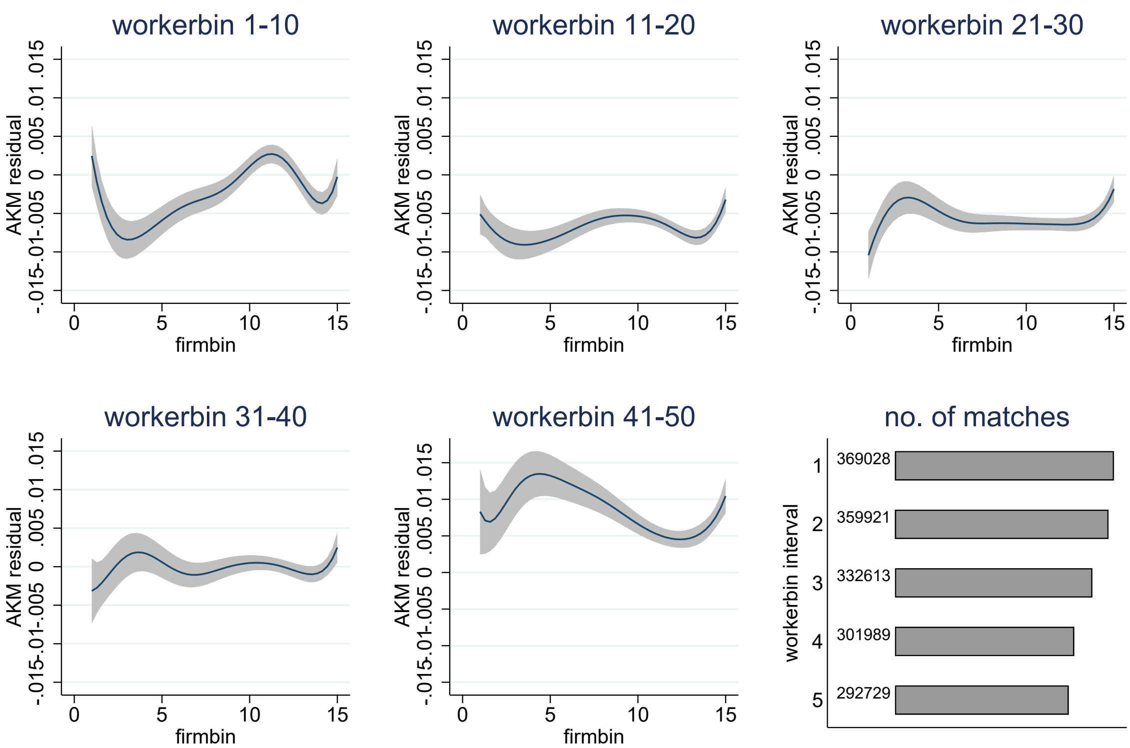
Source: Authors' calculations based on the BHP, EP, BeH.
Note: Plots show the estimated wage–productivity profiles across firm bins for all matches and new matches when using the AKM firm fixed effects as the wage variable. Based on kernel-weighted local polynomial regressions. Kernel: Gaussian. Bandwidth: 2. 95% confidence bands in gray.

Figure D.8: Wage–Productivity Profiles, Residuals Only

(a) All Matches



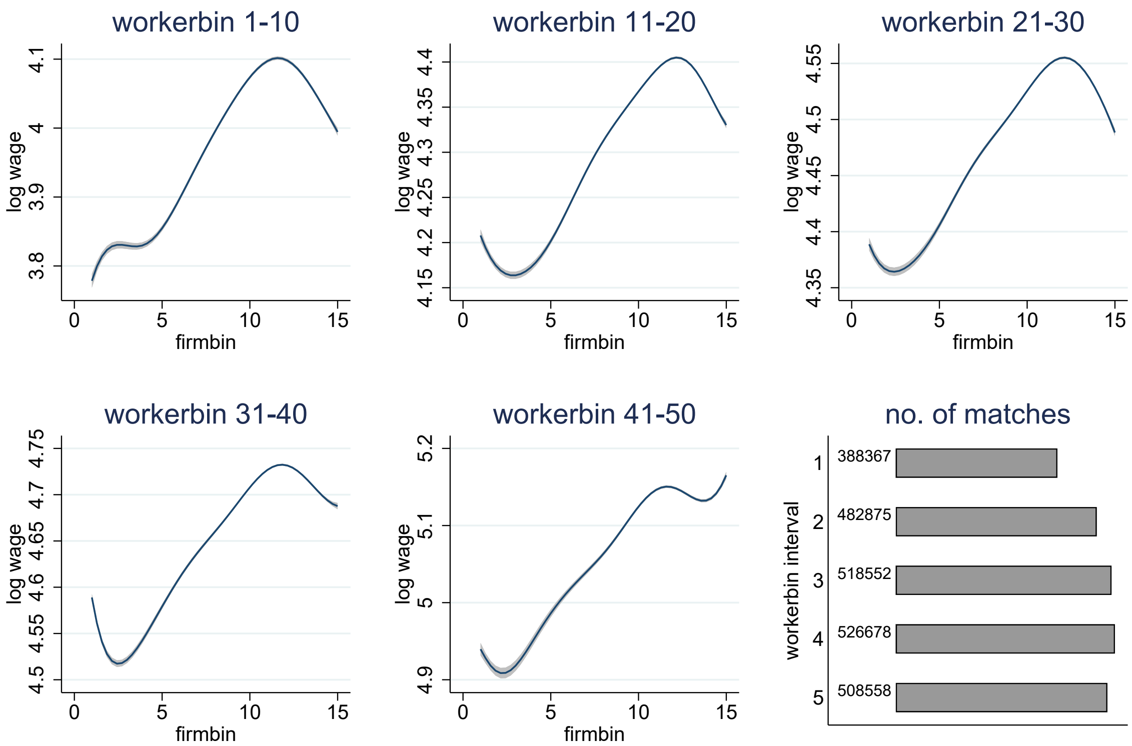
(b) New Matches



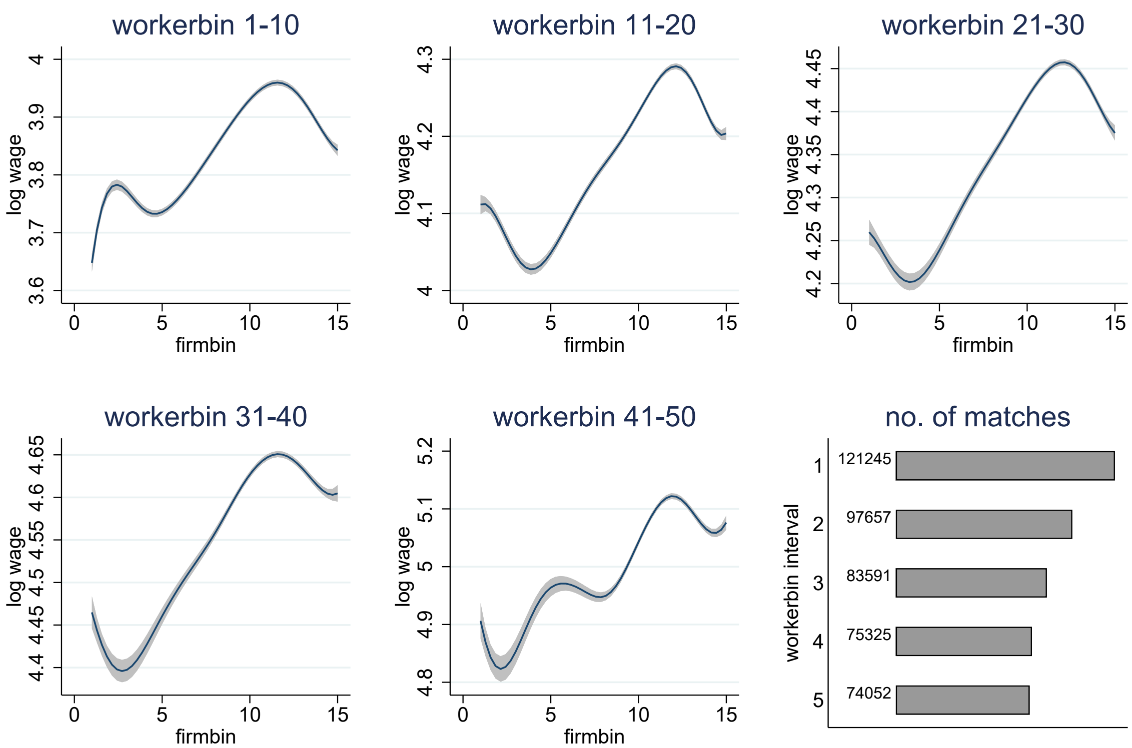
Source: Authors' calculations based on the BHP, EP, BeH.
Note: Plots show the estimated wage–productivity profiles across firm bins for all matches and new matches when using the AKM residuals as the wage variable. Based on kernel-weighted local polynomial regressions. Kernel: Gaussian. Bandwidth: 2. 95% confidence bands in gray.

Figure D.9: Wage–Productivity Profiles, Worker Subsamples

(a) All Matches



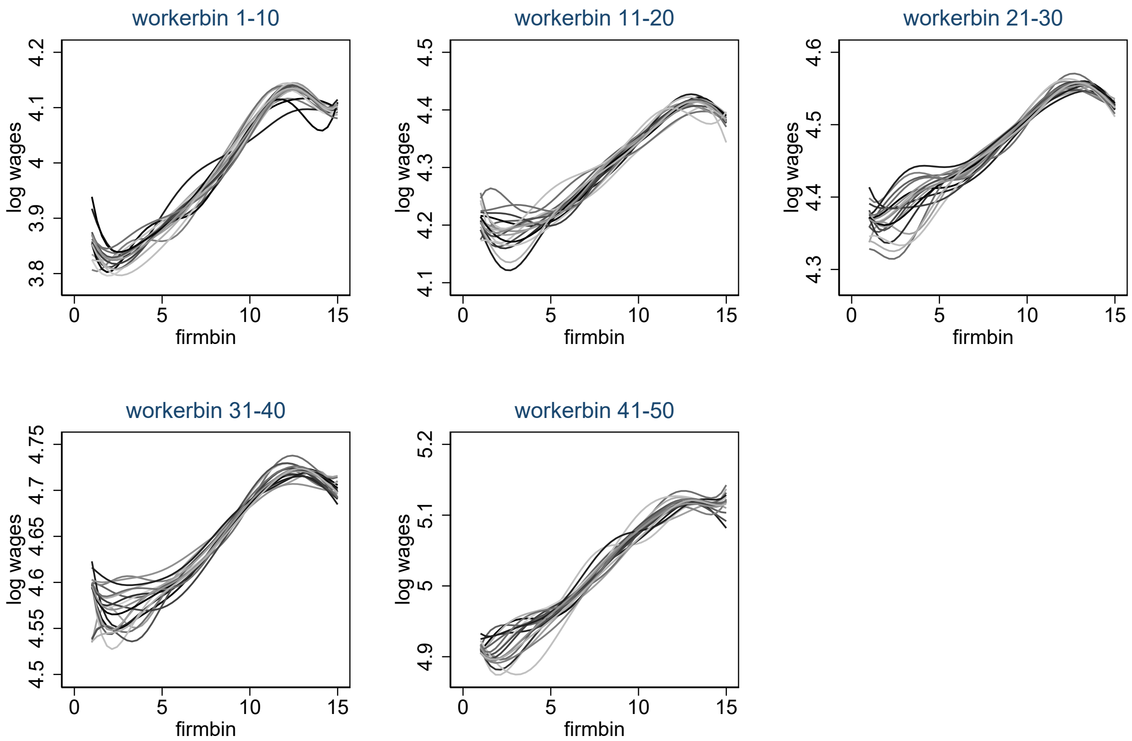
(b) New Matches



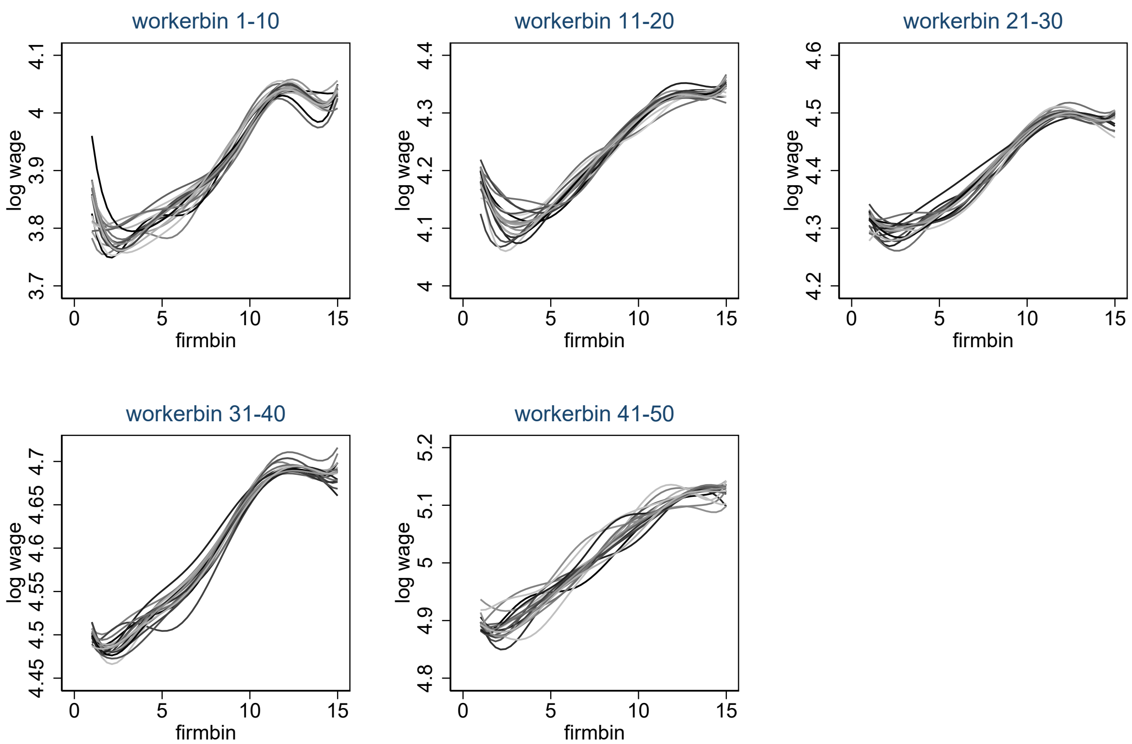
Source: Authors' calculations based on the BHP, EP, BeH.
Note: The plots show the estimated wage–productivity profiles across firm bins for all matches (a) and new matches (b). We use an alternative worker ranking based on the AKM worker effects from a shorter panel (2003–2008). Based on kernel-weighted local polynomial regressions. Kernel: Gaussian. Bandwidth: 2. 95% confidence bands in gray.

Figure D.10: Wage–Productivity Profiles, Firm Subsamples

(a) All Matches



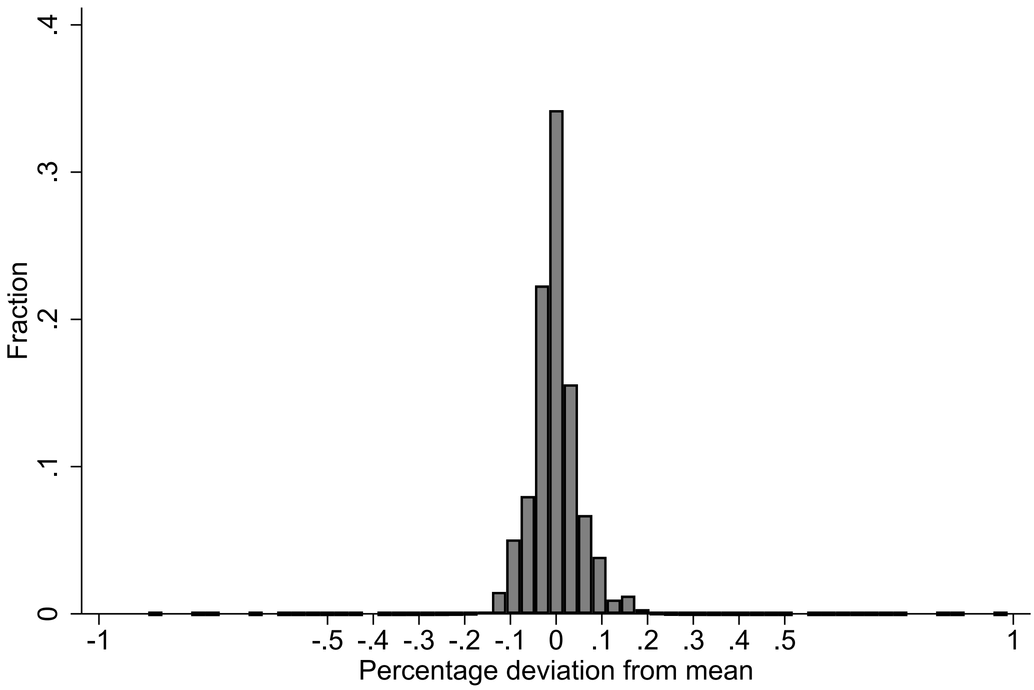
(b) New Matches



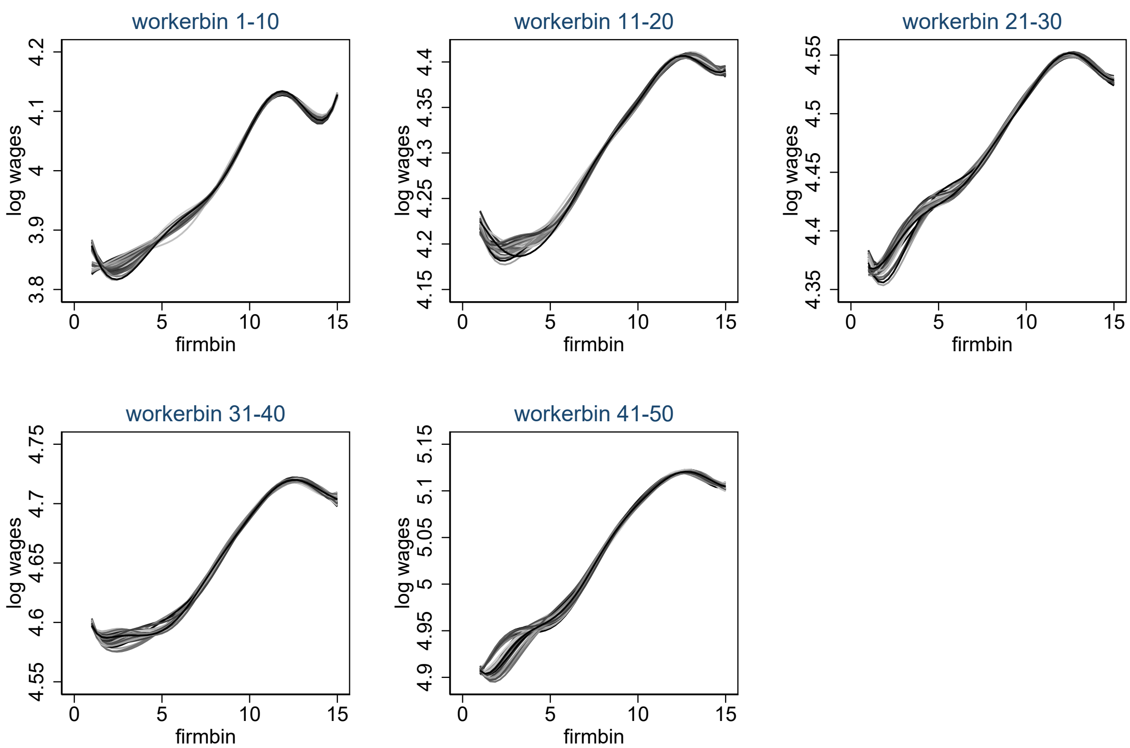
Source: Authors' calculations based on the BHP, EP, BeH.
Note: Plots show the estimated wage–productivity profiles across firm bins constructed from 20 random subsamples (with replacement, clustered by firm-year) from the original sample (size=N). The subsample size M varies randomly between $0.5 * N < M \leq N$. Based on kernel-weighted local polynomial regressions. Kernel: Gaussian. Bandwidth: 2.

Figure D.11: Robustness to Measurement Error

(a) Histogram of Deviations



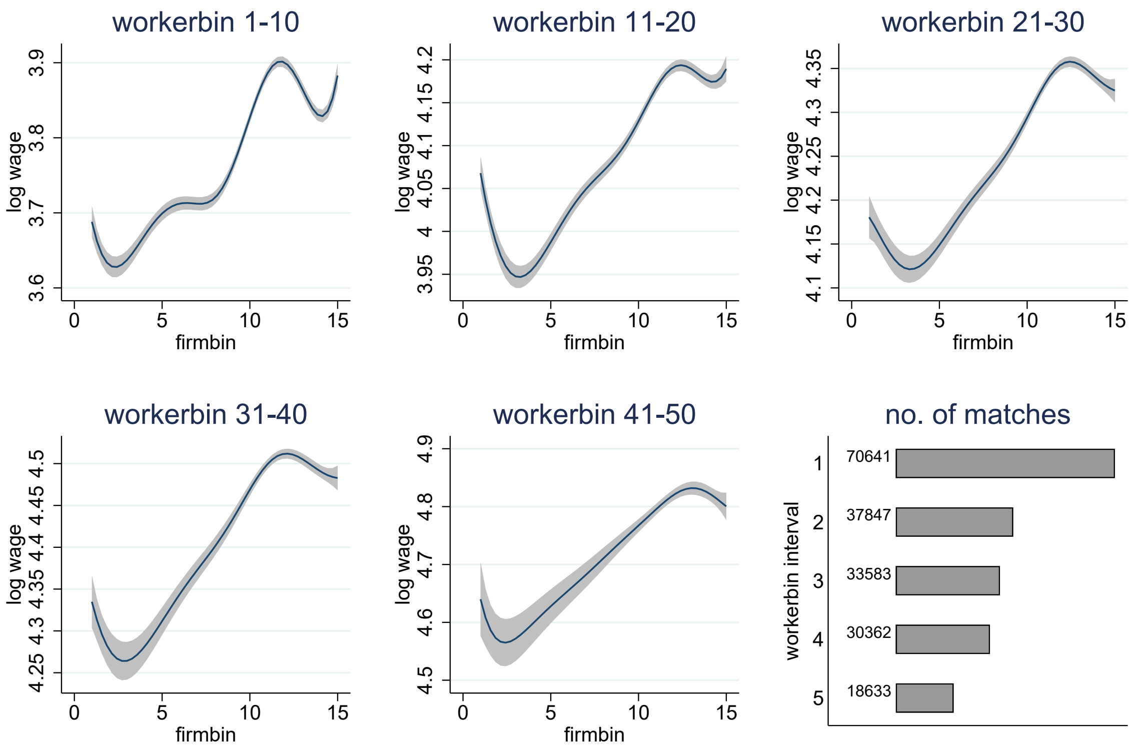
(b) Wage–Productivity Profiles, All Matches



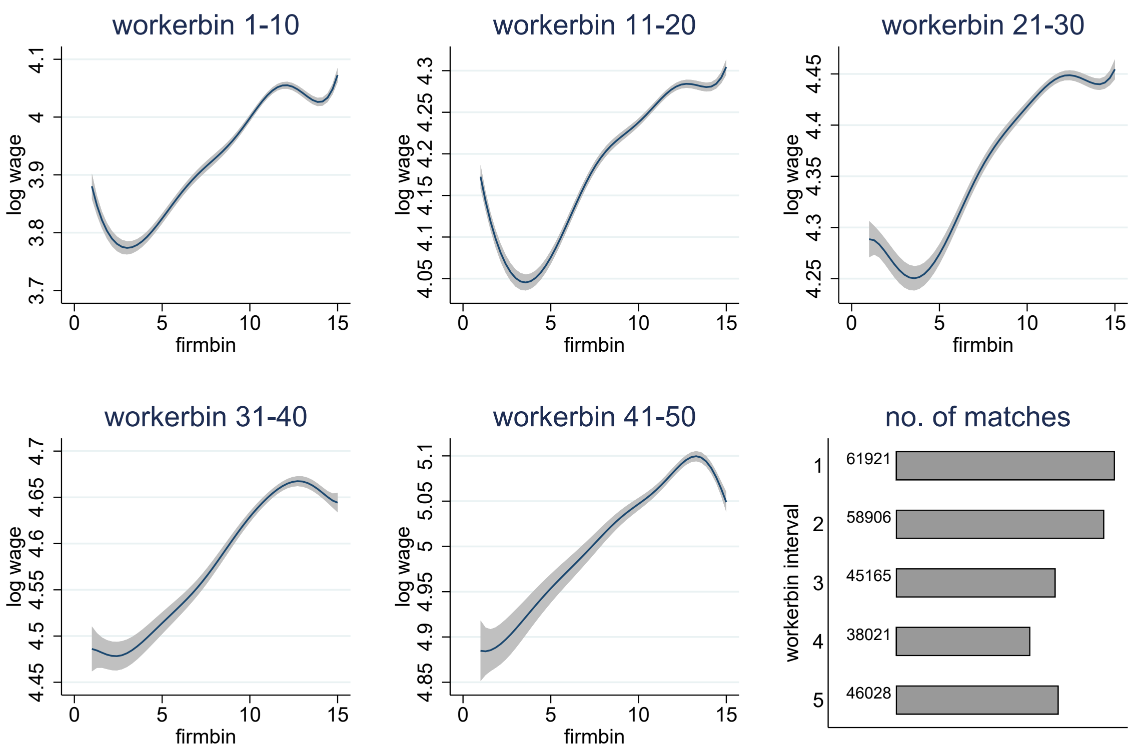
Source: Authors' calculations based on the BHP, EP, BeH.
Note: Panel (a) shows a histogram of deviations from the bootstrapped mean of $\hat{\omega}_{jt}$ measured in logs, so a deviation of 0.1 corresponds to a 10% deviation. Panel (b) shows the estimated wage–productivity profiles for all matches based on firm productivities that are drawn from the confidence band around every $\hat{\omega}_{jt}$ (± 1.96 standard errors, draws based on uniform random numbers). 50 repetitions. Based on kernel-weighted local polynomial regressions. Kernel: Gaussian. Bandwidth: 2.

Figure D.12: Wage–Productivity Profiles, New Matches, First Match-Year Only

(a) Out of Nonemployment



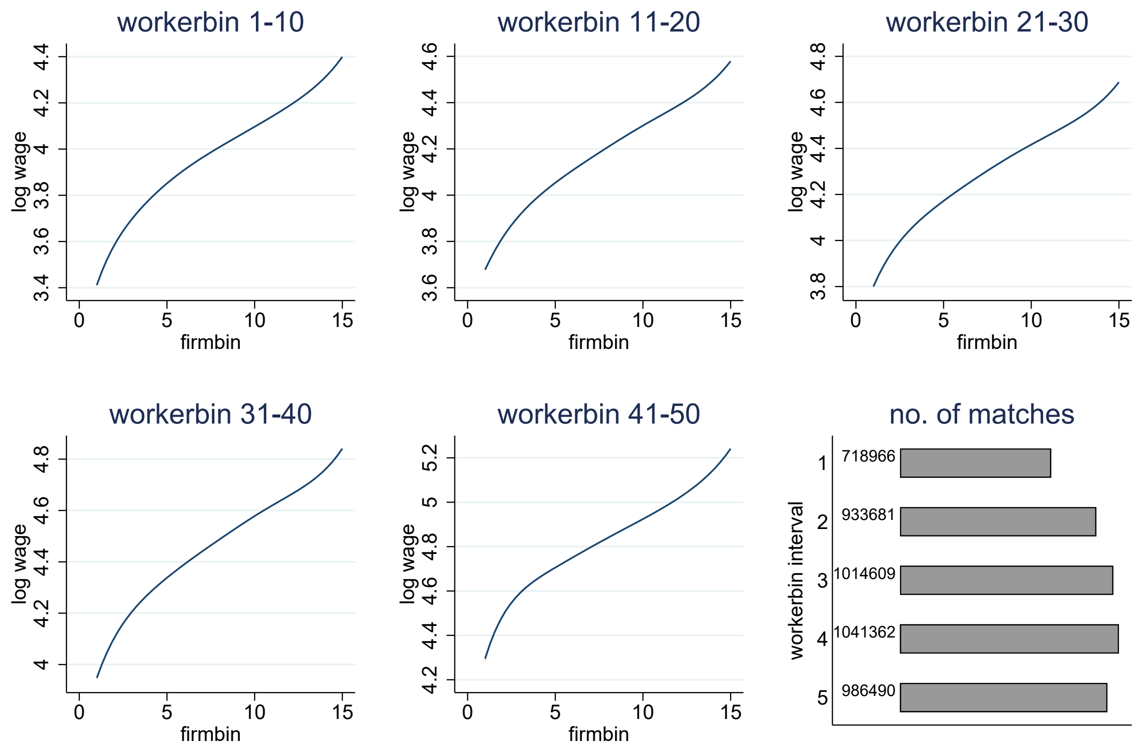
(b) Job-to-Job



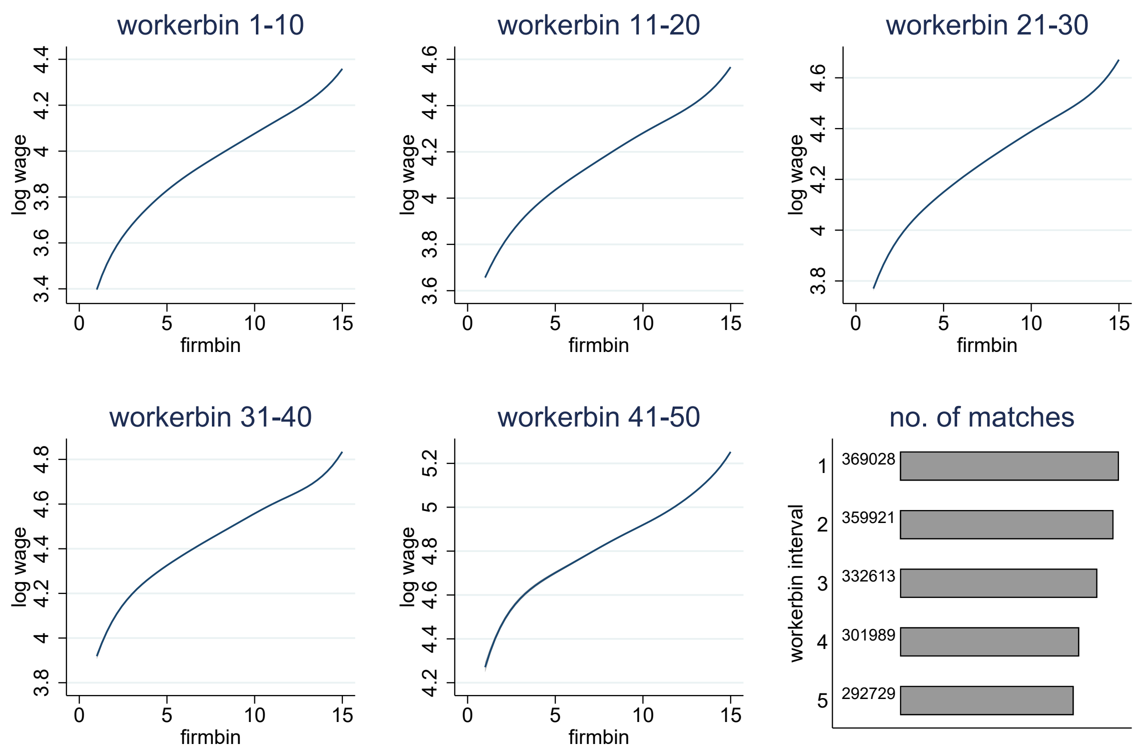
Source: Authors' calculations based on the BHP, EP, BeH.
Note: The plots show the estimated wage–productivity profiles across firm bins constructed using only the first yearly wage observation for all matches and for new matches to remove tenure effects. Based on kernel-weighted local polynomial regressions. Kernel: Gaussian. Bandwidth: 2. 95% confidence bands in gray.

Figure D.13: Wage Profiles with AKM-Based Firm Types

(a) All Matches



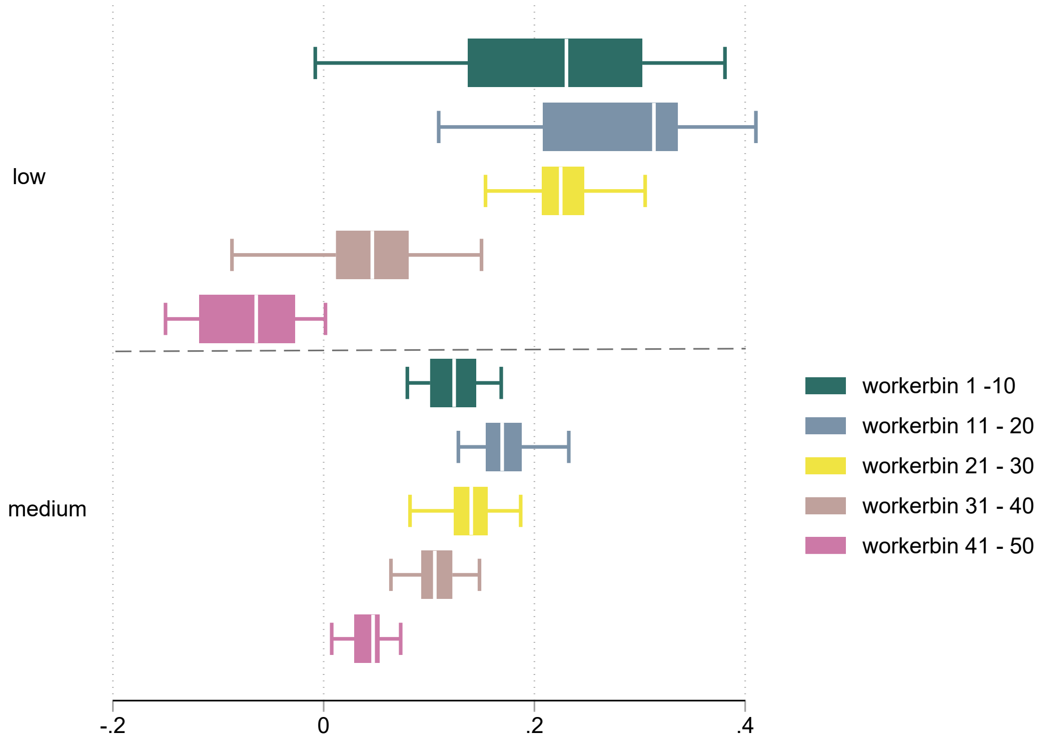
(b) New Matches



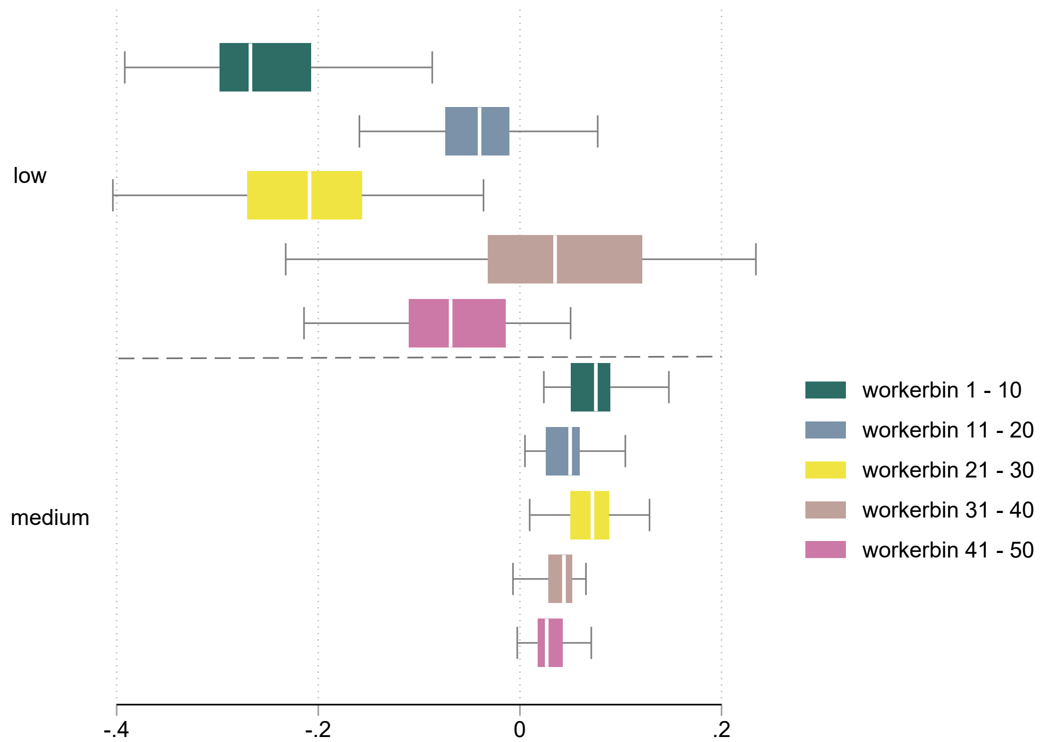
Source: Authors' calculations based on the BHP, EP, BeH.
Note: The plots show the estimated wage profiles across firm bins constructed using the AKM firm effects for all matches and for new matches. Based on kernel-weighted local polynomial regressions. Kernel: Gaussian. Bandwidth: 2. 95% confidence bands in gray.

Figure D.14: Wage Changes for Observed Transitions, Firm Subsamples

(a) To high

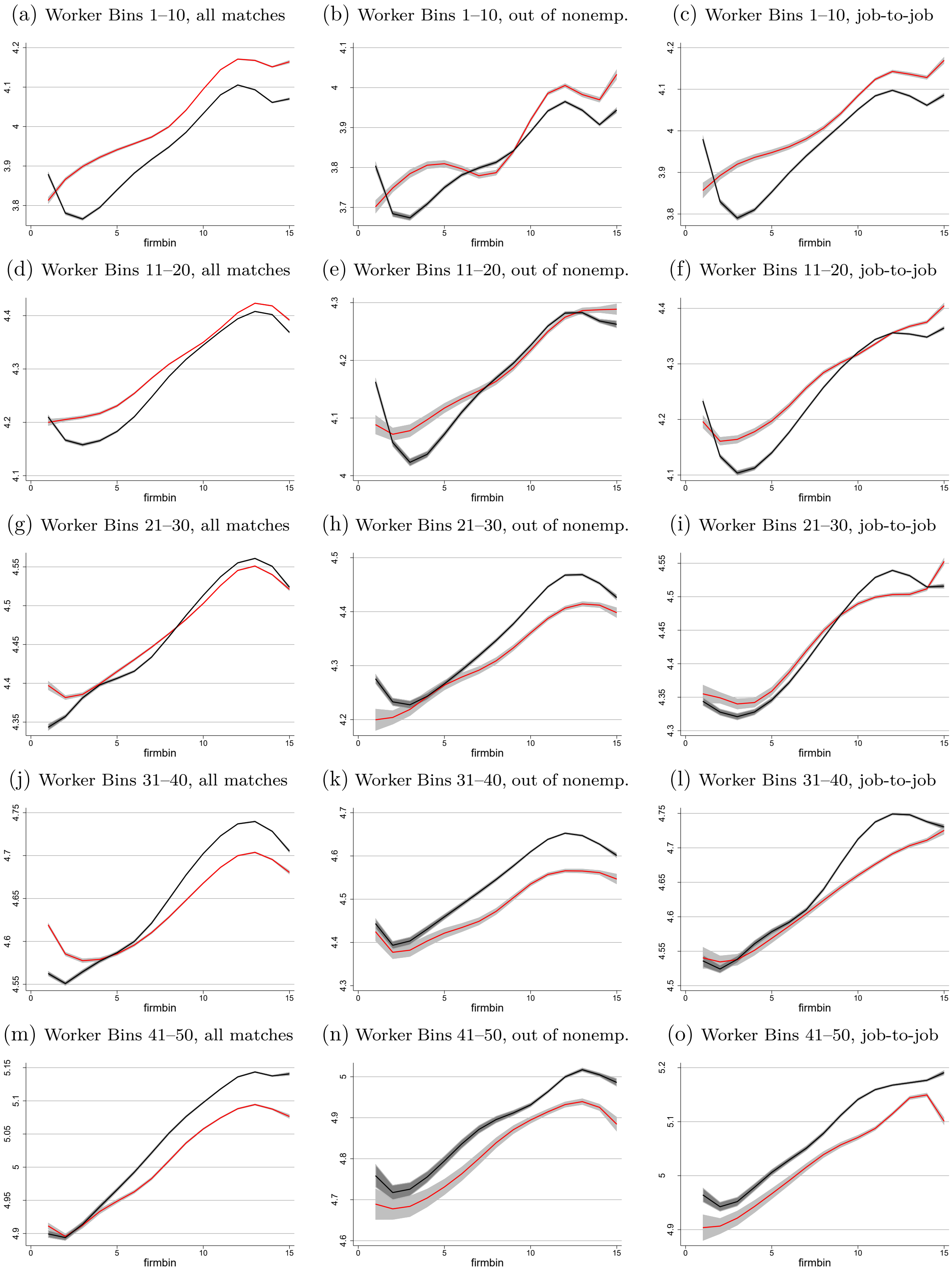


(b) From high



Source: Authors' calculations based on the BHP, EP, BeH.
Note: The plots show the estimated coefficients and 95% confidence intervals (based on robust standard errors) from a linear regression of the individual-level wage differences of transitioning workers on dummies for the origin and destination firm bins using 20 random subsamples (with replacement, clustered by firm-year) from the original sample (size=N). The subsample size M varies randomly between $0.5 * N < M \leq N$. The subsamples consist of new matches (job-to-job switches, no intermediate nonemployment spell) for five groups of worker types. The depicted coefficients are for transitions out of (Panel (a)) and into (Panel (b)) high-productivity firms (bins 13–15). The vertical axes capture the destination/origin firm bin groups: low (bins 1–3) and moderate (bins 4–12).

Figure D.15: Changes in Mean Wages across Worker and Firm Types: 1998–2002 (red) vs. 2003–2008 (black)

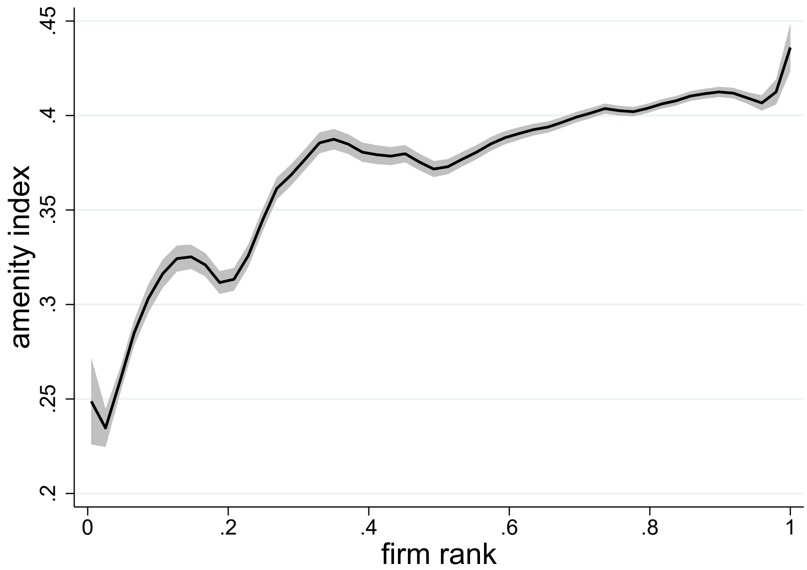


(a) Worker Bins 1–10, all matches
(b) Worker Bins 1–10, out of nonemp.
(c) Worker Bins 1–10, job-to-job
(d) Worker Bins 11–20, all matches
(e) Worker Bins 11–20, out of nonemp.
(f) Worker Bins 11–20, job-to-job
(g) Worker Bins 21–30, all matches
(h) Worker Bins 21–30, out of nonemp.
(i) Worker Bins 21–30, job-to-job
(j) Worker Bins 31–40, all matches
(k) Worker Bins 31–40, out of nonemp.
(l) Worker Bins 31–40, job-to-job
(m) Worker Bins 41–50, all matches
(n) Worker Bins 41–50, out of nonemp.
(o) Worker Bins 41–50, job-to-job

Source: Authors' calculations based on the BHP, EP, BeH.
Note: The plots show the estimated wage profiles across grouped firm bins during two time periods for all matches, new matches out of nonemployment, and job-to-job moves. Based on kernel-weighted local polynomial regressions. Kernel: Gaussian. Bandwidth: 2. 95% confidence bands in gray.

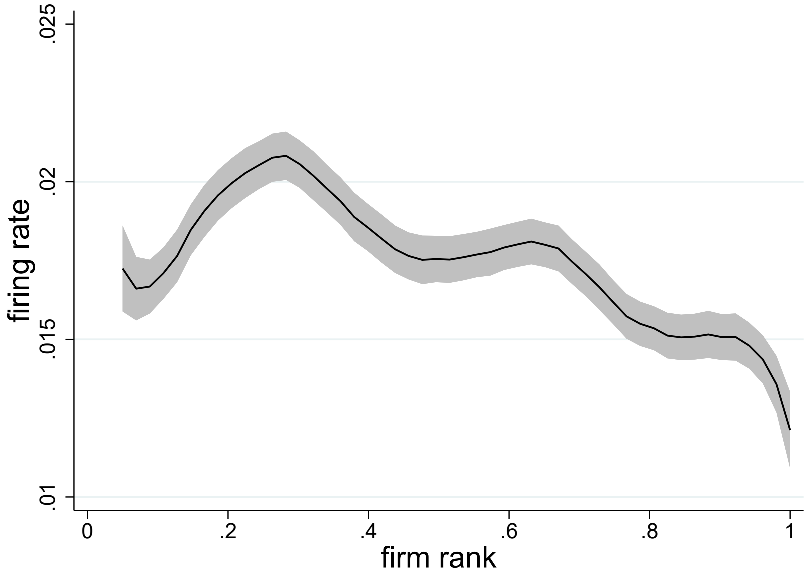
D.7 Extensions

Figure D.16: Amenities



Source: Authors' calculations based on the BHP, EP, BeH.
Note: Estimated univariate kernel density of the amenity index across estimated firm ranks normalized to be between zero and one. Kernel: Epanechnikov. Bandwidth: 0.05. 95% confidence bands in gray.

Figure D.17: Job security



Source: Authors' calculations based on the BHP, EP, BeH.
Note: Estimated univariate kernel densities for the semiannual rate of employer-initiated separations across estimated firm ranks, normalized to be between zero and one. Kernel: Epanechnikov. Bandwidth: 0.1. 95% confidence bands in gray.

Online Appendix References

Andrews, M. J., L. Gill, T. Schank, and R. Upward. 2008. High wage workers and low wage firms: Negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 171 (3): 673–697.

Andrews, M.J., L. Gill, T. Schank, and R. Upward. 2012. High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias. *Economics Letters* 117 (3): 824–827.

Bagger, Jesper and Rasmus Lentz. 2019. An empirical model of wage dispersion with sorting. *Review of Economic Studies* 86 (1): 153–190.

Bond, Stephen and Måns Söderbom. 2005. Adjustment costs and the identification of Cobb Douglas production functions. IFS Working Paper W05/04, Institute for Fiscal Studies, London, UK.

Borovičková, Katarína and Robert Shimer. 2020. High wage workers work for high wage firms. Unpublished manuscript, Department of Economics, New York University.

Bruns, Benjamin. 2019. Changes in workplace heterogeneity and how they widen the gender wage gap. *American Economic Journal: Applied Economics* 11 (2): 74–113.

Cahuc, Pierre, Francois Marque, and Etienne Wasmer. 2008. A theory of wages and labor demand with intra-firm bargaining and matching frictions. *International Economic Review* 49 (3): 943–972.

Card, David, Ana Rute Cardoso, and Patrick Kline. 2016. Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics* 131 (2): 633–686.

Dustmann, Christian, Johannes Ludsteck, and Uta Schönberg. 2009. Revisiting the German wage structure. *The Quarterly Journal of Economics* 124 (2): 843–881.

Fitzenberger, Bernd, Aderonke Osikominu, and Robert Völter. 2006. Imputation rules to improve the education variable in the IAB employment subsample. *Schmollers Jahrbuch: Journal of Applied Social Science Studies* 126 (3): 405–436.

Gandhi, Amit, Salvador Navarro, and David A. Rivers. 2020. On the identification of gross output production functions. *Journal of Political Economy* 128 (8): 2973–3016.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten. 2020. Leave-out estimation of variance components. *Econometrica* 88 (5): 1859–1898.

Lachowska, Marta, Alexandre Mas, Raffaele Saggio, and Stephen A. Woodbury. Forthcoming. Do firm effects drift? Evidence from washington administrative data. *Journal of Econometrics*.

Levinsohn, James and Amil Petrin. 2003. Estimating production functions using inputs to control for unobservables. *Review of Economic Studies* 70 (2): 317–341.

Lochner, Benjamin, Stefan Seth, and Stefanie Wolter. 2020. Decomposing the large firm wage premium in Germany. *Economics Letters* 194, p. 109368.

Olley, G. Steven and Ariel Pakes. 1996. The dynamics of productivity in the telecommunications equipment industry. *Econometrica* 64 (6): 1263–1297.

Petrongolo, Barbara and Christopher A. Pissarides. 2001. Looking into the black box: A survey of the matching function. *Journal of Economic Literature* 39 (2): 390–431.

Pissarides, Christopher A. 2000. *Equilibrium Unemployment Theory.* Boston: MIT Press, 2nd edition.

Postel-Vinay, Fabien and Jean-Marc Robin. 2002. Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica* 70 (6): 2295–2350.

Smith, Eric. 1999. Search, concave production, and optimal firm size. *Review of Economic Dynamics* 2 (2): 456–471.

Stole, Lars A. and Jeffrey Zwiebel. 1996. Intra-firm bargaining under non-binding contracts. *The Review of Economic Studies* 63 (3): 375–410.

Stüber, Heiko and Stefan Seth. 2019. The FDZ sample of the administrative wage and labor market flow panel 1976–2014. FDZ-Datenreport 01/2019, The Research Data Centre (FDZ) at the Institute for Employment Research (IAB), Nürnberg, Germany.

Taber, Christopher and Rune Vejlin. 2020. Estimation of a Roy/search/compensating differential model of the labor market. *Econometrica* 88 (3): 1031–1069.