# Firm Productivity, Wages, and Sorting

Benjamin Lochner[1] and Bastian Schulz[2]

[1]*Institute for Employment Research (IAB), University of Erlangen-Nuremberg*
[2]*Aarhus University, the Dale T. Mortensen Centre, and CESifo*

April 25, 2023

**Abstract**

We study the link between firm productivity and the wages that firms pay. Guided by a search-matching model with large firms, worker and firm heterogeneity, and production complementarities, we infer firm productivity by estimating firm-level production functions. Using German data, we find that the most productive firms do not pay the highest wages. Worker transitions from high- to medium-productivity firms are on average associated with wage gains. Productivity sorting, that is, the sorting of high-ability workers into high-productivity firms, is less pronounced than the sorting into high-wage firms.

This paper is a substantially revised version of CESifo Working Paper No. 6066 (August 2016), which was circulated under the title "Labor Market Sorting in Germany". First draft: February 2016. We thank John Abowd, Michele Battisti, Anja Bauer, Stéphane Bonhomme, Carlos Carillo-Tudela, Jason Faberman, Bernd Fitzenberger, Christian Hellwig, Timo Hener, Christian Holzner, Leo Kaas, Philipp Kircher, Patrick Kline, Thibaut Lamadon, Rasmus Lentz, Jonas Maibom, Christian Merkl, Annaïg Morin, Andreas Mueller, Fabien Postel-Vinay, Helmut Rainer, Jean-Marc Robin, Johannes Schmieder, Heiko Stüber, Coen Teulings, Robert Ulbricht, and Rune Vejlin for their feedback and many helpful suggestions. The input from seminar and conference participants at Aarhus, Cambridge, CBS, the ECB, Glasgow, Nuremberg, LMU Munich, Regensburg, TSE, the ifo Institute, the IAB, SMYE 2016, SaM 2016, ES-NASM 2016, SED 2016, AEA/ASSA 2017, the DFG SPP 1764 Conference 2017, SOLE 2017, and ES-WC 2020 has also greatly improved the paper. We are indebted to Stefan Seth for helping us with data access. All errors are our own. Benjamin Lochner gratefully acknowledges financial support from the Graduate Program of the IAB and the School of Business and Economics of the University of Erlangen-Nuremberg (GradAB), as well as the Emerging Talents Initiative (ETI) of the University of Erlangen-Nuremberg. Bastian Schulz thanks the ifo Institute, the Dale T. Mortensen Centre, and the Aarhus University Research Foundation, particularly the Associate Professor Starting Grant awarded to Rune Vejlin, for financial support.
Corresponding author: Bastian Schulz, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark. Tel.: +4542836363. Email: bastian.schulz@econ.au.dk.

# 1 Introduction

Using the seminal Abowd et al. (1999) (AKM) approach to decompose wages into worker- and firm-specific components, Card et al. (2013) (CHK) and Song et al. (2019) show that increases in wage inequality are driven by increases in the sorting of high-wage workers into high-wage firms. Presumably, the wages that these firms pay are related to their productivity. However, the mapping from productivity to wages is far from clear. Highly productive firms may share profits with their workers and pay high wages. However, the wages these firms pay could also be lower due to labor market imperfections or compensating differentials.

The primary contribution of this paper is to shed light on the link between firm productivity and the wages that firms pay. To this end, we estimate worker- and firm-specific wage components following AKM and firm productivity following Ackerberg et al. (2015) (ACF). We present a parsimonious search-matching model to elucidate the minimal set of assumptions necessary to combine the AKM and ACF approaches to study sorting. The model is compatible with both the discrete-time model of dynamically optimizing firms underlying ACF and the log-linear wage equation underlying AKM. It features matching frictions, decreasing returns, multiworker firms, intrafirm bargaining (following Cahuc et al., 2008), firm heterogeneity and worker-firm complementarities.[1]

Our empirical analysis proceeds in two steps. First, we estimate an AKM model for the full population of private sector employees in Germany and merge the estimated AKM wage components with detailed German establishment survey data.[2] Second, we estimate establishment-level production functions and calculate firm productivity using the ACF approach. We separate firm productivity from the effect of heterogeneous worker ability on output using the estimated AKM effects for all workers at the surveyed establishments.

We present empirical results concerning (i) productivity sorting, i.e., the sorting between worker ability (inferred from the AKM worker effects) and firm productivity (inferred using the ACF approach); (ii) the link between firm productivity and wages; and (iii) the role of productivity sorting in increasing wage inequality. (i) Productivity sorting is less pronounced than wage sorting; the average rank correlation between wage-based worker types and productivity-based firm types is 0.07, while the correlation between AKM worker and firm fixed effects is 0.27. Moreover, the AKM correlation increases over time, while our measure of productivity sorting remains relatively flat. This flat trend is the result of two opposing developments, namely, increasing sorting of medium-ability workers into medium-productivity firms and decreasing sorting of both high-ability workers into high-productivity firms and low-ability workers into low-productivity firms.

[1]Theoretical sorting models highlight production complementarities as the primary reason for positive sorting (Becker, 1973; Shimer and Smith, 2000).

[2]We do not observe firms in the legal sense; rather, we observe establishments, i.e., individual production units. We use the terms "firm" and "establishment" interchangeably throughout the paper.

(ii) The empirical relationship between wages and firm productivity helps to explain these developments: it is hump-shaped. Specifically, the most productive firms pay lower wages than somewhat less productive firms, and this tendency has become more pronounced over time. Consistently, worker transitions from high- to medium-productivity firms are on average associated with wage gains. The most productive firms do not have the highest headcount or wage bill, despite very high labor productivity (value added per worker). Taken together, findings (i) and (ii) suggest that increases in wage sorting could be accompanied by a decrease in the allocative efficiency of the labor market.

The hump-shaped, nonmonotonic relation between wages and productivity poses a challenge for widely used structural models of wage dispersion. To make progress toward replicating this nonmonotonic relation with structural models, we discuss possible extensions of the parsimonious model that we use to develop our empirical strategy. We focus on the link between wages and productivity and comment on the compatibility of extended models with the AKM approach (in the sense of retaining a log-linear wage equation). We consider on-the-job search, general production functions, and multidimensional firm and worker heterogeneity (firms: productivity and nonwage characteristics of jobs (amenities); workers: ability and preferences for amenities).

(iii) To understand the role of productivity sorting in increasing wage inequality, we decompose the wage variance into its within- and between-firm components. The contribution of the between-firm component rose by almost 10% in Germany between 1998 and 2008 (in line with Song et al., 2019). It is comparable in magnitude to the relatively stable within-firm component. However, when we decompose the wage variance using the estimated firm productivity and worker ability types we find that the variance share explained by the between-firm productivity component is low and barely increases over time. Its contribution is dwarfed by the variance shares explained by the within-firm productivity types and between-worker ability types. We conclude that productivity sorting contributes less to rising wage inequality than does wage sorting.

We show how the wage determination mechanism of a search-matching model with multiworker firms, intrafirm wage bargaining, decreasing returns, and worker-firm complementarities can be used to facilitate the estimation of unobserved firm productivity. We build on Cahuc et al. (2008), who embed the Stole and Zwiebel (1996) intrafirm bargaining framework into the canonical search-matching model. With multiworker firms, our approach differs from existing structural work in the sorting literature that mostly assumes one-worker–one-firm matches. Under this assumption, the focus lies on worker *quality*, while the *quantity* dimension (i.e., the number of workers) is not considered. Eeckhout and Kircher (2018) relax this assumption and study both the quantity and the quality dimensions of production from a theoretical perspective. The firm decides which worker type(s) to hire and, additionally, how many workers of each type to employ. Their main result is that firms optimally hire multiple workers of exactly one type. This result holds for both frictionless matching and competitive search. In our data, firms are simultaneously matched with multiple worker types. Thus, we use a model with random search and a production function that is geared toward our empirical approach.

Our findings contribute to the empirical literature on wage dispersion. The aforementioned study of Germany by CHK and that of the U.S. by Song et al. (2019) follow the AKM approach. These studies decompose wage dispersion into the contributions from unobserved worker ability, firm wage premia, and wage sorting, thus measuring the extent to which workers who receive high wages are matched with firms that pay high wages. We show that the way in which one measures firm heterogeneity, i.e., by the wages that firms pay or by firm productivity, affects these decompositions.

The main difference between AKM-inspired and model-based analyses of labor market sorting is typically the implied (non)monotonicity of the wage equation. Due to complementarities, wages are not a monotonic function of firm type in many models (e.g., Gautier and Teulings, 2006; Eeckhout and Kircher, 2011; Hagedorn et al., 2017; Lopes de Melo, 2018). However, the log-linear AKM wage equation implies that wages are strictly increasing in the firm fixed effect. Abowd et al. (2018) find common ground by formally relating the equilibrium quantities derived from the directed-search sorting model of Shimer (2005) to the worker and firm heterogeneity components that can be identified with the AKM approach. We follow their example and discuss how our model's wage equation maps onto AKM and which assumptions are critical for this mapping. Moreover, we contribute to the discussion about wage (non)monotonicity. We find quantitatively important deviations from monotonicity among high- and low-productivity firms.

The literature has developed multiple ways to capture firm heterogeneity when analyzing labor markets, and the productivity-based firm ranking that we propose extends this toolbox. Bagger et al. (2014) estimate firm-level production functions with heterogeneous labor inputs to study wage dispersion. Hagedorn et al. (2017) rank firms based on the value of a vacancy, which can be identified from wage data in the context of their model. Bartolucci et al. (2018) use balance sheet data to rank firms by observed profits. Taber and Vejlin (2020) and Bagger and Lentz (2019) rank firms by the share of workers that they poach from other firms. Sorkin (2018) ranks firms by amenities based on a revealed-preference argument. Haltiwanger et al. (2018) study worker flows using gross output per worker as a measure of firm productivity. Other related papers focus on rent sharing and imperfect competition. Card et al. (2018) show that rent-sharing elasticities are largely driven by worker quality (measured by AKM worker effects). Nevertheless, they find a sizable positive correlation between value added per worker and AKM firm effects, which might be driven by compensating differentials. Lamadon et al. (2022) estimate a model with compensating differentials, worker preference heterogeneity, and monopsony power. In line with our findings, they show that high-productivity firms pay relatively low wages.

# 2 Model

This section introduces a model of multiworker firms with decreasing returns, intrafirm wage bargaining, worker–firm complementarities, and matching frictions. It is designed to elucidate the minimal set of assumptions necessary to combine the AKM and ACF approaches. We show that it is compatible with the model of dynamically optimizing firms underlying ACF and explain how our model's wage equation maps onto the log-linear AKM wage equation. Using this mapping and the estimated AKM wage components, we construct model-consistent firm-level labor inputs for production function estimation. Model details are in appendix 2, available online.

Consider an economy in which atomistic firms produce a numeraire good using multiple heterogeneous labor inputs. Worker heterogeneity is summarized by $n > 1$ ability types indexed by $x$. Worker types are time-invariant. Firm productivity, denoted as $\Omega$, may change due to idiosyncratic firm-level shocks that evolve according to a stochastic process characterized by the conditional CDF $G(\Omega'|\Omega)$. Workers and firms meet randomly. Conditional on meeting, a match is not guaranteed because the match surplus may be too low. Appendix A.1 provides details on the assumed matching mechanism.

To set up the firm's problem, we build on Cahuc et al. (2008), who generalize the canonical search-and-matching model to allow for multiworker firms with heterogeneous labor inputs and decreasing returns, strategic interactions, and intrafirm bargaining in the spirit of Stole and Zwiebel (1996). We add heterogeneous firm productivity and a production structure that is consistent with positive worker–firm sorting. To facilitate a simple and transparent link to the empirical models that we use (i.e., the AKM and ACF models), we abstract away from complementarities between worker types within the same firm. Furthermore, we make six simplifying assumptions. First, workers and firms are risk neutral. Second, both worker and firm heterogeneity are one-dimensional.[3] Third, worker ability and firm productivity are known to all market participants and are cardinally measurable. Fourth, infinitely lived workers supply one infinitesimally small unit of labor (no extensive/intensive margin choice) so that labor input is a continuous variable.[4] Fifth, we abstract away from capital inputs when discussing the model.[5] Sixth, we present the model in discrete time because the ACF approach also relies on a discrete-time model of dynamically optimizing firms. Time indices are omitted.

The general concave, firm-level production function is as follows:

$$Y = F(L, \Omega), \tag{1}$$

[3]We discuss how multidimensional firm/worker heterogeneity would affect our approach in Section 7.

[4]We take part-time labor into account when we estimate the production functions in Section 4.2.

[5]Cahuc et al. (2008) explicitly consider a predetermined capital stock; thus, our model could be extended in this direction. ACF assume a similar capital accumulation mechanism; see Section 4.2.

where $Y$ is value added and $L = \sum_x xL_x$ is a scalar composite labor input measured in units of worker ability. It combines all heterogeneous labor inputs $L_x$, which is the number of type $x$ workers employed by the firm. $\Omega$ is the current productivity of the firm. Our focus is on worker–firm sorting. We assume that output is (log-)supermodular at the match level, i.e., the complementarity between firm productivity $\Omega$ and worker ability determines the contribution of every single match to firm-level output. The marginal product of an additional unit of worker ability is firm specific. We interpret firm productivity as a "nonrival" resource; i.e., we do not consider the span-of-control problem in the optimal allocation of resources to heterogeneous workers (Eeckhout and Kircher, 2018).

The following simple production structure is in line with our assumptions:

$$F(L,\Omega) = \left(\sum_{x=1}^{n}(x \times \Omega)L_x\right)^{\beta_l}, \tag{2}$$

where $0 < \beta_l < 1$ is the output elasticity of the composite labor input. This production function is (weakly) log-supermodular at the match level, which is in line with the sufficient conditions for positive assortative matching (PAM) derived by Shimer and Smith (2000).[6] Under our assumptions, worker ability units are perfect substitutes at the firm level. That is, output depends on $\Omega$ and the number of efficiency units of labor employed. The marginal product of an additional unit of type $x$ labor is as follows:

$$F_x(L,\Omega) = \frac{\partial F(L,\Omega)}{\partial L_x} = x\beta_l\Omega^{\beta_l}L^{\beta_l-1}. \tag{3}$$

Worker ability $x$ scales the marginal product at an $(L,\Omega)$ firm. We use this property below. The marginal product is increasing in firm productivity and the output elasticity of labor but decreasing in the composite labor input $L$ due to decreasing returns.

As in Cahuc et al. (2008) and ACF, employment is a state variable. The firm's problem is to optimally choose how many vacancies to post given its expected profits from hiring. We assume that vacancies cannot be targeted to specific worker types and are subject to a flow cost $c$, which could depend on the firm type. In appendix A.2, we solve the firm's problem and derive the relevant optimality conditions in the steady state.

The bargained wage satisfies the Nash sharing rule. The firm's surplus consists of its marginal profit from hiring an additional worker of type $x$ (equation A.11). Its threat point is to renegotiate wages with all other employees (Stole and Zwiebel, 1996). The worker's surplus is the difference between the option values of employment and unemployment (equations A.14 and A.15). Wage (re)negotiations happen instantaneously so

[6]Shimer and Smith (2000) establish the existence of an equilibrium in this environment. PAM arises with a log-supermodular match-level production function, while log-submodularity leads to negative assortative matching (NAM). In this empirical paper, we do not attempt to generalize the Eeckhout and Kircher (2018) large-firm sorting conditions for random search models.

that firm-level employment remains fixed. In appendix A.3, we show that the outcome of intrafirm wage bargaining in the model is described by the following differential equation:

$$w(x, L, \Omega) = \alpha \left( F_x(L, \Omega) - L \frac{\partial w(x, L, \Omega)}{\partial L} \right) + (1 - \alpha)(1 - \beta)U(x). \tag{4}$$

This is a discrete-time version of the wage bargaining outcome derived by Cahuc et al. (2008) for the "single labor case". We can utilize their result due to our assumption of perfectly substitutable ability units. $\alpha$ is the workers' bargaining power parameter, and $\beta$ is the common discount factor. The first term in parentheses shows that the wage of a type $x$ worker at an $(L, \Omega)$ firm is a function of the worker–firm-specific marginal product, $F_x(L, \Omega)$. The second term captures the inframarginal effect that hiring the marginal worker has on all other workers' wages. This mirrors the finding that firms can reduce incumbent workers' wages by increasing employment in the presence of decreasing returns.[7] Absent complementarities between worker types, the inframarginal adjustment reflects decreasing returns only and is unambiguously negative. Finally, the third term in equation (4) captures the worker-specific outside option, which we return to below. The solution[8] to the differential equation (4) is as follows:

$$w(x, L, \Omega) = (1 - \alpha)(1 - \beta)U(x) + \int_0^1 z^{\frac{1-\alpha}{\alpha}} F_x(Lz, \Omega) \mathrm{d}z, \tag{5}$$

which indicates that the wage depends on the worker's outside option and an integral expression that combines the worker's marginal product and the inframarginal effect. The latter is weighted by the worker's bargaining power and decreases in the distance from the margin. With the assumed production function, equation (5) also implies that wages are monotone in firm productivity, which we show formally in appendix A.4.

## 2.1 From Theory to Estimation

To estimate $\Omega$, we use a two-step approach. First, we estimate the AKM model with German data (Section 3, appendix D.1). The estimation yields, for all individual workers and firms in the largest connected set, worker fixed effects, firm fixed effects, and a residual. Second, we estimate the production function (Section 4.2, appendix C). We merge rich survey data on inputs and output for a representative sample of firms with the estimated AKM wage components for their full workforce to predict the firm-level labor input (wage bill). According to the model's separability property, the worker fixed effects capture the full effect of worker ability on wages. Thus, this wage component allows

[7] Cahuc et al. (2008) allows for unrestricted substitutability/complementarity patterns between worker types. In this case, the effects on coworker wages can be either positive or negative. Firms may strategically over/underemploy specific worker types depending on their contribution to the total wage bill.

[8] We follow Stole and Zwiebel (1996) and Cahuc et al. (2008); see online appendix A.3.

us to separate firm productivity from the effect of worker ability on output. To predict labor inputs, we include the effect of observable worker characteristics (not part of the model). Moreover, we include either the firm fixed effects or the AKM residuals, or both. This allows us to investigate how firm-specific and time-(in)variant wage components matter for the inferred productivity measure.

The model's wage equation maps onto the AKM model as follows. Both the outside option and the integral expression in (5) are functions of worker ability. Their effects on wages are absorbed by the AKM worker fixed effects when both wage components are linear with respect to worker ability $x$. Given our production structure, the integral expression is indeed linear in $x$ because worker ability scales the marginal product of labor. This feature is crucial for the applicability of the AKM approach.[9] Plugging the marginal product of labor into the wage equation (5) yields:

$$w(x, L, \Omega) = (1 - \alpha)(1 - \beta)U(x) + x \int_0^1 z^{\frac{1-\alpha}{\alpha}} \beta_l \frac{F(Lz, \Omega)}{Lz} \mathrm{d}z, \tag{6}$$

where $x$ can be written in front of the integral sign. We show in appendix A.5 that the worker's outside option in the first term is also linear in $x$ under two additional assumptions. First, the flow value of unemployment for the worker, $b(x)$, must be proportional to $x$. This is a standard assumption. Second, matching sets must cover the whole type space; that is, there are no unacceptable combinations of worker and firm types. An implication of this assumption combined with our matching technology is that all worker types are unemployed for the same (expected) duration. In Section 4.1, we verify that this holds in our data (conditional on observables). Another implication is that there is no endogenous sorting despite the complementarity in the production function.

Next, consider firm fixed effects. How the integral expression in (6) is captured in the AKM model depends on the underlying data-generating process (DGP). Suppose that our wage equation describes reality without error. The economy is in a steady state, and firm productivity does not change. In this case, (a log-affine transformation of) the firm-specific integral, which represents the effects of firm productivity and decreasing returns on the wage, is absorbed by the time-invariant firm fixed effect, and the residuals are zero. With a more complex DGP, the value of the integral changes over time. Suppose productivity $\Omega$ follows an AR(1) process. In this case, the AKM firm fixed effect captures a time average of the effect of changing productivity on the wage. Thus, the more persistent the productivity process, the more informative is the time-invariant firm effect. The AKM wage component that remains is the residual, which absorbs fluctuations due to transitory shocks around the time average of the effect of productivity on the wage.[10]

[9] Linearity in $x$ implies that the wage equation is additive in logs (see appendix A.5).

[10] The mapping that we describe could be obscured when using real data, e.g., due to measurement/sampling error in short panels. We assess the robustness of our results in Section 6.

# 3 Data

Our analysis combines two data sets provided by the Institute for Employment Research (Institut für Arbeitsmarkt- und Berufsforschung, IAB) of the German Federal Employment Agency (Bundesagentur für Arbeit, BA). The first is the "IAB Employee History File" (BeH), which comprises the universe of employment spells recorded by the German social security system. The second data set is the "IAB Establishment Panel" (EP), which is a representative establishment-level survey that can be linked to the BeH.[11] The BeH data contain information on worker gender, age, and education (5 categories); start and end dates for the employment spells; total earnings; and occupation/industry codes. They cover the vast majority of the German workforce. Only civil servants and self-employed workers who do not make social security contributions are excluded. Each worker and each establishment have a unique identification number, which allows us to follow workers over time and across establishments. Regarding sample selection, we follow CHK and Lochner et al. (2020). We start with the universe of employment spells observed between 1998 and 2008. There is no exact information on hours worked, so we restrict our sample to full-time employees (males and females) aged 20 to 60. We define the job with the highest total wage as each worker's main job.

The EP is a comprehensive annual survey of establishments.[12] It provides us with the data needed to estimate production functions; we observe revenue, intermediate goods purchases (reported as revenue shares), value added (calculated as revenue minus intermediate goods purchases), and net investments in four categories of capital goods (buildings, production machinery, IT, and transport equipment). We restrict our analysis to establishments with nonmissing data for revenue and intermediate goods and trim the data by dropping establishments below the 1st and above the 99th percentile of the revenue distribution. We supplement the EP data with covariates from the "Establishment History Panel" (BHP; see Spengler, 2008). These include average wages, headcounts, shares of full-time/part-time workers, and worker shares by education group. Moreover, the BHP provides administrative information on firm age and a consistent industry classification.

We observe nominal gross daily wages, which we deflate using the consumer price index. Earnings are tracked up to a threshold, so we follow Dustmann et al. (2009) and impute the upper tail of the wage distribution by running a series of tobit regressions; see appendix B.1. We impute missing and inconsistent education data using the methodology proposed in Fitzenberger et al. (2006); see appendix B.2. To estimate the capital stock, we use a perpetual inventory method (Müller, 2008). This method approximates the establishment-level capital stock by combining net investments and depreciation rates (available from the national accounts) for different types of capital goods.

[11] The EP is a random sample of all establishments in Germany, stratified by size, industry, and federal state. See Kölling (2000) and Fischer et al. (2009) for a detailed description of the EP data.

[12] These are single production units. A total of 80% report that they are single-establishment firms.

Following CHK, we estimate an AKM-type wage regression on the largest connected set in the BeH for our period of analysis, namely, 1998–2008. We include both men and women. The connected set contains more than 233 million person-years, corresponding to 35 million individual workers at 3.3 million establishments; 55% of the observations are classified as movers between establishments, and the regression model is identified for such workers.[13] We discuss the estimation results in online appendix D.1. We use the estimated AKM wage effects for men, women, and both movers and stayers from the full BeH sample (column (a) in Table D.1) to construct the labor input for the production function estimation at the firm level. Based on the discussion in Section 2.1, we include (exponentiated) worker fixed effects, the effects of observable characteristics, and, in some specifications, firm fixed effects and AKM residuals to distinguish among the different channels through which wages and firm productivity are related according to our theory.

We construct two final samples for analysis. The first sample includes *all matches*, i.e., matches that formed both before and after 1998. We do not distinguish between job-to-job (J2J) flows and matches formed out of nonemployment (OON). There are 4,695,108 employment spells for 1,344,382 workers employed at 10,004 EP establishments. The second sample includes *new matches*, i.e., matches formed from 1999 onward. We distinguish between J2J flows and matches that formed OON and use the initial observation (1998) to determine employment status before the new match. There are 1,656,280 employment spells for 633,831 workers at 9,659 establishments. A total of 64% (1,055,151) of new employment spells are J2J moves from one employer to another, while 36% (601,129) are spells formed OON, which also include young workers who enter the labor market.[14]

# 4 Ranking Workers and Firms

## 4.1 Worker Ranking

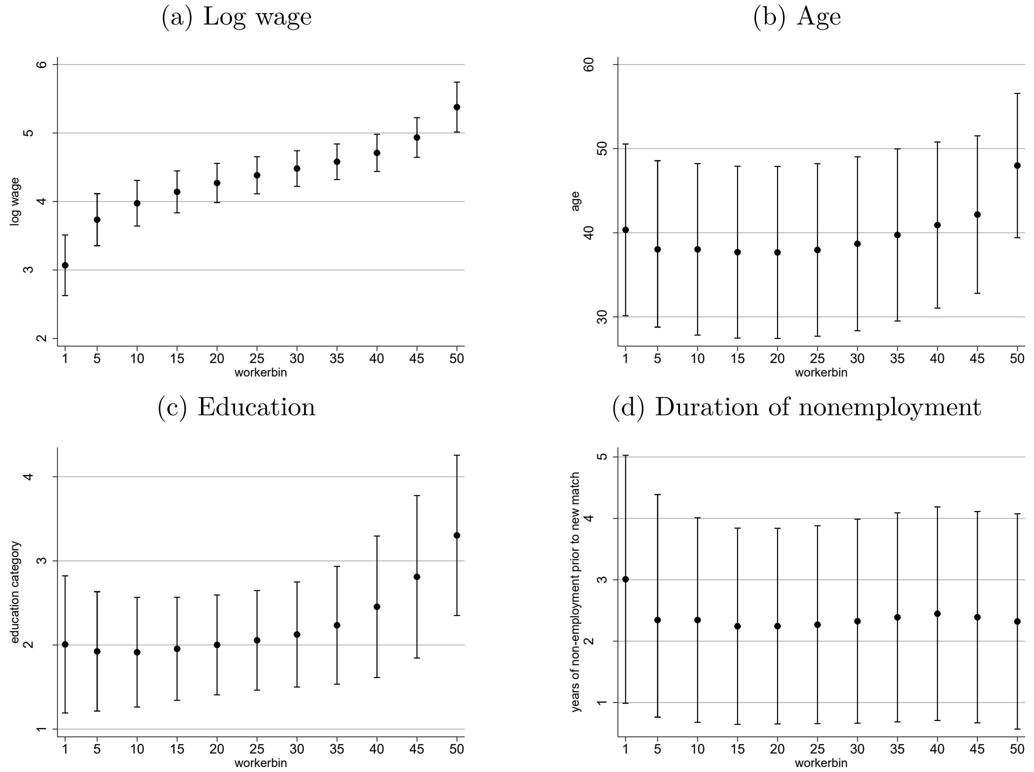
According to our model, the estimated AKM worker fixed effects fully capture the effect of unobserved worker ability on wages. This result can serve as a "microfoundation" for using the AKM effects to rank workers. We create 50 ability bins of equal size.[15] To show how the bins summarize worker heterogeneity, we decompose the variances of observed wages, age, and education into the shares explained within and between bins. Little within-bin variance implies homogeneity among workers along the corresponding dimension. The rankings are based on the worker fixed effects, which explain the majority of the wage variation in the data. Thus, the between-bin share of the wage variance is

[13]Following CHK, we impute worker fixed effects and residuals for "stayers" (workers who do not change their employer over the entire sample period).

[14]Our definition of nonemployment includes marginal employment, unemployment (benefit receipt), and inactivity.

[15]There are 702,540 individual workers in each of the 50 bins.

Figure 1: Log Wage, Age, Education, and Nonemployment Duration across Worker Bins




Source: Authors' calculations based on the BeH.
Note: The figures plot the means ± one standard deviation for log wages, age, education, and nonemployment duration for every fifth worker bin. The ages of the individual workers in our sample range from 20 to 60. There are 5 education categories, which are defined in online appendix B.2.

large (70%). For age and education, the picture is quite different. We find that 96% of the age variation and 74% of the education variation are within bins. Because we control for time-varying age and education effects in the AKM regression, this finding is a direct reflection of the low correlation between the estimated worker effects and the time-varying observables (see Table D.1).

Figure 1 illustrates how log wages, age, and education vary across worker bins. Panel (a) shows that mean wages increase monotonically across worker bins, as expected. In contrast, Panel (b) shows that the mean age across worker bins is relatively flat. Only the highest bin has a somewhat higher mean age. However, due to the large standard deviation, this difference is not significant. Panel (c) shows that mean education increases above bin 35 but is flat below it. Note that the dispersion of education increases among the highest-ranked workers. Thus it is more common to observe high-ranked workers with little education than low-ranked workers with tertiary degrees. Finally, Panel (d) shows the duration of nonemployment across worker types. The mean is virtually flat. Recall that we control for the effect of observable characteristics. The finding that, conditional on observables, nonemployment durations do not change with worker type supports the assumption that the matching sets cover the whole type space (see Section 2).

## 4.2 Firm Ranking

We rank firms based on their unobserved productivity, which we infer by estimating production functions at the establishment level. The empirical industrial organization literature shows that this approach is susceptible to two challenges. The first is that input choices are likely correlated with firm productivity. To address this challenge, we rely on the ACF version of the control function approach (Olley and Pakes, 1996; Levinsohn and Petrin, 2003). The key assumption is that intermediate input demand (the control function) is strictly increasing in (scalar and unobserved) firm productivity. This implies that the control function can be inverted, thus effectively controlling for unobserved firm productivity by substituting it out of the production function.

The second challenge is that the quality of the labor inputs varies across firms. Physical units, e.g., headcounts or hours worked, do not reflect differences in worker ability and imply measurement error. This complicates the separate identification of firm productivity and the output elasticity of labor. The literature addresses this challenge by controlling for labor quality differences (e.g., Fox and Smeets, 2011; Irarrazabal et al., 2013). We construct model-based labor input measures that isolate the effect of worker ability on output (see Section 2.1).

To estimate the production function, we rewrite equation (2) using the composite labor input, $L_{jt}$. We add capital, $K_{jt}$, and indices for firms ($j$) and time ($t$):

$$Y_{jt} = \left(\Omega_{jt} L_{jt}\right)^{\beta_l} \left(\Omega_{jt} K_{jt}\right)^{\beta_k}. \tag{7}$$

$\beta_l$ and $\beta_k$ are the output elasticities of labor and capital. Both inputs are scaled by the firm's current productivity $\Omega_{jt}$. Note that $Y_{jt}$ is value added (revenue minus expenditures on intermediate goods). Without assuming constant returns to scale, we obtain

$$Y_{jt} = \Omega_{jt}^{\beta_l + \beta_k} L_{jt}^{\beta_l} K_{jt}^{\beta_k}. \tag{8}$$

The sum of the output elasticities in the exponent on $\Omega_{jt}$ is irrelevant for the purpose of ranking the firms. Thus, we define $\omega_{jt} = (\beta_l + \beta_k) \ln \Omega_{jt}$ when taking logs. We estimate the following value-added production function, where lowercase letters indicate logarithms:

$$y_{jt} = \beta_0 + \beta_l l_{jt} + \beta_k k_{jt} + \omega_{jt} + z'_{jt}\gamma + \epsilon_{jt}. \tag{9}$$

We add a constant $\beta_0$, the residual $\epsilon_{jt}$, a vector of controls $z'_{jt}$ (indicators for West German establishments and employee representation in management, four firm age categories, and the share of part-time workers), and time and sector fixed effects.

ACF assume that the capital stock is predetermined by investment $i_{j,t-1}$:

$$k_{jt} = \kappa(k_{j,t-1}, i_{j,t-1}). \tag{10}$$

We abstract away from capital in Section 2, but our model builds on that of Cahuc et al. (2008), who explicitly consider a capital accumulation process similar to equation (10), so the models are compatible in the capital dimension. The demand for other inputs, labor and intermediate goods, may change in response to firm productivity shocks.[16] The conditional expectation of firm productivity is modeled using a first-order Markov process:

$$\omega_{jt} = E(\omega_{jt}|\omega_{j,t-1}) + \xi_{jt} = \rho\omega_{j,t-1} + \xi_{jt}, \tag{11}$$

where $\xi_{jt}$ is an innovation that is assumed to be uncorrelated with $\omega_{jt}$ or $k_{jt}$. For the estimation, we follow the ACF model and assume that $\omega_{jt}$ follows an AR(1) process with parameter $\rho$. In our model, both the firm's vacancy-posting decision and its matching set depend on current productivity; see appendix A.2. Moreover, the firm's dynamic problem involves the expected future evolution of productivity. Recall that we abstract away from productivity shocks when deriving the wage equation (4) and establishing its linearity in worker type $x$. This simplifies the derivation but is not critical for the result. As long as the evolution of firm productivity is independent of the employed worker types, the wage equation will still be (log-)linear in $x$ under our assumed production structure.[17]

Finally, the ACF approach allows for (exogenous) across-firm differences in labor adjustment costs and wage setting. This flexibility makes it a good fit for search-matching models. ACF use the intermediate input demand conditional on labor as the control function. This conditioning eliminates the link between intermediate input demand and the labor input, which, in turn, depends on wage conditions and adjustment costs.[18] In the context of our model, differences in adjustment costs and wage setting across firms arise for two reasons. First, vacancy posting costs can be heterogeneous, and firms have limited control over the hiring process.[19] Second, due to decreasing returns to labor and intrafirm bargaining, wages have a firm-size component that varies across firms.[20]

We describe the technical details of the ACF production function estimation in appendix C. Table 1 presents the results. It shows five specifications in which we vary the labor input, i.e., the way we control for worker ability. All specifications include the aforementioned firm-level controls, as well as year and sector fixed effects. Specifications (a)–(d) use predicted wage bills based on different combinations of the estimated AKM

[16] The firm's information set when making input choices includes all past and present productivity shocks $\{\omega_{j\tau}\}_{\tau=0}^{t}$, but it does not include future productivity shocks.

[17] Consider an environment with productivity shocks and, accordingly, changing labor demand and endogenous separations. In this case, the wage equation contains an additional forward-looking component that captures the expected evolution in firm productivity. Worker type $x$ can still be written in front of the integral sign under the mild assumption that $x$ and the firm productivity process are independent.

[18] Earlier approaches that use unconditional intermediate input demand are more restrictive. As explained in ACF (p. 2431), their approach can handle serially correlated, exogenous, and unobserved shocks to the prices of capital and labor. Levinsohn and Petrin (2003) can only allow for such shocks to the price of capital inputs. Neither approach allows for shocks to the price of intermediates.

[19] With random search, the time that passes until a suitable worker arrives is stochastic.

[20] Our model implies a nondegenerate firm-size distribution due to heterogeneous productivity.

Table 1: Production Function Estimation Results

| | *Value Added* | | | | |
|---|---|---|---|---|---|
| | AKM Predicted Wage Bill | | | | Headcount |
| | (a) | (b) | (c) | (d) | (e) |
| *Labor input* | 0.7129 | 0.7131 | 0.7125 | 0.7099 | 0.8573 |
| | (0.0141) | (0.0141) | (0.0141) | (0.0135) | (0.0133) |
| *Capital input* | 0.1706 | 0.1703 | 0.1707 | 0.1512 | 0.1358 |
| | (0.0117) | (0.0117) | (0.0117) | (0.0116) | (0.0110) |
| *Variance of* $\hat{\omega}_{jt}$ | 0.074 | 0.073 | 0.072 | 0.064 | 0.050 |
| *Variance of* $\epsilon_{jt}$ | 0.628 | 0.627 | 0.626 | 0.597 | 0.583 |
| *Worker fixed effects* | Yes | Yes | Yes | Yes | – |
| *Worker observables* | Yes | Yes | Yes | Yes | – |
| *Firm fixed effects* | No | Yes | No | Yes | – |
| *Residual* | Yes | No | No | Yes | – |
| Observations | 38,598 | 38,598 | 38,598 | 38,598 | 39,808 |

Source: Authors' calculations based on the BHP, EP, BeH.
Note: Bootstrapped standard errors (500 iterations) in parentheses. Coefficients and standard errors are rounded to four decimal places. All specifications include firm-level controls as well as year and sector fixed effects (32 categories).

wage components. We always include worker fixed effects, which, according to our model, capture the effect of worker ability on output, along with the effects of worker observables. In addition, we include the AKM residual in (a), the firm fixed effect in (b), neither the residual nor the firm fixed effect in (c), and both in (d). The labor input in (d) is thus equal to the total observed wage bill for the firm. For the sake of comparison, we also show the results of a specification that uses the worker headcount (e).

The first interesting finding is the large difference in the estimation results for (a)–(d) and those for (e). The headcount mismeasures the labor input in the presence of worker heterogeneity; thus, the output elasticity of labor is biased upward in specification (e), while the capital elasticity and the estimated productivity $\hat{\omega}_{jt}$ are biased downward. Across specifications (a)–(d), the results are very consistent, with an estimated output elasticity of labor of approximately 0.71. Together with the capital elasticity, these estimates imply decreasing returns to scale, in line with our model's assumption.

The second interesting finding is that the AKM wage components beyond the worker fixed effects and observable effects have little effect on the estimation results. In light of our model, this is unsurprising. The worker fixed effects capture the effect of worker ability on output. The firm effect, however, does not reflect contributions to output. It captures the time average of the effect of firm productivity on wages (see Section 2.1). The AKM residual absorbs the effect of firm productivity shocks on wages according to our model. Such shocks have been shown to affect wages ("pass-through").[21] A correlation between the contemporaneous productivity realization and the labor input does not threaten our estimation strategy.[22] Notably, adding the residual has little impact on the estimated output elasticities. However, capturing the reaction of wages to productivity shocks could still be important. If worker ability and firm productivity are complements, the contribution of worker ability to firm output could shrink after a negative shock. For this reason, we select specification (a), which includes the AKM residuals, as our benchmark.

Finally, we examine the dispersion and persistence of estimated firm productivity $\hat{\omega}_{jt}$. The estimated variance of 0.074 in our benchmark specification is almost 50% higher than that in the headcount specification (e). It is, however, an order of magnitude smaller than the residual variance. The residual absorbs transitory shocks, whereas productivity $\omega_{jt}$ is the realization of a persistent stochastic process, and the estimated autocorrelation in $\hat{\omega}_{jt}$ in the benchmark specification is indeed high (0.75). Given the link between firm productivity and the AKM firm effects suggested by our model, this high degree of persistence implies that firm fixed effects should be quite stable over time.[23] Thus, we rank firms by the firm-level mean of productivity.

Productivity ranks are moderately positively correlated with headcount (0.40), value added (0.43), value added per worker (0.25), the capital stock (0.24), and profit (0.16). They are not correlated with capital per worker (-0.03). These correlations suggest that the largest firms in terms of output, workforce, and assets are not the most productive firms. The correlation with the share of workers with tertiary degrees is rather low (0.08), so productive firms do not necessarily have a highly educated workforce. In appendix D.2, we compare the productivity ranking with other rankings used in the literature.

Figure 2 shows estimated kernel densities for six firm performance measures across ranks. Revenue (measured by log sales, Panel (a)) mostly increases in rank but becomes relatively flat at the top. The most productive firms do not have greater revenue than firms around the 80th percentile of the productivity distribution. In Panel (b), the relation between firm rank and log employment (measured by the headcount) is shown. This size measure increases until approximately the 80th percentile and falls steeply thereafter. The most productive firms are comparable to median-productivity firms in terms of size.

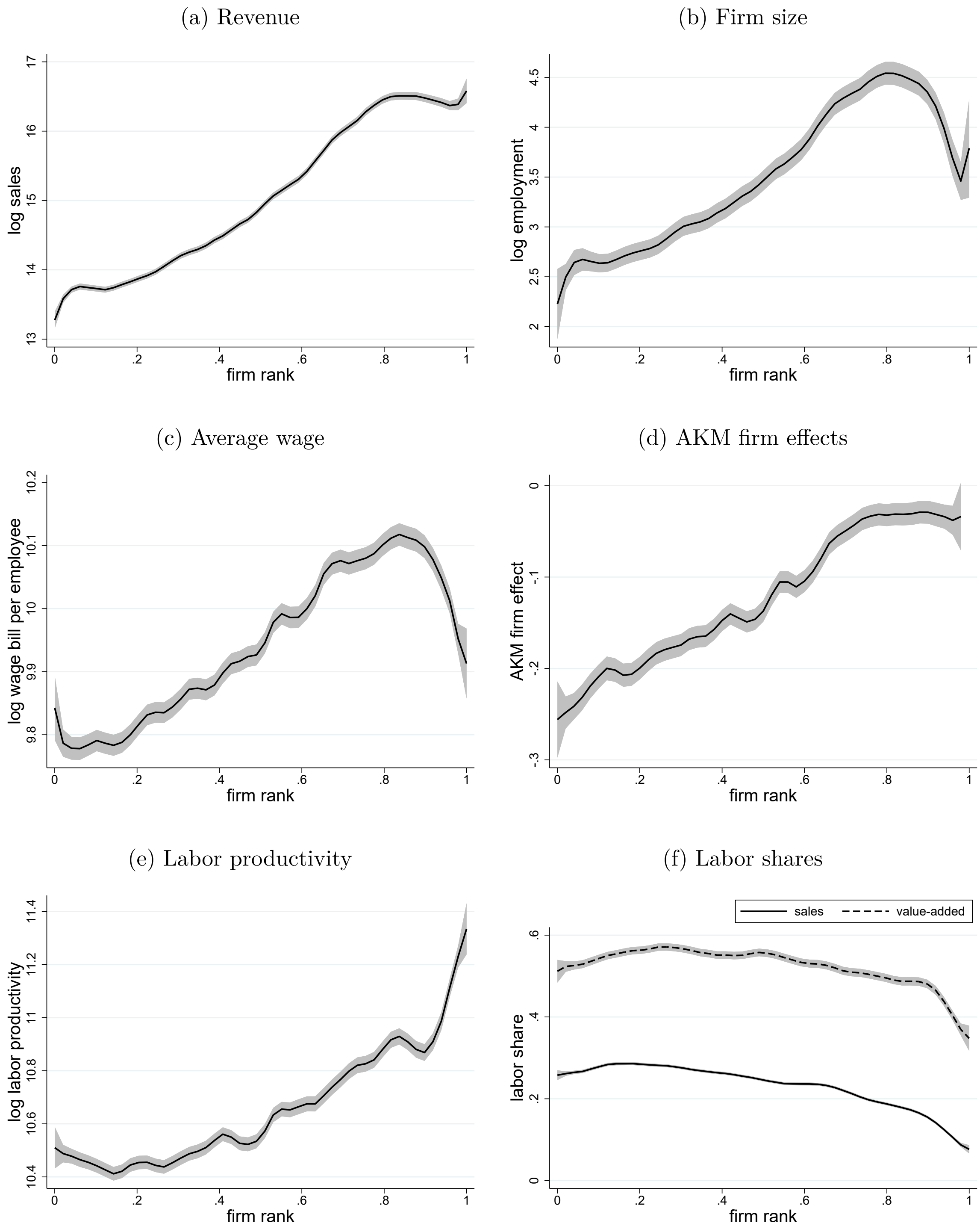
The log wage bill per employee (the average wage) is shown in Panel (c). Above the 80th percentile, average wages fall. Notably, the largest firms are also situated near the 80th percentile; therefore, average wages are closely related to size but less related to productivity. Interestingly, the least productive firms pay higher average wages than other firms below the 25th percentile. The AKM firm effect in Panel (d) is essentially flat above

[21] Guiso et al. (2005) show that firms fully insure their workers against transitory productivity shocks but only partially against the persistent productivity changes captured by $\hat{\omega}_{jt}$.

[22] See the ACF moment conditions, which we discuss in appendix C, equation C.5.

[23] Recent evidence by Lachowska et al. (Forthcoming) supports this conjecture.

Figure 2: Firm Performance Measures by Estimated Firm Rank




Source: Authors' calculations based on the BHP, EP, BeH.
Note: Estimated univariate kernel densities for selected firm performance measures across estimated firm ranks, normalized to be between zero and one. Kernel: Epanechnikov. Bandwidth: 0.05 for (a)–(c), (e) and 0.04 for (d) and (f). The qualitative findings are robust to the bandwidth choice. The 95% confidence bands are shown in gray.

the 80th percentile. Thus, firms within the top productivity quintile are indistinguishable in terms of AKM firm effects.

Panel (e) shows that log labor productivity (value added per worker) is relatively flat in the bottom half of the distribution but increases substantially for the most productive firms. At the most productive firms in the distribution, it is approximately twice as high as that of the firms at the bottom. Panel (f) shows the labor shares for revenue and value added. Both labor shares exhibit a hump shape but, for the most part, fall in the estimated firm rank. The labor share of revenue falls from just below 30% among the least productive firms to less than 10% among the most productive firms. For value added, the labor share falls from 50–60% to approximately 35%. Thus, the average worker obtains a significantly lower share of the output at high-productivity firms.

We group all individual firms into 15 productivity bins of equal size.[24] The bins exhibit a high within-bin wage variance (94% of the total). This is due to controlling for worker ability when estimating firm productivity, and it implies that our firms cannot be categorized into high-wage and low-wage firms. Within-bin heterogeneity in output (value added, value added per worker) and size (headcount, assets) is sizable, which implies that firms of similar sizes or production capacities are not necessarily in the same bin.

# 5 Productivity Sorting

To measure the degree of productivity sorting, we merge the two data sets containing our estimated worker and firm rankings. In this step, we lose all employment spells at firms that are not part of the EP sample. For this reason, we rebin workers and firms in the merged sample.[25] We rely on the final samples defined in Section 3.

## 5.1 Rank Correlations

We compute Spearman's $\rho$ to study the associations between the estimated worker and firm ranks in the data. For all matches in all years, we find a significantly positive but relatively low rank correlation of 0.07. For new matches, this correlation is 0.12. The correlation for J2J moves (0.11) is somewhat lower than the correlation for new matches OON (0.13).[26] These positive rank correlations reflect PAM based on worker ability and firm productivity. However, the degree of productivity sorting is low, both in an absolute sense and compared to measures of wage sorting.[27]

[24]There are 668 establishments in each bin.

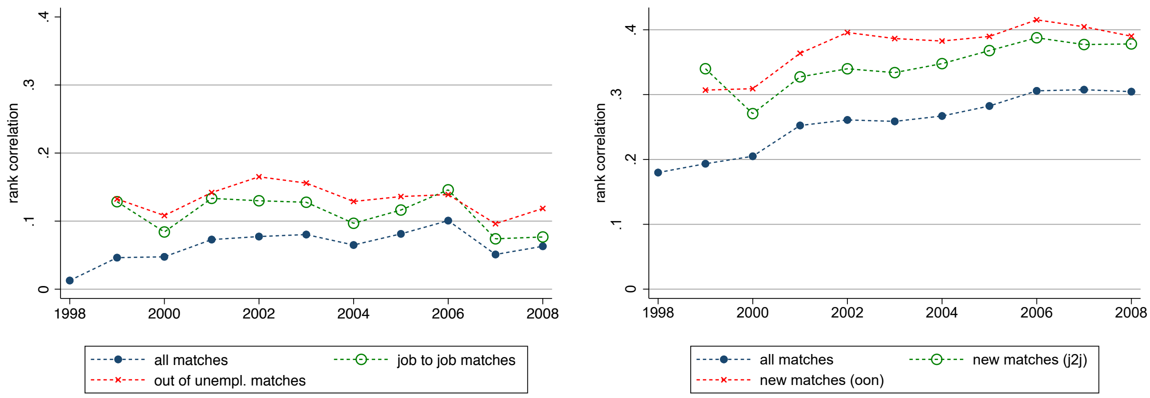
[25]The number of workers per bin decreases to 26,888. The number of firms per bin is stable (667).

[26]For clarity, the rank correlations presented here are calculated based on the worker and firm bins. Correlations based on worker and firm ranks (not reported) differ only marginally. Table D.2 provides an overview of all estimated rank correlations for different samples and time periods.

[27]Using German data, CHK and Hagedorn et al. (2017) find correlations of 0.17–0.25 and 0.71, respectively. CHK use the AKM model and compute their measure of wage sorting by correlating estimated

Figure 3: Rank Correlation Coefficients over Time (1998–2008)

(a) Firm ranking: productivity (b) Firm ranking: AKM firm effects




Source: Authors' calculations based on the BHP, EP, BeH.
Note: Panel (a) shows Spearman rank correlation coefficients based on our productivity-based ($\omega$) firm ranking. Panel (b) ranks firms using the AKM firm fixed effects. Both plots use the worker ability ranking introduced in Section 4.1.

Next, we consider how the estimated rank correlations change over time. Figure 3 presents rank correlations for all samples and all years. In Panel (a), the correlations are calculated using the productivity-based firm ranking. The solid circles depict the correlations for all matches. It first increases and then stabilizes just below 0.1 before dropping in 2007. For both types of new matches, the correlations are higher but also fall toward the end. The degree of sorting OON (crosses) decreases steadily after 2002. For J2J switches (hollow circles), the time dynamics closely mirror those for all matches because approximately 65% of all new matches are J2J switches (see also Table D.2). Thus, new matches drive up the overall rank correlation in the beginning; however, once their correlation decreases, the correlation for all matches also levels off.

Panel (b) of Figure 3 facilitates the comparison with wage-based approaches. We compute rank correlations using the same worker ranking but use the AKM firm fixed effects to rank firms. Notably, this implies higher rank correlations (more positive sorting) and a significant increase over time. The correlation for all matches reaches 0.3 near the end of our sample period, which is close in magnitude to that obtained by CHK. For new matches, the correlations are even higher, just below 0.4. New matches OON exhibit the highest rank correlations with both the AKM and the productivity-based firm rankings.

## 5.2 Distributional Dynamics

Rank correlations are only a summary statistic for the allocation of workers to firms. To investigate which worker–firm-type combinations contribute to changes in productivity sorting, we study how the matching of worker and firm types has changed over time.

worker and firm effects. Hagedorn et al. (2017) study sorting through the lens of a structural model with worker and firm heterogeneity. Both CHK and Hagedorn et al. (2017) rely on wages and observed worker mobility to rank firms, while we rank firms based on their estimated productivity.

To this end, we estimate univariate density functions for the employed worker types. Figure D.4 (online appendix) presents these densities for low-productivity firms (bins 1–2), medium-productivity firms (bin 8), and high-productivity firms (bins 14–15). We compare the first half of our sample period (1998–2002) to the second half (2003–2008) and show estimated densities for all matches, new matches OON, and J2J flows.

Low-type firms hire mainly low-ability workers OON. This is where the density of new matches is highest, and it clearly falls in the worker type. This is consistent with positive productivity sorting. However, the average quality of new hires in low-productivity firms has increased over time. Low-type firms now hire more medium-ability workers and fewer low-ability workers, both OON and through J2J flows. Considering all matches, the densities of the worker types hired by bin 1 firms have become more uniform over time. That is, in addition to making new matches of higher quality, these firms also have sizable outflows of high-ability workers. Taken together, these changes in low-productivity firms confirm the aggregate trend toward less productivity sorting.

For medium-type firms, we observe a strong increase in the hiring of medium-ability workers both through J2J flows and OON. For very productive firms, an increase in the hiring of medium-ability workers is also present but less pronounced. This increase is paired with a notable decrease in the hiring of high-ability workers for new matches made both through J2J flows and OON. Considering all matches, we see a significant reduction in the hiring of high-ability workers and an increase for medium-ability workers, although that increase among all firms is smaller than for medium-productivity firms alone. Only for the most productive firms can we discern a small but significant increase in new matches with high-ability workers through J2J flows, although the density for all matches indicates that the allocation of high-ability workers at the very top is quite stable.

In summary, high-ability workers have, to some extent, been replaced by medium-ability workers. Moreover, low-productivity firms have reduced their hiring of low-ability worker types and increased the quality of their average worker. The stable rank correlation is thus the result of two opposing trends: reduced productivity sorting at the bottom and the top (fewer low–low and high–high matches) and increased sorting in the middle (more medium–medium matches). In the final section, we analyze how these changes in allocations are related to the wages that workers earn at different firms.

# 6 Wages and Inequality

## 6.1 Wage Variation across Worker and Firm Bins

We investigate how the wages of different worker types vary across firm types. Figure 4 depicts the resulting wage–productivity profiles for five groups that include ten worker bins each. Panel (a) includes all matches, while new matches are shown in Panel (b).[28] The wage–productivity profiles exhibit a characteristic S-shape, i.e., low-productivity firms pay relatively high wages. The lowest wages are paid by firms in bins 3–4. Wages then increase monotonically up to firm bins 11–12 and decrease thereafter.[29] Moreover, the wage drop among the most productive firms decreases in worker ability. It is most pronounced for low- and medium-ability workers.[30] For high-ability workers, the wage levels off (bins 41–50, new matches).

This nonmonotonic, hump-shaped relation between wages and firm productivity poses a challenge for empirical models of wage dispersion. From a theoretical perspective, complementarities may imply that wages are a hump-shaped function of firm type (e.g., Gautier and Teulings, 2006; Eeckhout and Kircher, 2011). However, many structural models cannot generate a hump-shaped relation between wages and firm productivity. Consider two widely used models of on-the-job search (OJS). The most productive firm pays the highest wage in the wage-posting model of Burdett and Mortensen (1998). In the sequential auction model of Postel-Vinay and Robin (2002), wages decline in the poaching firm's productivity. This can explain the U-shaped relation at the bottom (firm bins 1–5) but not the hump-shaped relation at the top. We further discuss the potential for OJS models to rationalize our findings in Section 7.

In the model that we use to motivate our empirical strategy, wages are monotonic in firm productivity. They increase in firm productivity at a decreasing rate (see online appendix A.4). To generate a hump shape, wages have to be locally decreasing in firm productivity. To explore how this could arise in an extended model, we take a closer look at the two key features of the empirical wage–productivity profiles: the observed wage drop among the most productive firms and the fact that this drop decreases in worker ability. To this end, we rely on the estimated AKM wage components. Online appendix Figures D.7 and D.8 reproduce Figure 4 using the firm fixed effects and AKM residuals on the vertical axis. For the firm fixed effects (Figure D.7), we observe a hump shape at the top of the distribution across all worker types, including high-ability workers. This suggests that the wage decrease in the most productive firms is driven by firm-specific factors. According to our model, the firm fixed effect captures a time average of the effect of firm productivity on wages. In an extended model with multidimensional firm heterogeneity, it would also absorb the wage effect of, e.g., nonwage characteristics of jobs (amenities), which have the potential to reduce wages. We discuss how adding multidimensional heterogeneity would change our model and empirical approach in Section 7.

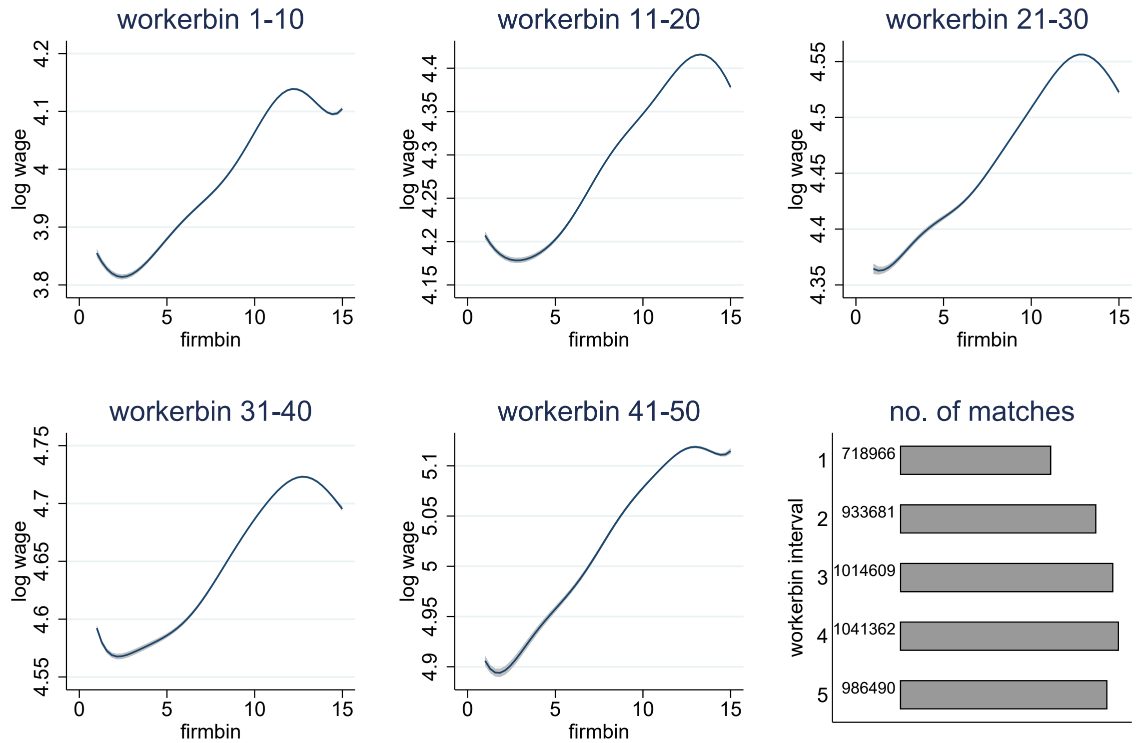
[28]The plots for matches OON and J2J flows are relegated to the online appendix, Figure D.6.

[29]The maximum around firm bin 12 is consistent with the firm performance measures shown in Figure 2, Section 4.2. These firms (around the 80th percentile of the productivity distribution) pay the highest average wages and are the largest firms in terms of headcount. The firms above are smaller, pay less on average, and exhibit greater labor productivity as well as lower labor shares.

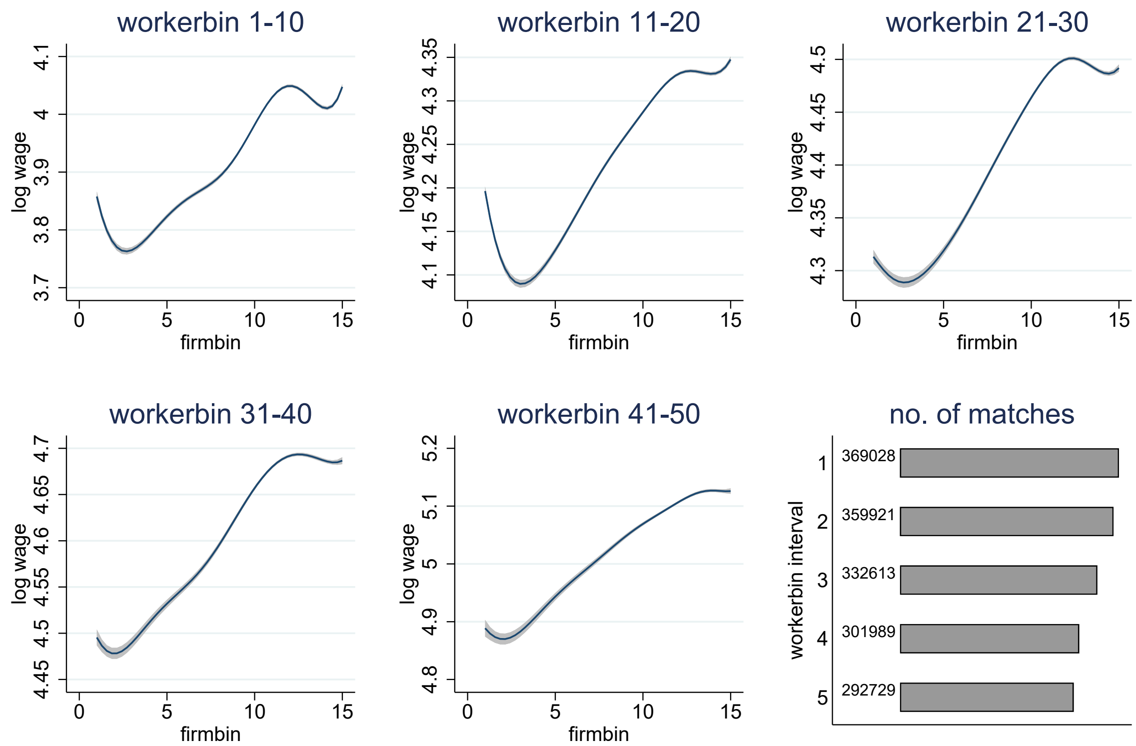
[30]The average wage difference between bin 15 and bin 12 firms for a bin 11–20 worker in terms of (deflated) log daily wages is approximately 4% (all matches), or approximately 1,177 euros annually.

Figure 4: Wage–Productivity Profiles

(a) All Matches




(b) New Matches

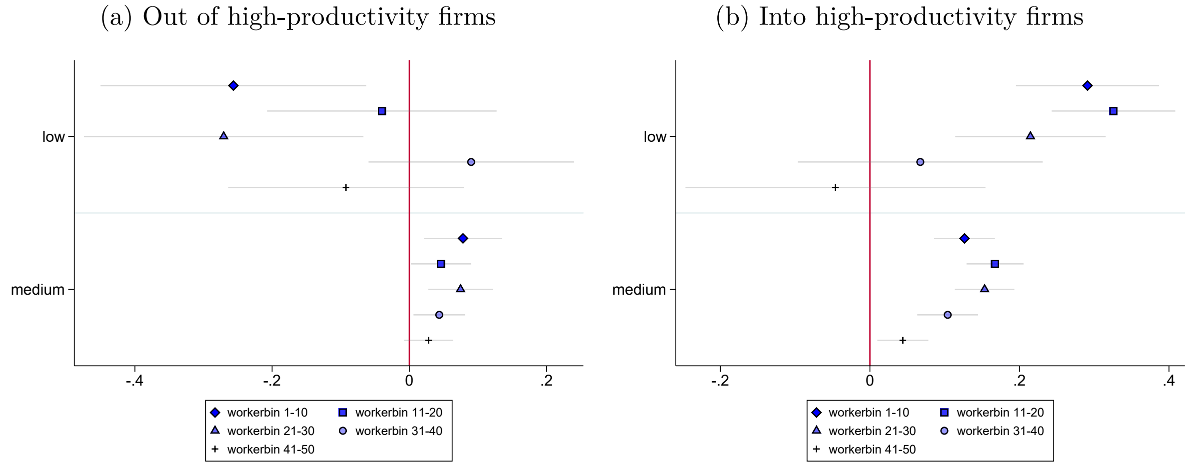



Source: Authors' calculations based on the BHP, EP, BeH.
Note: The plots show the estimated wage–productivity profiles across firm bins for all matches (a) and new matches (b). Based on kernel-weighted local polynomial regressions. Kernel: Gaussian. Bandwidth: 2. Shaded areas are 95% confidence bands.

For the AKM residuals (Figure D.8), we observe large positive values for the matches between the most productive firms and high-ability workers. For these worker types, the positive wage residuals compensate for the relatively small firm-specific pay components among high-productivity firms, so the hump shape is less pronounced. According to our model, the residual absorbs the effect of transitory productivity shocks on wages. The quantitative importance of this channel might be higher among the most productive firms, and it might also covary with worker ability. Moreover, the positive residuals could indicate that production complementarities are more complex than those in our simple model. For example, high-ability workers might increase the marginal product of other worker types, which could, in turn, add to their wage. Cahuc et al. (2008) study such mechanisms in detail, but they are absent in our simple model due to the assumptions needed to nest the AKM approach. In Section 7, we explore how more complex production functions relax the link between wages and productivity and how they would affect our empirical approach.

Finally, we check the robustness of the empirical wage–productivity profiles. Concerns about their validity might be related to measurement error in the data used to rank workers and firms. On the worker side, we use the universe of employment spells available through the German social security registers. The estimated AKM worker effects could be sensitive to the number of years included. On the firm side, we rely on survey data, which might be prone to sampling error. Moreover, the firm-year productivities $\hat{\omega}_{jt}$ could be measured with error. We first assess the sensitivity of the wage–productivity profiles to the inclusion of fewer years in the sample used to rank workers (2003–2008). The S-shape is robust (see Figure D.9). Second, we re-estimate the wage-productivity profiles using 20 random firm subsamples (drawn with replacement and clustered at the firm level). The S-shape is robust (see Figure D.10). Third, to alleviate concerns about measurement error in $\omega_{jt}$, we bootstrap the ACF estimation and assess the dispersion of estimated productivities at the firm-year level (see Figure D.11a). The mean deviation is less than 0.2% and more than 90% of errors lie within ± 9% of the bootstrapped firm-year-level mean. For every firm-year, we draw a new omega from the respective 95% confidence interval using uniform random numbers, rerank and rebin the firms, and reestimate the wage-productivity profiles. We repeat this 50 times. Again, the S-shape is robust (see Figure D.11b). Fourth, Figure D.12 confirms that the nonmonotonicities are not related to tenure effects. We re-estimate the wage–productivity profiles using the first match-year only. If anything, the wage drop at the top is more pronounced in this case. Additionally, Figure D.13 shows that firm types based on AKM firm fixed effects do not reveal any nonmonotonicities, as one would expect.

Figure 5: Wage Changes for Observed Transitions




Source: Authors' calculations based on the BHP, EP, BeH.
Note: The plots show estimated coefficients and 95% confidence intervals (robust standard errors) from a linear regression of individual-level wage differences among transitioning workers on dummies for origin and destination firm bins. The sample consists of new matches (J2J switches, no intermittent 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 axis captures the destination/origin firm bin groups: low (bins 1–3) and medium (bins 4–12).

## 6.2 Worker Transitions

The nonmonotonicities documented in Figure 4 reflect the mean wage differences across firm types for a given set of worker types. The next step is to check whether observed worker transitions between firm bins are consistent with these wage–productivity profiles; i.e., do workers transition toward higher wages, even if this implies switching to a less-productive employer? In this case, transitions out of the most productive firms should lead to wage gains, at least for a subset of destination firm bins. Similarly, when moving up the productivity ladder, wage gains should be decreasing in the origin firm bin and might even be negative for transitions to the most productive firms.

Figure 5 shows that the wage changes for observed J2J transitions largely support these conjectures. We regress log wage differences for the same five groups of worker bins on a set of origin and destination firm bin dummies.[31] To simplify the graphical illustration, we also group the firm bins into high (13–15), medium (4–12), and low (1–3) bins. Panel (a) shows the estimated wage changes for transitions out of high-productivity firms. Low- and medium-type workers experience significant wage gains of 7.8% (bins 1–10), 4.6% (bins 11–20), 7.5% (bins 21-30), and 4.4% (bins 31–40) when they transition from high-productivity to medium-productivity firms. For the highest-ability worker types, the wage differences are not significantly different from zero, which is consistent with the muted wage drop for these worker types. Wage differences for transitions from high-productivity to low-productivity firms are quite noisy and in most

[31] We measure the difference between the wage during the last spell in the pretransition firm and the wage during the first spell in the posttransition firm.

cases not significantly different from zero. Overall, we find that transitions down the firm productivity ladder lead to wage gains for many worker types and that even transitions into low-productivity firms do not always come with a negative wage change. These findings suggest that job mobility and sorting are guided by wages and not by firm productivity. This also explains the trend of decreasing productivity sorting at the top.

Panel (b) presents the estimated wage changes for transitions into high-productivity firms in a similar manner. As conjectured, we find significant positive wage changes for most worker types, and they decrease in the origin firm type. For workers in bins 31–50, we cannot reject the hypothesis that transitioning from a low- to a high-productivity firm yields no wage gain, which is in line with the S-shaped wage-productivity profiles. Low- to medium-ability workers (1–30) can expect significant wage gains of more than 20% when moving from a low-productivity to a high-productivity firm. When the origin firm is of medium productivity, wage gains are between 4.4% and 16.7% and decrease in worker ability. Based on the wage drops at the top in Figure 4, one might expect that some of these transitions lead to wage cuts. This is not the case, which reaffirms that workers tend to move toward higher wages.

In summary, workers select jobs to maximize their wages, even when this implies moving to a less productive firm. Upward transitions typically lead to wage increases as well, so workers do not accept wage cuts to move to the most productive firms. We again test the robustness of our findings by re-estimating the wage changes using random subsamples. All findings are robust (see Figure D.14).

## 6.3 Wage Inequality

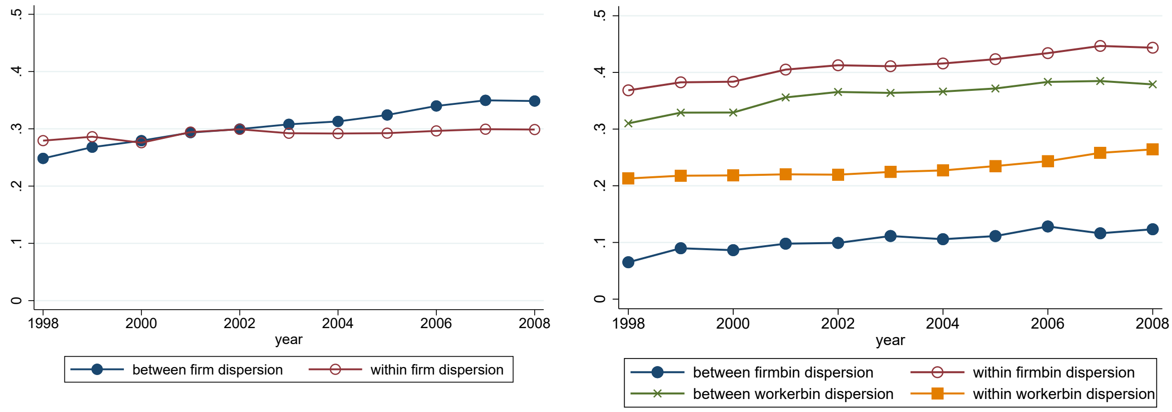
Song et al. (2019) show that two-thirds of the increase in wage inequality in the U.S. from 1978 to 2013 can be attributed to increasing pay differences between firms. Increasing wage sorting (high-wage workers into high-wage firms) and increasing segregation of workers contribute roughly equally to the rising contribution of firms to wage inequality. Panel (a) of Figure 6 shows that this trend is also present in the German data. We use establishment identifiers to decompose the variance in wages into the respective shares explained within and between establishments. Between 1998 and 2008, wage inequality grew because the between-firm component (filled circles) grew by approximately 10%. The within-establishment contribution (hollow circles) is of similar magnitude but stable.

In Panel (b), we decompose wage inequality based on firm productivity and worker ability. The variance between the firm bins (filled circles) does increase but only by approximately 3%. It contributes little to overall inequality. The variance within firm bins (hollow circles) is most important. Worker segregation, that is, wage differences between worker bins (crosses), is also a major contributing factor. The increase in these two components dominates the increase in between-firm inequality. The contribution of

Figure 6: Decomposition of Wage Dispersion over Time

(a) Song et al. (2019) replication

(b) Our approach




Source: Authors' calculations based on the BHP, EP, BeH.
Note: Panel (a) shows the yearly decompositions of the wage dispersion using establishment identifiers. Panel (b) shows the yearly decompositions of the wage dispersion into the respective contributions of within- and between-estimated worker bin variance and within- and between-estimated firm bin variance.

within-worker bin inequality (squares) is moderately large, but it increases only in the second half of the sample period. We conclude that increasing pay differences between firm productivity types are not the major reason for increasing wage inequality in Germany.

We link the increasing variance of the different pay components to the wage–productivity profiles in Figure 4 by analyzing how the components change over time (1998–2002 and 2003–2008); see Figure D.15. There are large differences in wage growth across worker types, and they correspond to increasing worker segregation (the between-worker component in Figure 6b). For low-ability workers, wages decrease by more than 10% for all firm types above the very bottom. For medium-ability workers, wages shrink among low-productivity firms but are relatively stable or somewhat increasing among more productive firms. For high-ability workers, wages increase among almost all firm types but most notably at the top. However, the hump shape has also become more pronounced over time, which explains the decreasing sorting of high-ability workers into high-productivity firms. Overall, the wage–productivity profiles become steeper, which implies an increase in between-firm wage inequality consistent with Figure 6b.

# 7 Extensions

We consider whether extending the model to allow for OJS, general production functions, and multidimensional heterogeneity would be compatible with the AKM approach and nonmonotonic wage–productivity profiles.

We have assumed that only unemployed workers search for jobs. If the DGP featured OJS, workers' wages could depend on their employment history; worker types would be identifiable only through wages out of unemployment. Researchers commonly use functional form assumptions that separate worker ability and labor market history to circumvent this issue. Postel-Vinay and Robin (2002) show the wage is log-linear in the worker type in the sequential auction model if the utility function is CRRA. Our model assumes that worker ability units are perfectly substitutable at the firm level, so the marginal product of labor scales with worker type, implying additive separability in logs. This similarity makes us suspect that there are production functions for which OJS is compatible with a wage equation that is log-linear in the worker type, and, therefore, with using the AKM approach.[32]

Adding OJS to our model is unlikely to generate hump-shaped wage–productivity profiles. In the sequential auction model, the productivities of both the poaching firm and the previous employer affect the wage. The wage increases in the previous employer's productivity, but decreases in the poaching firm's productivity due to the option value of future wage gains. Overall, the effect of firm productivity on wages is nonlinear. However, the magnitude of the second effect decreases in firm productivity because workers become increasingly unlikely to meet an even more productive firm as they move up the productivity ladder. This implies a U-shaped, not a hump-shaped, relation between firm productivity and wages. Thus, OJS models can yield a log-linear (in the worker type) wage equation, but not a hump-shaped wage–productivity profile.

The simple production function motivating our empirical strategy is inconsistent with a hump-shaped relation between productivity and wages because wages increase with firm productivity, albeit at a decreasing rate (see Appendix A.4). The reason is that high productivity increases the effect of worker ability on output, but decreasing returns to labor dampen this effect. As in Cahuc et al. (2008), firms overemploy workers to reduce their marginal product and pay lower wages. With complementarities among worker types, employing a worker of one type increases other types' marginal products and wages. This effect could outweigh the effect of decreasing returns and interact with firm productivity, heterogeneous worker bargaining power, or the firm's capital stock. Such mechanisms seem flexible enough to generate a nonmonotonic relation between productivity and wages and a less pronounced hump shape for high-ability workers. However, a more complex production function would probably make the wage equation nonlinear in the worker type, and, thus, inconsistent with the AKM model.

In addition, job amenities can create wage differences between equally skilled workers (Rosen, 1986; Mortensen, 2003). If the most productive firms offered better amenities than somewhat less productive firms, wages at the top firms would be relatively low. To

[32] Di Addario et al. (Forthcoming) show that an augmented AKM model with origin- and destination-firm effects resembles the wage equation of sequential auction models for specific utility functions. They find a small covariance between the worker fixed effect and the origin-firm effect in Italian data, which supports the separability of worker ability and labor market history.

test this hypothesis, we use the EP survey to construct a firm-level amenity index.[33] We regress this index on the productivity rankings, controlling for (2-digit) industry and firm size (headcount). The most productive firms indeed offer the most amenities (see Figure D.16). We also compute the rate of employer-initiated separations. The most productive firms have the lowest firing rates (see Figure D.17). This higher job security could entice workers to accept lower wages.

The finding that high-productivity firms provide a high level of amenities can explain the hump shape at the top. However, less productive firms pay low wages despite offering few amenities, so wages depend on both firm productivity and amenities (and worker ability). Whether AKM can be applied in a setting with two-dimensional firm heterogeneity depends on the production function. With homogeneous worker preferences for amenities and assuming that amenities do not directly affect the marginal product of labor, nothing changes. The wage effect of amenities would be absorbed by the firm fixed effect. If workers' preferences over amenities are also heterogeneous, the composition of the workforce may depend on amenities (as in Lamadon et al., 2022).[34] However, if the composition does not affect output, e.g., due to the perfect substitutability of ability units, nothing changes. If heterogeneous amenity preferences are combined with richer complementarities in the production function, the marginal product of a given worker type could depend on amenities because they might attract other types that affect the marginal product. In this case, we would again require a more flexible approach to ranking workers.

Another issue is whether amenities could change the estimated productivity rankings. Our labor inputs take into account ability heterogeneity, so the composition of the workforce may depend on amenities. Beyond the composition, amenities do not affect output. Thus, adding amenities as a second dimension of firm heterogeneity to our model should not affect the estimated firm rankings. Reassuringly, adding the amenity index as a control variable leaves the rankings virtually unchanged.[35]

# 8 Conclusions

We present a parsimonious search-matching model that elucidates the assumptions necessary to combine the AKM and ACF approaches to study labor market sorting and wage dispersion. The model is compatible with both the discrete-time model of dynamically optimizing firms underlying ACF and the log-linear wage equation underlying AKM.

[33] The survey contains questions about the flexibility of working hours, working time accounts, parental benefits, and other amenities. We define an indicator as 1 if an amenity exists and 0 otherwise. For each establishment year, we sum up the indicators and divide by the total number of indicators for that year.

[34] Lamadon et al. (2022) combine heterogeneous preferences for amenities with monopsonistic wage-setting in a model of imperfect competition. Consistent with our findings, they find that high-productivity firms offer a high level of amenities and pay relatively low wages.

[35] These results are available upon request.

Based on worker ability (inferred from the AKM worker effects) and firm productivity (inferred using the ACF approach), we study productivity sorting, wages across firm productivity types, and inequality trends in the German labor market. Productivity sorting is positive, low, and relatively stable over time. Among the most productive firms, sorting decreases as high-ability workers become more likely to work at somewhat less productive firms that pay higher wages. Thus, sorting into high-productivity firms is quantitatively less important for rising wage inequality than wage sorting. If workers move away from high-productivity firms to increase their wages, a side effect of this increasing wage sorting could be decreasing allocative efficiency in the labor market. Thus, studying productivity sorting is a useful complement to the wage-based analysis of sorting.

We discuss several model extensions in regard to their potential to match the facts that we document. Canonical OJS models are compatible with log-linear (in the worker type) wage equations but incompatible with the hump-shaped relation between wages and productivity. With more general production functions, our framework could generate the hump shape, but the AKM model would no longer be nested. Models with multidimensional firm and/or worker heterogeneity can generate a hump shape with relatively simple production functions and thus remain compatible with the AKM approach.

# References


Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. High wage workers and high wage firms. *Econometrica* 67 (2): 251–333.

Abowd, John M., Francis Kramarz, Sébastien Pérez-Duarte, and Ian M. Schmutte. 2018. Sorting between and within industries: A testable model of assortative matching. *Annals of Economics and Statistics* 129: 1–32.

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. Identification properties of recent production function estimators. *Econometrica* 83 (6): 2411–2451.

Bagger, Jesper, Bent Jesper Christensen, and Dale T. Mortensen. 2014. Productivity and wage dispersion: Heterogeneity or misallocation? Unpublished manuscript, Department of Economics, Royal Holloway, University of London.

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

Bartolucci, Cristian, Francesco Devicienti, and Ignacio Monzón. 2018. Identifying sorting in practice. *American Economic Journal: Applied Economics* 10 (4): 408–38.

Becker, Gary S. 1973. A Theory of Marriage: Part I. *Journal of Political Economy* 81 (4): 813–846.

Burdett, Kenneth and Dale T. Mortensen. 1998. Wage differentials, employer size, and unemployment. *International Economic Review* 39 (2): 257–273.

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, Joerg Heining, and Patrick Kline. 2018. Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics* 36 (S1): S13–S70.

Card, David, Jörg Heining, and Patrick Kline. 2013. Workplace heterogeneity and the rise of West German wage inequality. *The Quarterly Journal of Economics* 128 (3): 967–1015.

Di Addario, Sabrina, Patrick Kline, Raffaele Saggio, and Mikkel Sølvsten. Forthcoming. It ain't where you're from, it's where you're at: Hiring origins, firm heterogeneity, and wages. *Journal of Econometrics*.

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

Eeckhout, Jan and Philipp Kircher. 2011. Identifying sorting–in theory. *Review of Economic Studies* 78 (3): 872–906.

——— 2018. Assortative matching with large firms. *Econometrica* 86 (1): 85–132.

Fischer, Gabriele, Florian Janik, Dana Müller, and Alexandra Schmucker. 2009. The IAB Establishment Panel - things users should know. *Schmollers Jahrbuch: Journal of Applied Social Science Studies* 129 (1): 133–148.

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.

Fox, Jeremy T. and Valérie Smeets. 2011. Does Input Quality Drive Measured Differences In Firm Productivity? *International Economic Review* 52 (4): 961–989.

Gautier, Pieter A. and Coen N. Teulings. 2006. How large are search frictions? *Journal of the European Economic Association* 4 (6): 1193–1225.

Guiso, Luigi, Luigi Pistaferri, and Fabiano Schivardi. 2005. Insurance within the firm. *Journal of Political Economy* 113 (5): 1054–1087.

Hagedorn, Marcus, Tzuo Hann Law, and Iourii Manovskii. 2017. Identifying equilibrium models of labor market sorting. *Econometrica* 85 (1): 29–65.

Haltiwanger, John, Henry Hyatt, and Erika McEntarfer. 2018. Who moves up the job ladder? *Journal of Labor Economics* 36 (S1): S301–S336.

Irarrazabal, Alfonso, Andreas Moxnes, and Karen Helene Ulltveit-Moe. 2013. Heterogeneous firms or heterogeneous workers? Implications for exporter premiums and the gains from trade. *Review of Economics and Statistics* 95 (3): 839–849.

Kölling, Arnd. 2000. The IAB-Establishment Panel. *Schmollers Jahrbuch: Journal of Applied Social Science Studies* 120 (2): 291–300.

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

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler. 2022. Imperfect competition, compensating differentials, and rent sharing in the US labor market. *American Economic Review* 112 (1): 169–212.

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.

Lopes de Melo, Rafael. 2018. Firm wage differentials and labor market sorting: Reconciling theory and evidence. *Journal of Political Economy* 126 (1): 313–346.

Mortensen, Dale. 2003. *Wage dispersion: why are similar workers paid differently?* Boston: MIT press.

Müller, Steffen. 2008. Capital stock approximation using firm level panel data. *Jahrbücher für Nationalökonomie und Statistik* 224 (4): 357–371.

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

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

Rosen, Sherwin. 1986. Chapter 12: The theory of equalizing differences. In *Handbook of Labor Economics*, eds. O. Ashenfelter and R. Layard. Amsterdam: Elsevier.

Shimer, Robert. 2005. The assignment of workers to jobs in an economy with coordination frictions. *Journal of Political Economy* 113 (5): 996–1025.

Shimer, Robert and Lones Smith. 2000. Assortative matching and search. *Econometrica* 68 (2): 343–369.

Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. 2019. Firming up inequality. *The Quarterly Journal of Economics* 134 (1): 1–50.

Sorkin, Isaac. 2018. Ranking firms using revealed preference. *The Quarterly Journal of Economics* 133 (3): 1331–1393.

Spengler, Anja. 2008. The Establishment History Panel. *Schmollers Jahrbuch: Journal of Applied Social Science Studies* 128 (3): 501–509.

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

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