Sorting and Wage Inequality*

Kory Kantenga[†]
Tzuo-Hann Law[‡]

This Version: February 12, 2016

Abstract

We measure the roles of the permanent component of worker and firm productivities, complementarities between them, search frictions, and equilibrium sorting in driving German wage dispersion. We do this using a standard assortative matching model with on-the-job search. The model is identified and estimated using matched employer-employee data on wages and labor market transitions without imposing parametric restrictions on the production technology. The model's fit to the wage data is comparable to prominent wage regressions with additive worker and firm fixed effects that use many more degrees of freedom. Moreover, we propose a direct test that rejects the restrictions underlying the additive specification. We use the model to decompose the rise in German wage dispersion between the 1990s and the 2000s. We find that changes in the production function and the induced changes in equilibrium sorting patterns account for virtually all the rise in the observed wage dispersion. Search frictions are an important determinant of the level of wage dispersion but have had little impact on its rise over time.

JEL: J24, J31, O15

Keywords: Wage Inequality, Sorting, Assortative Matching, Search Frictions

*We thank Marcus Hagedorn, Dirk Krueger, and Iourii Manovskii for invaluable suggestions. Comments and discussions at various seminars at the University of Pennsylvania, Pennsylvania State University, Boston College, Carnegie Mellon University, Federal Reserve Bank of Minneapolis, Federal Reserve Bank of St Louis, National University of Singapore, Singapore Management University, Temple University and Rutgers University were greatly appreciated. This study uses the Linked-Employer-Employee Data (LIAB) longitudinal model 1993-2010 (LIAB LM 9310) from the IAB. Data access was provided via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB), the IAB Research Data Center at the Cornell Institute for Social and Economic Research (CISER), Harvard Univerity, and subsequently remote data access under Project fdz777. All errors are ours.

†Department of Economics, University of Pennsylvania, 160 McNeil Building, 3718 Locust Walk, Philadelphia, PA, 19104-6297 USA. E-mail: kantenga@sas.upenn.edu

‡Department of Economics, Boston College, 333 Maloney Hall, 140 Commonwealth Avenue, Chestnut Hill, MA, 02467-3859 USA. E-mail: tzuo.law@bc.edu

1 Introduction

Cross-sectional wage dispersion increased substantially in the US between the 1970s and the 1980s. Lagging the experience of the US by about a decade, Germany experienced a similar, dramatic increase in wage dispersion from the 1990s to the 2000s.[1] Until recently, the literature mainly focused on understanding the increase in wage dispersion across observable dimensions of worker skills, such as education, age, experience, and occupation. It is well known, however, that these observable dimensions account for a relatively small share of the wage variance. In contrast, worker and firm fixed effects included in standard log wage equations are typically found to account for a larger amount of wage variance than all the observables combined. Moreover, changes in the dispersion of the estimated fixed effects and their correlation are found to be very important in accounting for the increase in wage variance over time. Card, Heining, and Kline (2013) document this for Germany, while Barth, Bryson, Davis, and Freeman (2014) report related evidence for the US. The empirical literature lacks a structural interpretation for these fixed effects. However, the findings that fixed effects are important for fitting both the level and increase in wage dispersion suggests that permanent heterogeneity across workers and firms is an important feature of the data.

Motivated by this descriptive evidence, we assess the role of the dispersion of the permanent component of worker abilities, the dispersion of firm productivities, and complementarities between the two in the production technology in determining the level and the rise in German wage dispersion. While these changes are exogenous from the point of view of our theory, they induce an endogenous response in wages and in the sorting of workers across employers. Moreover, even if they are fixed, the extent of frictions in the assortative matching process might change over time, for example due to the spread of new information technologies, generating an endogenous response of wages and sorting patterns. We attempt to disentangle and separately measure these ef-

[1] See Dustmann, Ludsteck, and Schönberg (2009).

fects. The key challenge, of course, is that neither workers' abilities nor firm productivities nor the production technology are directly observable in the data.

We conduct our analysis using the standard theory of assignment problems in heterogeneous agent economies which traces its roots to Becker (1973). Specifically, we use the state-of-the-art version of the model that allows for time-consuming search as introduced by Shimer and Smith (2000) and on-the-job search as in Hagedorn, Law, and Manovskii (2014).[2] The key distinction of our approach is that the identification strategy of Hagedorn, Law, and Manovskii (2014) does not impose parametric restrictions on the shape of the production technology. This is important because the production technology is the key object of interest in our analysis. In this model, sorting of workers across firms is guided by wages which reflect complementarities in the production technology. The data we use comes from a large matched employer-employee sample provided by the German Institute for Employment Research (IAB). To measure the changes in wage dispersion over time, we consider a sample spanning the 1990s and another spanning the 2000s.

In this model, the production technology is a production function that takes worker ability and firm productivity as inputs. The first step in our analysis involves nonparametrically estimating this production function. To do so, we implement the identification strategy in Hagedorn, Law, and Manovskii (2014).[3] First, we use the result that workers hired from unemployment can be ranked based on their wages within firms. Within-firm rankings are partial because each firm hires and therefore ranks a subset of workers in the economy. Workers who move between firms link these partial rankings. This enables us to solve a rank aggregation problem to effectively maximize the

[2]Related structural models of labor market sorting were estimated in Lise, Meghir, and Robin (2011) and Lopes de Melo (2013). Gautier and Teulings (2012), Abowd, Kramarz, Pérez-Duarte, and Schmutte (2014) and Bagger and Lentz (2014) estimated sorting models that are more fundamentally different.

[3]Lamadon, Lise, Meghir, and Robin (2014) study nonparametric identification of a related model that introduces a different model of on-the-job search into the environment of Shimer and Smith (2000).

likelihood of a correct global ranking. Second, we rank firms exploiting the result that the value of a vacant job increases in firm productivity. We measure this value using only data on wages and labor market transition rates. Third, we recover the production function. The production function can be recovered, because the observed out-of-unemployment wages of a match between a particular worker and firm in the model are a function of the match output and the value of a vacancy for that firm. Thus, the out-of-unemployment wage equation can simply be inverted for output. Although each worker is typically observed working at only a few firms, we estimate his output at other firms by considering how much similarly ranked workers (who actually work at the other firms) produce.

We make three empirical contributions. First, we show that a structural model can fit the data as well as prominent wage regressions do. Second, we show through a series of decompositions that the production function is primarily responsible for the increase in German wage dispersion. Third, we provide evidence in support of wages driving sorting patterns in the data.

As many of our decompositions will involve counter-factual experiments, it is important to verify that our model fits the wage data well. To do so, we use the estimated production function and the parameters describing search frictions to simulate equilibrium wages and ask if the simulated wages fit wages in the data. The model's fit to wage data is comparable to that achieved by prominent regressions with a fixed effect for every worker and a fixed effect for every firm in the dataset. In addition, the model fits mobility rates and sorting between workers and firms while using far fewer degrees of freedom.

To disentangle the contributions to wage dispersion due to changes in production complementarities from the induced endogenous response of sorting, we use the model to conduct counter-factual experiments that involve changing the estimated production function holding the match distribution fixed and changing the match distribution for a fixed production function. These experiments imply that the joint effect of changing complementarities and sorting account for almost all the increase in wage variance, while the direct effect of technological change and indirect endogenous response of sorting each account for about half of the increase.

Similar experiments that involve changing the estimated parameters governing search frictions imply that these changes have had only a minor effect on the change in wage dispersion. However, search frictions play a very important role in determining the level of wage variance. Given the estimated production function we can compute the wage dispersion that would arise in a frictionless model. We find that eliminating search frictions would *increase* wage dispersion. This finding may appear surprising given the standard result that search frictions tend to generate wage dispersion among homogeneous workers.[4] However, in our analysis, workers are heterogeneous and search frictions prevent them from fully exploiting the complementarities in the production process, which lowers the cross-sectional wage variance in equilibrium.

As mentioned, a prominent alternative approach to studying wage dispersion in the literature estimates log wage equations that include additive worker and firm fixed effects. Gautier and Teulings (2006) and Eeckhout and Kircher (2011) show that standard assortative matching models based on comparative or absolute advantage do not give rise to log wages that are linear in worker and firm fixed effects. Instead, it is the nonlinearities of wages, reflecting in part the production complementarities, that guide the sorting process in the model. Yet, log wage regressions with worker and firm fixed effects fit the wage data very well with the R^2 often in excess of 0.9 across many datasets (e.g. Germany, Denmark, France, and US). We explore whether the nonlinearities at the core of the theory can be directly detected in the wage data.[5] Consider two workers of different ability x and x' working at a firm with productivity y who both move to a firm with productivity $y' \neq y$ and earn log wages $\log w(\cdot, \cdot)$. Linearity in fixed effects restricts log wage differentials to be equal

[4]For example, see Burdett and Mortensen (1998).

[5]Nonlinearities refer to deviations from log additive separability.

when workers switch firms

$$\log w(x,y) - \log w(x',y) = \log w(x,y') - \log w(x',y').$$

We develop a statistical test based on this restriction and find that log wage differentials vary across firms, indicating the presence of nonlinearities. The additive specification rules out nonlinearities that we detect in the data, because it imposes that firms pay a firm-specific wage premium to all workers. Thus, the firm effect cancels out when considering the log wage difference for the workers in the same firm.

The same test applied to model-generated data yields comparable results. Hence, in this aspect, this structural model (in which nonlinearities in wages drive sorting) fits the data while the additive specification does not. Moreover, the linear regression with additive worker and firm fixed effects yields an R^2 above 0.9 when estimated on model-generated wages. This is the case even though complementarities in production induce substantial nonlinearities in model-generated wages. This suggests that a high R^2 with an additive specification is insufficient to conclude that the additive specification is a meaningful description of the data. To the contrary, nonlinearities feature prominently in the data and drive sorting.

The remainder of the paper is organized as follows. In Section 2, we present the model and summarize the identification strategy. In Section 3, we describe the data. The empirical performance of the model is assessed in Section 4. In Section 5, we use the estimated model to perform the counter-factual experiments that isolate the sources of the rise in the German wage dispersion. In Section 6, we test whether wage differentials differ across firms, which regressions with worker and firm fixed effects preclude. Section 7 concludes.

2 Model and Identification

We use the on-the-job search model presented in Hagedorn, Law, and Manovskii (2014). Their identification strategy relies only on wages and job transitions observable in standard matched employer-employee datasets. The model's theoretical foundation is Becker (1973), where wages and allocations reflect production complementarities in a frictionless setting. Becker's framework stresses the role of wages in guiding the assignment of workers to firms but is not well suited for labor market applications due to its lack of frictions. To extend Becker's framework, this model incorporates the frictional search and vacancy posting environment of Shimer and Smith (2000). Only unemployed workers search in Shimer and Smith (2000), whereas job-to-job moves are common in the data. To accommodate this, Hagedorn, Law, and Manovskii (2014) build in on-the-job search in the spirit of Cahuc, Postel-Vinay, and Robin (2006).

2.1 Model

Time is discrete. Agents are risk neutral, live infinitely, and maximize present value of payoffs discounted by a common discount factor $\beta \in (0,1)$. A unit mass of workers are either employed (e) or unemployed (u) while p_f mass of firms are producing (p) or vacant (v). Workers and firms have heterogeneous productivities. Without loss of generality, their productivity *rank* is denoted by $x \in [0,1]$ and $y \in [0,1]$, respectively. When matched, worker x and firm y produce $f(x,y)$ where $f : [0,1]^2 \to \mathbb{R}_+$. Consistent with x and y being productivity ranks, $f_x > 0$ and $f_y > 0$. There are no other restrictions on f. We call f the production function and refer to the quantity $f(x,y)$ as the match output.[6]

[6]Defining productivity on ranks is without loss of generality. The rank of a worker (firm) is given by the fraction of workers (firms) who produce weakly less with the same firm (worker). In this paper, *productivity*, *rank*, or *type* have identical meanings. Therefore, the distributions of worker and firm types are both uniform. If the "original" (non-rank) distributions of worker and firm types are G_x and G_y respectively, and the "original" production

The functions characterizing the distributions of employed workers, unemployed workers, producing firms, and vacant firms are denoted $d_e(x), d_u(x), d_p(y)$ and $d_v(y)$, respectively. Since productivities are defined on ranks, $d_e(\cdot) + d_u(\cdot) = 1$ and $d_p(\cdot) + d_v(\cdot) = p_f$. The function describing the distribution of producing matches is $d_m : [0,1]^2 \to \mathbb{R}_+$. Aggregate measures of this economy are employment, E; unemployment, U; producing firms, P; and vacant firms, V. Specifically, $\int d_m = \int d_e = \int d_p = E = P$, $1 - \int d_m = \int d_u = U = \int d_v = V$. All these equilibrium objects that characterize distributions are constant in the steady state.

There are two stages in each period. In the first stage, matched workers and firms produce and the output is split into wages and profits. There is free entry. Entrant firms draw a fixed number of vacancies and type y from a uniform distribution. Entry costs c^e per vacancy.[7] Once in the market, firms pay maintenance cost c per unfilled vacancy per period. In the second stage, all workers and all vacancies engage in random search. The total search effort is $s = U + \phi E$ where $\phi \in [0,1]$ is an exogenous search intensity of employed workers (relative to unemployed workers). V denotes the number of vacancies. Meetings are generated by $m : [0,1] \times [0,1] \to [0, \min(s,V)]$ which takes the pair (s,V) as inputs. The probabilities that an unemployed or an employed worker meets a vacancy are given by $\mathbb{M}_u = \frac{m(s,V)}{s}$, and $\mathbb{M}_e = \phi\frac{m(s,V)}{s}$, while the probability of a vacancy meeting a potential hire (employed or unemployed) is $\mathbb{M}_v = \frac{m(s,V)}{V}$. Conditional on the meeting, the vacancy meets

function is $\tilde{f}(\tilde{x}, \tilde{y})$, then we transform the production function

$$f(x,y) = \tilde{f}(G_x^{-1}(x), G_y^{-1}(y))$$

and the distributions are $G_x(\tilde{x}) = x$, $G_y(\tilde{y}) = y$. It is easier to see this in one dimension. Let the "true productivity" of workers be given by $\tilde{x}$ distributed $G_x(\cdot)$ with support $[0, \bar{x}]$. The "true production function" is $\tilde{f}$ and hence, the output of a worker is $\tilde{f}(\tilde{x})$. Then, worker $\bar{x}$ produces $\tilde{f}(\bar{x})$, i.e. a worker with rank $x = 1$ produces $f(1) = \tilde{f}(\bar{x})$. Because $\tilde{x}$ and $\tilde{f}$ are unobserved, it is not possible to separately identify $\tilde{x}$ and $\tilde{f}$, e.g. $\tilde{f} = 3\tilde{x}$ with $\tilde{x} \in [0,1]$ and $\tilde{f} = \tilde{x}$ with $\tilde{x} \in [0,3]$ are observationally identical. This observation extends to two dimensions. Hence, the relevant object to measure in the data is $f(x,y)$ with $(x,y) \in [0,1]^2$.

[7] c^e is assumed to be such that the mass of jobs in the economy is equal to the mass of workers. That is, $p_f = 1$.

an employed worker with probability $\mathbb{C}_e = \frac{\phi E}{U+\phi E}$ and meets an unemployed worker with probability $\mathbb{C}_u = \frac{U}{U+\phi E}$.[8] Not all meetings result in matches, because some unemployed workers prefer continuing searching to matching with the vacancy they met and some employed workers prefer remaining in their existing matches. At the end of the period a match is destroyed with exogenous probability δ.

Denote the surplus received by an employed worker by S^o. The worker's surplus received depends on search history, as will become clear when we describe wage setting. Let $V_u(x)$ denote the value of unemployment for a worker of type x. $V_e(x, y, S^o)$ is the value of employment for a worker of type x at a firm of type y when the worker receives S^o. $V_v(y)$ is the value of a vacancy for firm y, and $V_p(x, y, S^o)$ is the value of firm y employing a worker of type x when the worker receives S^o. S^o does not affect the size of match surplus $S(x, y)$. It only determines the split of the surplus between the worker and the firm. Formally,

$$
\begin{aligned}
V_e(x, y, S^o) &:= V_u(x) + S^o && (1)\\
V_p(x, y, S^o) &:= V_v(y) + (S(x, y) - S^o) && (2)\\
S(x, y) &:= V_p(x, y, S^o) - V_v(y) + V_e(x, y, S^o) - V_u(x) && (3)
\end{aligned}
$$

We now describe wage setting, which determines S^o. An unemployed worker who meets a vacancy makes a take-it-or-leave-it offer and extracts the full surplus. As in Cahuc, Postel-Vinay, and Robin (2006), when a worker of type x employed at some firm $\tilde{y}$ meets a firm y which generates higher surplus, the two firms engage in Bertrand competition such that the worker moves to firm y. At the new firm y, the worker obtains the full surplus generated with firm $\tilde{y}$, $S(x, \tilde{y})$, while the new firm y retains $S(x, y) - S(x, \tilde{y})$. Small, unmodelled costs of writing an offer deter potential poaching firms from engaging in Bertrand competition unless they know the poaching attempt will succeed.

[8]We estimate $\mathbb{M}_v$, $\mathbb{M}_e$, $\mathbb{M}_u$, $\mathbb{C}_e$, and $\mathbb{C}_u$ directly without imposing functional form assumptions on m.

These modeling choices on the wage setting protocol are restrictive. However, as we demonstrate later, they enable us to use the non-parametric identification strategy in Hagedorn, Law, and Manovskii (2014) and are flexible enough to deliver a good fit to the data.

These modeling choices also imply certain assumptions on wage dynamics. First, wages are constant over a job spell. This happens because firms poach only when they know they will succeed. Empirically, we attribute within-job spell wage growth to experience accumulation. Second, workers move job-to-job to firms with whom they generate higher *surplus*. These firms *may* be less productive (lower y). Third, wages may decline upon a job-to-job transition like in Cahuc, Postel-Vinay, and Robin (2006), especially after the first job-to-job transition in an employment spell. This may happen because a worker accepts a lower wage in anticipation of the potential surplus (and wage) gain from future successful job-to-job moves. Fourth, the take-it-or-leave-it offer by the unemployed worker does not mean that wages equal the entire match output. Firms profit from poached workers and hence, unemployed workers who make the take-it-or-leave-it offer must compensate the firm for the option value of poaching. Fifth, the take-it-or-leave-it offer implies that for any (x, y) match, wages out-of-unemployment are higher than wages which arise from a job-to-job move (for the same worker type x). This happens because the continuation value of the match is identical regardless of S^o. Hence, the surplus premium that a worker who moved out-of-unemployment commands over a worker who moved job-to-job, must be reflected in wages.[9] We confront these implications with the data in Section 4.

Matching takes place when both the worker and the firm find it mutually acceptable. To formalize this, we describe the set of firms (workers) that workers (firms) are willing to match with. This set depends on whether the worker is moving out-of-unemployment or job-to-job. $B^w(x)$ is the set of firms

[9] The surplus premium can be seen from $S^o(x, y^{current}, U) = S(x, y) > S^o(x, y^{current}, y^{previous})$.

that a worker of type x moving out-of-unemployment is willing to match with:

$$B^w(x) = \{y : S(x,y) \geq 0\}.$$

Likewise, $B^f(y)$ is the set of workers moving out-of-unemployment that firm y is willing to match with:

$$B^f(y) = \{x : S(x,y) \geq 0\}.$$

$B^e(x,y)$ is the set of firms whom worker x employed at y is willing to move to via a job-to-job transition:

$$B^e(x,y) = \{\tilde{y} : S(x,\tilde{y}) \geq S(x,y)\}.$$

$B^p(y)$ refers to the set of matches where firm y can successfully poach a worker from:

$$B^p(y) = \{(\tilde{x},\tilde{y}) : S(\tilde{x},y) \geq S(\tilde{x},\tilde{y})\}.$$

A match (x,y) forms between a vacancy and an unemployed worker when $y \in B^w(x)$ and $x \in B^f(y)$. A worker in match $(x,\tilde{y})$ moves job-to-job and forms a new match with y when $(x,\tilde{y}) \in B^p(y)$ and $y \in B^e(x,\tilde{y})$. We denote the complement of a set X by $\overline{X}$.

Thus, the worker's value of unemployment reflects the surplus that the worker claims from the take-it-or-leave-it offers:

$$V_u(x) \quad = \quad \beta V_u(x) + \underbrace{\beta(1-\delta)\mathbb{M}_u \int\limits_{B^w(x)} \frac{d_v(\tilde{y})}{V} S(x,\tilde{y}) d\tilde{y}}_{\text{expected surplus from successful matching}}. \tag{4}$$

The firm's value of vacancy reflects the expected profits from poaching only.

Firms extract no surplus from hiring unemployed workers:

$$V_v(y) = -c + \beta V_v(y) + \underbrace{\beta(1-\delta)\mathbb{M}_v\mathbb{C}_e \int\limits_{B^p(y)} \frac{d_m(\tilde{x},\tilde{y})}{E}(S(\tilde{x},y) - S(\tilde{x},\tilde{y}))\, \mathrm{d}\tilde{x}\mathrm{d}\tilde{y}}_{\text{expected profits from poaching}}. \quad (5)$$

However, the maintenance cost to unfilled vacancies (c) provides the incentive to firms to hire unemployed workers. Free entry implies

$$c^e = \int_0^1 V_v(\tilde{y})d\tilde{y},$$

because firms enter and exit until the expected value of a vacancy equals the entry cost of posting a vacancy. Employed workers extract S^o from their *current* match if the current match is maintained and stand to extract the *current* match surplus in the event of a successful job-to-job move:

$$\begin{aligned} V_e(x,y,S^o) &= w(x,y,S^o) + \beta V_u(x) \\ &+ \underbrace{\beta(1-\delta)\left[1 - \mathbb{M}_e + \mathbb{M}_e \int\limits_{\overline{B^e(x,y)}} \frac{d_v(\tilde{y})}{V}\, \mathrm{d}\tilde{y}\right] S^o}_{\text{retains } S^o \text{ when not successful at on-the-job search}} \\ &+ \underbrace{\beta(1-\delta)\left[\mathbb{M}_e \int\limits_{B^e(x,y)} \frac{d_v(\tilde{y})}{V} d\tilde{y}\right] S(x,y)}_{\text{captures } S(x,y) \text{ when successful at on-the-job search}}. \end{aligned} \quad (6)$$

A special case of this will be when workers move out-of-unemployment. Here,

$S^o = S(x,y)$ and the value of employment is

$$\begin{aligned}
V_e(x,y,S(x,y)) &= w(x,y,S(x,y)) + \beta V_u(x) \\
&+ \beta(1-\delta)\left[1 - \mathbb{M}_e + \mathbb{M}_e \int\limits_{\overline{B^e(x,y)}} \frac{d_v(\tilde{y})}{V}\,\mathrm{d}\tilde{y}\right] S(x,y) \\
&+ \beta(1-\delta)\left[\mathbb{M}_e \int\limits_{B^e(x,y)} \frac{d_v(\tilde{y})}{V} d\tilde{y}\right] S(x,y) \\
&= w(x,y,S(x,y)) + \beta V_u(x) + \beta(1-\delta) S(x,y) \\
&= V_u(x) + S(x,y)
\end{aligned}$$

where the last equality is from Equation (1). Rearranging yields

$$\begin{aligned}
w(x,y,S(x,y)) &= S(x,y) + (1-\beta)V_u(x) - \beta(1-\delta)S(x,y) \\
&= (1-\beta(1-\delta))S(x,y) + (1-\beta)V_u(x). \qquad (7)
\end{aligned}$$

Finally, the value of a producing job is

$$\begin{aligned}
V_p(x,y,S^o) &= f(x,y) - w(x,y,S^o) + \beta V_v(y) \\
&+ \underbrace{\beta(1-\delta)\left[1 - \mathbb{M}_e + \mathbb{M}_e \int\limits_{\overline{B^e(x,y)}} \frac{d_v(\tilde{y})}{V}\,\mathrm{d}\tilde{y}\right] (S(x,y) - S^o)}_{\text{retains profits from workers who did not move job-to-job}} . \qquad (8)
\end{aligned}$$

In a *steady state search equilibrium* (SE), all workers and firms maximize expected payoff, taking the strategies of all other agents as given. A SE is then characterized by the density $d_u(x)$ of unemployed workers, the density $d_v(y)$ of vacant firms, the density of formed matches $d_m(x,y)$ and wages $w(x,y,S^o)$. The density $d_m(x,y)$ implicitly defines the matching sets as it is zero if no match is formed and is strictly positive if a match is consummated. Wages are set as described above and match formation is optimal given wages w, i.e. a match is formed whenever the surplus (weakly) increases. The densities $d_u(x)$

and $d_v(x)$ ensure that, for all worker-firm type combinations in the matching set, the numbers of destroyed matches (into unemployment and to other jobs) and created matches (hires from unemployment and from other jobs) are the same.

2.2 Nonparametric Identification and Estimation

A constructive nonparametric identification proof is provided in Hagedorn, Law, and Manovskii (2014). Here, we briefly describe their strategy.

Plugging Equations (4), (5), (6), and (8) into (3) and using (7), we obtain that wages out-of-unemployment can be written as

$$w(x, y, S(x, y)) \quad = \quad f(x, y) - (1 - \beta)V_v(y). \tag{9}$$

From this equation, three key identification and implementation steps follow.

2.2.1 Ranking and Binning Workers

Recall that x is the productivity rank of workers. Since $f_x > 0$, we immediately see that out-of-unemployment wages within firms rank workers. We use the rank aggregation procedure described in Hagedorn, Law, and Manovskii (2014) to obtain a global ranking of workers initialized with lifetime expected wages of workers.[10] The rank aggregation algorithm combines the partial ranking of out-of-unemployment wages. For instance, at Firm 1, wages out-of-unemployment reveal that workers are ranked $a > b$ and wages at Firm 2 reveal that $b > c$. Worker b, by being ranked at two separate firms, reveals that $a > b > c$. Repeating this aggregation of rankings across more firms yields a global ranking of workers. Of course, rankings in the data may be inconsistent due to stochastic processes such as measurement error. The full

[10] Assuming that within firm wages are indeed increasing in true worker rank x, we prove in Appendix A that worker and firm fixed effects in the two-way fixed effects linear regression identifies these ranks of workers and firms only when the underlying match density is uniform. However, the identification of ranks is not guaranteed in presence of sorting that leads to a nonuniform match density.

procedure as described in Hagedorn, Law, and Manovskii (2014) maximizes the likelihood of the correct global ranking.

Once workers are ranked, they are binned. Workers are ranked from lowest to highest rank and partitioned (binned) to form bins. For example, the bottom 5% workers are in the lowest bin. Given the large number of workers available data we use, closely ranked workers that are very similar are put in the same bin. We then use wage observations for all workers in a bin as if they were a single worker's observations and compute the relevant statistics accordingly. For example, out-of-unemployment wages that workers in bin x at some firm j will simply be $w(x, j, S(x, j))$. Binning is advantageous because it averages out stochastic processes like measurement error. Binning also provides a good estimate of wages (and output) of matches between workers and firms that are not observed in the data. All this information can be inferred from wages of similarly ranked workers within the same bin.

2.2.2 Ranking and Binning Firms

Having ranked and binned workers, we first observe that by ranking and binning firms in a similar fashion, we will be able to nonparametrically estimate out-of-unemployment wages between workers in x and firms in y, $w(x, y, S(x, y))$. This can simply be done by averaging wages between workers in bin x and similarly job-to-job wage can be obtained. Furthermore, with $V_v(y)$ known, we can simply invert the wage equation and obtain $f(x, y)$.

Hagedorn, Law, and Manovskii (2014) show that the value of vacancy is monotone in y. Further, they show that the value of vacancy can be computed from wage and transition data alone.

To see this, rearrange Equation (7) and replace the surplus, $S(\cdot,\cdot)$, terms in the value of vacancy

$$V_v(y) = \frac{-c}{1-\beta} + \frac{\beta(1-\delta)\mathbb{M}_v\mathbb{C}_e}{1-\beta} \int\limits_{B^p(y)} \frac{d_m(\tilde{x},\tilde{y})}{E}(S(\tilde{x},y) - S(\tilde{x},\tilde{y}))\, \mathrm{d}\tilde{x}\mathrm{d}\tilde{y}$$

to obtain

$$
\begin{aligned}
V_v(y) &= \frac{-c}{1-\beta} \\
&+ \frac{\beta(1-\delta)\mathbb{M}_v\mathbb{C}_e}{(1-\beta)(1-\beta(1-\delta))} \times \\
&\underbrace{\int_{B^p(y)} \frac{d_m(\tilde{x},\tilde{y})}{E}(w(\tilde{x},y,S(\tilde{x},y)-w(\tilde{x},\tilde{y},S(\tilde{x},\tilde{y}))\,\mathrm{d}\tilde{x}\mathrm{d}\tilde{y}\,.}_{\text{out-of-unemployment wage premium}}
\end{aligned}
$$

This implies that firms can be ranked according to the out-of-unemployment wage premium they pay to workers that they expect to poach. This is not straightforward to compute in practice. Consider the naive approach of computing this statistic from the wages of workers that are actually poached by some firm j from other firms. The statistic requires the out-of-unemployment wages of these poached workers at j and their previous firms. However, out-of-unemployment wages at the previous firm may not be observed in the data. To overcome this problem, we utilize the fact that we have out-of-unemployment wages, $w(x, j, S(x, j))$, after ranking workers, which we do prior to computing V_v. This provides an estimate of the needed out-of-unemployment wages.[11]

Still, we may not accurately observe the distribution of workers moving into a given firm due to short samples. To solve this problem, we utilize the fact that for a given worker type, wages out-of-unemployment are greater when surplus is greater. This is immediate from Equation (7). Hence, we infer which matches a firm *would have* poached from by comparing $w(x, j, S(x, j))$ across firms, i.e. firm j will poach workers in bin x from other firms if $w^u(x, j, S(x, j)) > w^u(x, j', S(x, j'))$ where $j' \neq j$. Summing the wage premium weighted by the observed match density gives the expected out-of-unemployment wage premium which ranks firms. Once firms are ranked, they

[11] If no measure of $w(x, j, S(x, j))$ is available, then we use the average wages of all workers in bin x, $\mathbb{E}_S w(x, j, S)$.

can be binned in the same way workers were binned.

2.2.3 Recovering the Production Function and Search Parameters

To compute the value of vacancy of individual firms, $V_v(j)$ we now need to estimate $\mathbb{M}_v$, $\mathbb{C}_e$ and δ. The probability that firm j fills the vacancy conditional on meeting an unemployed worker ($\tilde{q}_j^u$) is the share of unemployed workers that j is willing to hire.[12] Denoting the number of observed new hires out-of-unemployment in firm j by $H^u(j)$ and the number of unobserved vacancies posted by $v(j)$, we have $H^u(j) = (1-\delta)\mathbb{M}_v\mathbb{C}_u\tilde{q}_j^u v(j)$. In other words, the observed number of new hires equals the probability the match is formed times the number of vacancies. Aggregating over firms, we can solve for $\mathbb{M}_v\mathbb{C}_u$ since total vacancies V equal U in the steady state, which overcomes the need to observe vacancies at the firm level. $\mathbb{M}_v\mathbb{C}_e$ can be estimated in a similar way. Next, we estimate on-the-job search intensity using the fact that $\phi = \frac{U}{E} \cdot \frac{\mathbb{C}_e}{\mathbb{C}_u}$. Finally, with ϕ we compute $\mathbb{C}_e$ or $\mathbb{C}_u$ (since unemployment U is known) and then back-solve for $\mathbb{M}_v$. The average length of employment spells identify δ. To do this, we use employment spells that are observed without truncation due to the sample period.

Recovering the production function is straightforward after workers and firms are ranked and binned. We first compute $w(x, y, S(x, y))$. Averaging $V_v(j)$ yields $V_v(y)$. Finally, we solve for $f(x, y)$ using Equation (9).

3 Data

We use the Linked Employer-Employee (LIAB) M3 panel covering 1993-2007 provided by the German Institute for Labor Research (IAB) to estimate the model. This panel includes about 1.8 million unique individuals and over 500,000 establishments out of which over 2,300 establishments are surveyed between 1996 and 2005. The IAB builds the LIAB survey panel through strati-

[12] This can be measured from the unemployment rates of worker types that firm j hires. See Hagedorn, Law, and Manovskii (2014) for details.

fying over industries, so the establishments represent the cross-section of industries in Germany. Large establishments are oversampled.[13] The work history of workers includes records from the Employment History (Beschäftigten-Historik - BeH) and records from the Benefit Recipient History (Leistungsempfänger-Historik - LeH). BeH records cannot be longer than a year since annual notification is required for all jobs in progress on December 31, but LeH records can span multiple years. We observe the complete work history between 1993 and 2007 of every worker recorded to have worked at any one of the surveyed establishments for at least a day between January 1st 1993 and December 31st 2007. While the work history we observe also includes employment spells at establishments outside the surveyed panel, we observe the complete workforces (that an establishment reports) at surveyed establishments only. Wage records are based on notifications submitted by employers to various Social Security agencies upon a change in the conditions of employment. Hence, this panel excludes individuals not subject to Social Security contributions, e.g. civil servants and full-time students.

The panel consists of continuous job spells and unemployment records. Start and ends of spells are reported at a daily frequency and the IAB splits unemployment spells spanning multiple years so that all spells fall within a year. We impute missing education values using the IP1 procedure described in Fitzenberger, Osikominu, and Völter (2006).[14] An important limitation of the data is the censorship of about 9% of the earnings at the Social Security maximum. Our structural analysis does not suffer much from this limitation, because the estimation procedure we use relies mainly on out-of-unemployment wages, of which only 2% are censored. We impute censored wages following closely the imputation procedure in Card, Heining, and Kline (2013).[15]

We consider full-time employed men aged 20-60 employed by West Ger-

[13] We show the establishment size distribution Appendix B and show in Appendix B.1 that this dataset reflects aggregate wage trends reported in the literature.

[14] Details in Appendix B.

[15] Details are provided in Appendix B. Our imputation procedure adapts the procedure in Card, Heining, and Kline (2013) to the limitations of our sample.

man establishments. Mini-jobs which appear past 1999 are dropped. We only consider workers with more than one job spell and less than 150 job spells. We also only consider jobs with a real daily wage above 10 Euros with 1995 as the base year. We drop all apprentice and self-employed workers as well. We define out-of-unemployment spells in our sample as individuals (1) whose first observed job is prior to age 26, (2) whose start of a new job is preceded by compensated unemployment in the past 28 days, or (3) who have an uncompensated gap between two jobs longer than one month.

Next, we split the sample into data from 1993 to 2000 (1990s) and 2001 to 2007 (2000s), and estimate the model separately on each subsample. Our sample contains 383,772 establishments, 889,307 workers, and 6,254,287 job spells for the 1990s; and 321,756 establishments, 818,967 workers, and 5,269,024 job spells for the 2000s. We aggregate it to a monthly frequency to estimate our model. We aggregate to a yearly frequency to perform our test of additive separability and estimate worker and firm effects to be consistent with Card, Heining, and Kline (2013). In the case of several concurrent jobs in a given month (year), we define the main job to be the job in which the worker earns the most in that month (year).

The worker ranking procedure we use relies on workers moving between establishments. Thus, we restrict the ranking of workers as well as regressions to remove the effects of observable characteristics to the largest connected set (see Abowd, Creecy, and Kramarz (2002)) containing 359,643 establishments, 871,533 workers and 6,176,894 job spells for the 1990s and 272,632 establishments, 780,347 workers, and 5,070,658 spells for the 2000s. We rank workers using data from the full sample.

We treat establishments in the data as firms as described in the model. We use the terms interchangeably for the rest of the paper.

For our test of additive separability, we restrict our sample to LIAB-surveyed establishments that employ at least 2 workers, because we cannot fully observe coworkers relationships at non-surveyed firms. This sample consists of 1,225,892 unique coworkers pairs observed at 2 different establishments, 11,120 workers, and 793 establishments. For estimation of the production function and report of fit, we restrict our sample to LIAB-surveyed establishments that employ at least 10 workers. We only rank establishments and estimate the production function on this sample, because ranking establishments requires observing their entire workforce history. This sample consists of 1,658 (1,512) establishments, and 720,762 (535,091) workers for the 1990s (2000s). We have a total of 3,442,577 job spells for the 1990s, and 2,501,472 job spells for the 2000s with which we estimate the model. The dropping algorithm we use for misranked workers (described in Hagedorn, Law, and Manovskii (2014)) drops 7,869 workers for the 1990s and 8,803 workers for the 2000s.

4 Estimating the Model

As described earlier, we estimate the model with wages net of the effects of observables on each subperiod (1990s and 2000s) of the data. The model is estimated on residual wages. To construct residual wages, we follow Card, Heining, and Kline (2013) in including an unrestricted set of year dummies as well as quadratic and cubic terms in age fully interacted with educational attainment in our set of time-varying observable characteristics. In particular, we regress individual log real daily wage $\log w_{it}$ of individual i in month t on a worker fixed effect α_i and an index of time-varying observable characteristics $z_{it}^{'}$

$$\log w_{it} = z_{it}^{'}\gamma + \alpha_i + r_{it},$$

where r_{it} is an error component. The residual wage which serves as input into the analysis is then defined as $w_{it} = \exp(\log w_{it} - z_{it}'\hat{\gamma})$. Card, Heining, and Kline (2013) (CHK) also include establishment fixed effects in the regression. This difference is inconsequential for our purposes, as the inclusion of establishment fixed effects has virtually no impact on $\hat{\gamma}$. In particular, over the combined 1993-2007 sample, $corr(z_{it}^{'}\hat{\gamma},\, z_{it}^{'}\hat{\gamma}_{CHK}) = 0.9952$ and $corr(\log w_{it},\, \log w_{it,CHK}) = 0.9995$, where $w_{it,CHK} = \exp(\log w_{it} - z_{it}^{'}\hat{\gamma}_{CHK})$.

Table 1 shows that residual wages, $\log w_{it} - z_{it}'\hat{\gamma}$, capture a large portion of cross-sectional variance in log wages in the data given our set of observables.

Table 1: Covariance matrix of log wages in the 1990s and 2000s
1993-2000 (1990s)

	$\log w$	$\hat{\alpha}$	$z'\hat{\gamma}$	$\hat{r}$
$\log w$	0.1811	0.1538	0.0097	0.0176
$\hat{\alpha}$		0.1516	0.0021	0.0000
$z'\hat{\gamma}$			0.0076	0.0000
$\hat{r}$				0.0176

$\text{var}(\hat{\alpha} + r) = 0.1516 + 0.0176 = 0.1692$
represents 93% of wage variance

2001-2007 (2000s)

	$\log w$	$\hat{\alpha}$	$z'\hat{\gamma}$	$\hat{r}$
$\log w$	0.2295	0.1906	0.0182	0.0207
$\hat{\alpha}$		0.1843	0.0063	0.0000
$z'\hat{\gamma}$			0.0120	0.0000
$\hat{r}$				0.0207

$\text{var}(\hat{\alpha} + r) = 0.1843 + 0.0207 = 0.2050$
represents 89% of wage variance

The overall change in wage variance is $0.2295 - 0.1811 = 0.0484$. The change in residual wage variance, $\text{var}(\hat{\alpha} + r)$, is $0.2050 - 0.1692 = 0.0358$. The model includes wage variation which accounts for $0.0357/0.0484 = 74\%$ of the increase in wage inequality.

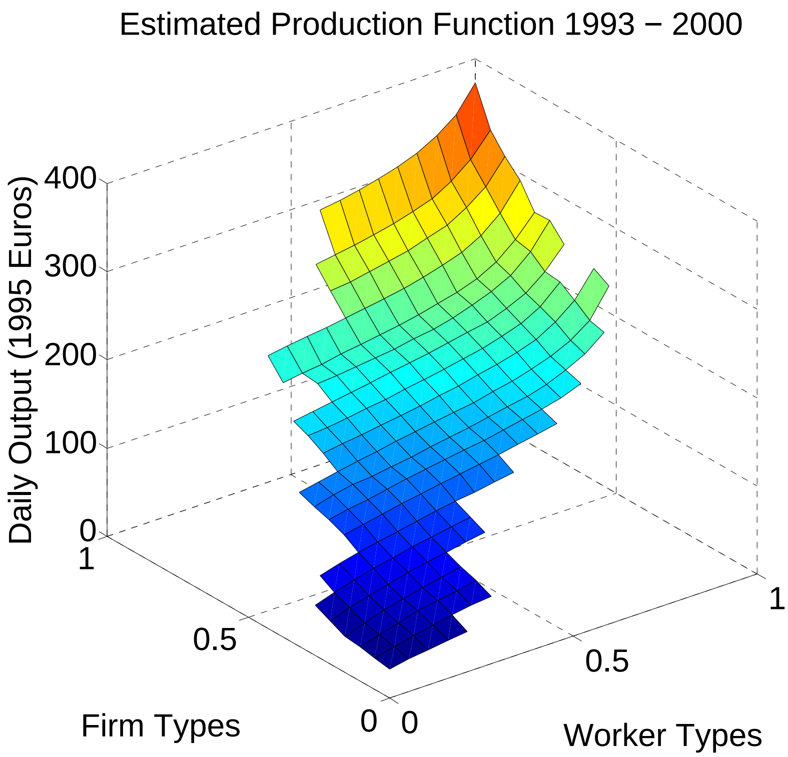
Having ranked workers and establishments, we bin workers into 20 bins with an equal number of unique individuals in each bin. Establishment bins are selected so that each bin contains approximately the same number of unique jobs. It is not possible to have exactly the same number of jobs in each bin because establishments in the data differ greatly in size.[16]

[16] We overcome this problem by maximizing over small (adjacent) perturbations to the

The inputs to the model (the production function, $f(x, y)$ and search parameters $\mathbb{M}_v$ (search intensity), ϕ (on-the-job search intensity), and δ (match destruction probability)) are estimated following the steps described in Section 2. We fix the gross interest rate at 1.04 to pin down the discount factor β and estimate production function up to an additive constant. The last step of our estimation involves estimating the additive constant to the production function to minimization the squared deviations of mean log wages and the variance of log wages.

We estimate this additive constant by simulating wages using the production function (Figure 1) and search parameters (Table 2) estimated from the data. Irregularities in the matching set arise due to firm size heterogeneity. For instance, some firm bins contain less than five firms with one of the firms being very large. The large firm influences the matching set greatly and this results in roughness of the matching set on the edges. To overcome these irregularities, we calibrate the matching set by perturbing the matching set obtained directly from the data by 1 bin from its edge. In practice, perturbation amounts to including and excluding worker types on the edge of the matching set. The perturbation which fits the data best is used. The fit of the model to the data is evaluated using the resulting wage and density functions. Note that the fit of the model generated wages and match *density* to the data does not arise by construction. The mobility of workers in the simulation arises endogenously in the model from the production function and search parameters. These primitives do not guarantee generating wages or mobility identical to what is observed in the data. We compare the resulting wage functions and the match densities later in this section.

estimated matching set while preserving positive assortative matching (PAM) where the upper and lower bounds of the matching sets are increasing in worker/firm type.



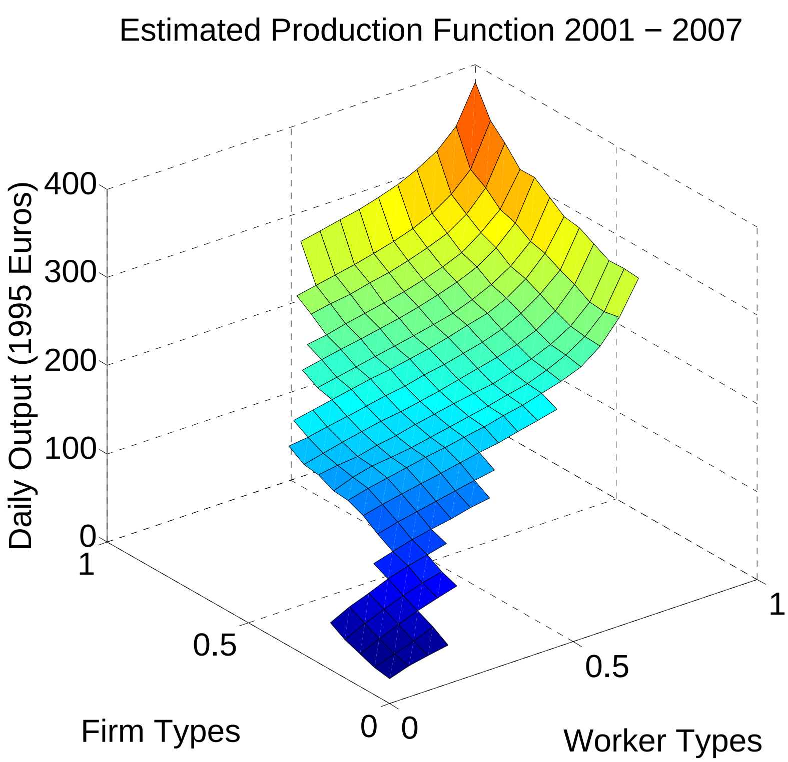


Figure 1: Estimated production functions.

Table 2: Parameters

	1990s		2000s	
	Data	Model	Data	Model
Externally Selected Parameters				
Annual Gross Interest Rate	1.04		1.04	
Estimated Parameters				
Monthly Meeting Probability, $\mathbb{M}_v$	0.34		0.41	
On-the-job Search Intensity, ϕ	0.42		0.23	
Monthly Job Separation Probability, δ	0.012		0.0098	
Calibrated Parameter				
Additive Constant to $f(x, y)$		1.254		5.799
Target Quantities				
Mean Log Wage	4.40	4.38	4.50	4.46
Variance Log Wages	0.169	0.169	0.205	0.208

4.1 Model Fit

We simulate the model using the model primitives, $f(x, y)$ and search parameters, obtained in the previous subsection. We simulate the model for the same number of years as in the subsample periods. The model is simulated at a weekly frequency and model-generated data is aggregated the same way as done in the real data. Table 3 summarizes the fit of the model. In both periods, the model replicates the job-to-job transition rate and the employment rate. In a steady state, aggregate employment and the separation rate (δ) define the job finding rate, so the model replicates overall job mobility rates. The model also generates comparable quantities of sorting (as measured by a rank correlation of types) between workers and establishments.[17] We can see from Table 3 that highly ranked workers tend to sort with highly ranked firms and that this correlation has increased from 0.7621 to 0.7919 in the data.

[17]Sorting is measured on surveyed establishments only as it requires ranking firms.

Table 3: Model Fit

	1990s		2000s	
	Data	Model	Data	Model
Fit to Mobility and Sorting				
Probability of Monthly Job-to-Job Move	0.0118	0.0116	0.0107	0.0091
Employment Rate	0.8916	0.8801	0.9002	0.9223
Correlation of Worker and Firm Type	0.7621	0.7129	0.7919	0.7651
$\mathbf{corr}(\mathbf{w}^{\mathbf{model}}, \mathbf{w}^{\mathbf{data}})$				
Overall	0.9996		0.9983	
Below Median	0.9974		0.9991	
Above Median	0.9995		0.9974	
Explanatory Power				
R^2 using $w^{model}(x, y)$	0.919		0.918	
R^2 using Worker and Firm Fixed Effects	0.942		0.941	

Next, we correlate the non-parametrically estimated wage function from the data, w^{data}, with the wage function generated by the model, w^{model}. The wage function refers to wages averaged across all workers and establishments in match (x, y) in the data and in the model simulated data. The model fit to the wage function in the data does not arise by construction. We only target the overall mean and variance of log wages in our calibration, and use the production function estimated using wages out-of-unemployment. We report the overall correlation of the wage function from the data and the model simulation. We also report this same correlation restricted to the lower and upper half of wages. We see from these correlations that the estimated production function along with the estimated search parameters replicate the nonparametric wage function in the data.

To assess the overall fit to raw wages, we predict wages and compute the R^2 arising from our prediction. We only assess the fit on wages earned at surveyed establishments, because we can only rank surveyed firms. Every worker i and establishment j in the economy has an estimated type given by $\hat{x}(i)$ and $\hat{y}(j)$, respectively. Our prediction of wages is $\log \hat{w}_{it} = z_{it}\hat{\gamma} +$

$\log w^{model}(\hat{x}(i), \hat{y}(j))$. w^{model} is the equilibrium wage function simulated from the model. For comparison, we display the R^2 of the regression including a fixed effect for every worker and every establishment. This regression is run on wage data including surveyed and non-surveyed establishments using an identical set of observable characteristics. The R^2 is displayed for wages paid by surveyed establishments only.

Regarding wage dynamics, the average residual wage of a worker declines about half of the time when the worker moves between jobs in the data. In model-generated data, wages decline 46% of the time. Note that the model implies that wages out-of-unemployment are greater than wages arising from a job-to-job move for a given (x, y) pair. Average *residual* wages from job-to-job moves exceed *residual* wages out-of-unemployment by only a tenth of a standard deviation. Hence, the data does not outright reject this implication of the model.

Finally, we provide a visual comparison of wages and match densities from the simulation described above. Figure 2 compares wages obtained in the data to wages that are simulated using the production function and search parameters estimated from the data. The wage function simulated from the model is almost uniformly above or below the wage function from the data. This is consistent with the average values reported in Table 2, because the wage function is weighted by the match densities to calculate its average. Figure 3 provide two views of the match densities arising from the same exercise. Overall, we find that our nonparametrically estimated model fits the data along key dimensions with three search parameters and a non-parametrically estimated production function over 20 worker types and 20 establishment types. The model's fit to wages is comparable to wage regressions which assign a fixed effect to every worker and to every establishment. This is true for the hundreds of thousands of workers and over a thousand establishments in this analysis. In addition to wages, the model replicates mobility and sorting over productivity in the data. Given this fit, we can confidently exploit the structure of the model to understand the rise in German wage inequality in Section 5.

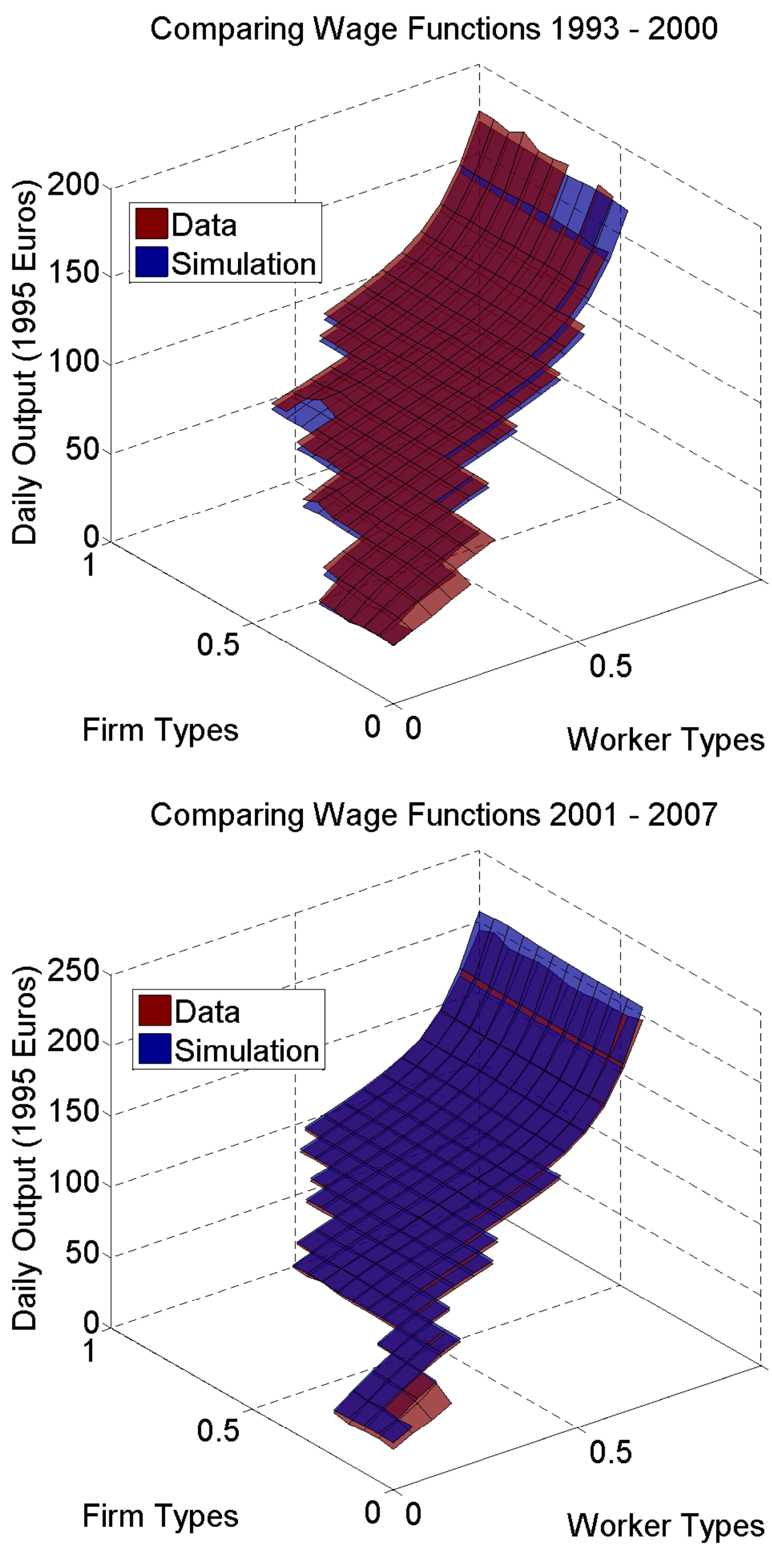


Figure 2: Wage functions from the data and simulated from the model using production function and search parameters from the data.

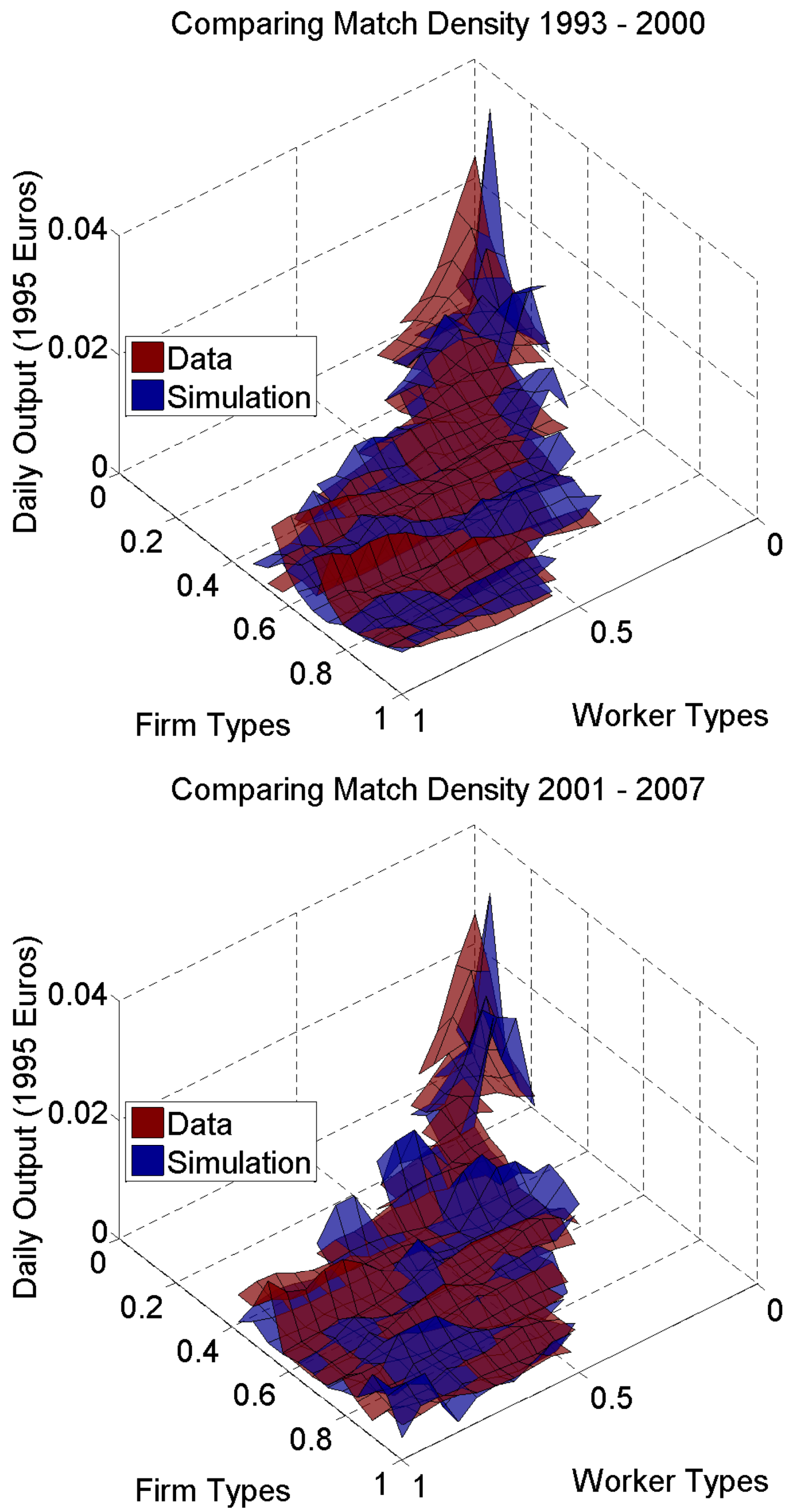


Figure 3: Match densities from the data and simulated from the model using production function and search parameters from the data.

4.2 Log-Linear Variance Decompositions

We also consider the model's ability to generate the decomposition of wages as described in Abowd, Kramarz, and Margolis (1999). To do so, we take the wage function estimated from the data and the wage function estimated from model simulated data and assess how well we reproduce the moments from a worker-firm fixed effects variance decomposition.[18] We also examine the extent to which aggregation and mobility bias affect replicating the decomposition.

In their study of West German inequality, Card, Heining, and Kline (2013) decompose log wages into variance contributions due to observables and worker and firm fixed effects on a superset of our data.[19] They find that firm fixed effects account for around 20% of wage variance over 1996 to 2001 and 2002 to 2009. However, variance contributions emerging from this method have been shown to be biased due to noisy estimates of fixed effects for small firms where few workers move (Andrews, Gill, Schank, and Upward, 2012, 2008). We follow Card, Heining, and Kline (2013) and aggregate the data to the annual level to estimate regressions with worker and firm fixed effects.

Estimating this regression on raw log wages in the data, we find evidence that much of the firm variance contribution in our sample comes from small firms. We only observe a few workers at these firms moving, because these firm are relatively small (e.g. less than 20 workers). Restricting to smaller firms, the firm contribution rises to 58% whereas it is only 41% for firms for larger firms.[20] The covariance between worker and firm fixed effects for these smaller firms drop to −11% compared −4% for larger firms. This feature of the data is consistent with Andrews, Gill, Schank, and Upward (2012) who argue that mobility bias causes an upward bias in the variance contributions

[18] The weekly simulation contains 24,000 workers and 240 firms over 8 years. The production function and search parameters are the only inputs into the model.

[19] Card, Heining, and Kline (2013) have access to the entire universe of firms and workers whereas our dataset only contains a subset of firms. Hence, their data features substantially more mobility.

[20] Firms are smaller in the sense that we observe fewer than 3000 worker-firm matches over 1993-2007.

and downward bias in the covariance contribution of worker and firm fixed effects.

To understand how our model performs, in Table 4, we present worker-firm fixed effect variance decompositions for (1) wages in the data, (2) fitted wages from binning worker-firm fixed effects, (3) fitted wages from our estimated wage function in the data, and (4) fitted wages from a wage function from model simulated data. For (2), we take fitted residual wages via worker and firm fixed effects and average them in bins defined by ordering fixed effects to see how aggregation affects variance contributions. The resulting new fitted wage averages out noise in firm effects due to limited mobility associated with small firms. With such aggregation, we see that the variance contribution of firms goes down to 21% from 54%. Thus, we find that aggregating removes a substantial amount of estimated firm heterogeneity, because much of the firm contribution is due to noisy estimates of fixed effects from smaller firms. Similarly, we construct a fitted wage for every observation in the data using the wage function we estimate in the data and a wage function we simulate from the model. Our fitted wages estimated in the data and simulated from the model, $z_{it}'\hat{\gamma} + \log w(\hat{x}(i), \hat{y}(i))$, match the firm contribution that emerges from (2). Hence, our fitted wages reproduce the firm contribution to wage variance in the data once we take into account the effect of aggregation on the firm variance contribution. The correlation between worker and firm productivity types reported in Table 3 is much higher than the correlation of estimated worker and firms fixed effects.[21]

Performing the AKM decomposition on model generated data and model generated mobility of workers to firms results in the estimated worker type share of wage variance to explain almost all (>94%) the variance in wages.[22]

[21] See Abowd and Kramarz (1999) and Eeckhout and Kircher (2011) for discussions on the relationship between the correlation of worker and firm fixed effects and sorting over fundamental quantities such as productivity.

[22] We concatenate model simulations from 1993-2000 and 2001 - 2007. Hence, the variance in the model lies in between the variance of wages in both halves of the data. This variance does not match exactly the variance of the full sample in the data due to factors (such as panel balance) that we do not account for. This does not affect the well known observation

Firm fixed effects explain a negligible quantity of wage variance. Our results here suggests that this is largely due to firm sizes, aggregation and heterogeneous mobility of workers across firms in the data. With those elements equalized between the model and the data (Table 4 and 5, lines 2 and 4), we find that the model in fact replicates the firm contribution.

Table 4: Variance Decomposition on Fitted Wages (1993-2007)

		Var($\log w$)	Var(α_i)	Var(ψ_j)	Cov(α_i, ψ_j)	Var($z'\hat{\gamma}$)
1.	$\log w_{ijt}$	0.192	0.090	0.103	-0.013	0.004
2.	$z_{it}'\hat{\gamma} + w^{akm}(x(i), y(j))$	0.133	0.090	0.028	-0.001	0.012
3.	$z_{it}'\hat{\gamma} + w^{data}(x(i), y(j))$	0.165	0.076	0.037	0.019	0.013
4.	$z_{it}'\hat{\gamma} + w^{model}(x(i), y(j))$	0.149	0.072	0.030	0.018	0.005
5.	$w^{model}(x, y)$	0.189	0.178	0.0001	0.002	N/A

Table 5: Variance Contribution on Fitted Wages (1993-2007)

		Var(α_i)	Var(ψ_j)	Cov(α_i, ψ_j)	Var($z'\hat{\gamma}$)
1.	$\log w_{ijt}$	47%	54%	-14%	2%
2.	$z_{it}'\hat{\gamma} + w_{akm}(x(i), y(j))$	68%	21%	-2%	9%
3.	$z_{it}'\hat{\gamma} + w_{data}(\hat{x}(i), \hat{y}(j))$	46%	22%	22%	8%
4.	$z_{it}'\hat{\gamma} + w_{model}(\hat{x}(i), \hat{y}(j))$	49%	20%	22%	4%
5.	$w^{model}(x, y)$	94%	0.1%	2.2%	N/A

Note: $w^{data}(i, j)$ are log wages in the data. $w^{akm}(x, y)$ are fitted log wages $z'\hat{\gamma} + \log w(x, y)$ where $w(x, y)$ is determined ranking workers according to worker and firm fixed effects instead of our estimation method. $w^{data}(x, y)$ are fitted log wages $z'\hat{\gamma} + \log w(\hat{x}, \hat{y})$ where $w(\hat{x}, \hat{y})$ is the wage function estimated in the data based on estimated worker ($\hat{x}$) and firm ($\hat{y}$) types. $w^{model}(x, y)$ are fitted log wages $z'\hat{\gamma} + \log w(x, y)$ where $w(x, y)$ is the wage function emerging from a simulation where the production function and search parameters are inputs.

---

that worker fixed effects explains most wage variance in Beckerian models.

5 Decomposing the Rise in German Wage Dispersion

We now perform decompositions to understand why wage dispersion has increased in Germany. Our first decomposition uses the model to tease apart the contributions to the increase in residual wage variance of model primitives − search and production technology. Then, we separate the direct and indirect effects of changes in the production technology. Finally, we evaluate the importance of search frictions for cross-sectional wage dispersion.

5.1 Separating the Contributions of Search Frictions and Production Technology to Rising Wage Dispersion

We first ask which model primitive affects the increase in wage variance. To answer this question, we turn to the structural model, which we demonstrated to be a good fit to the data. In this model, the search parameters are the job destruction rate (δ), the aggregate search intensity ($\mathbb{M}_v$), and the on-the-job search intensity (ϕ). We change these parameters one at a time and recompute the model while maintaining $f(x, y)$ to measure their contribution to the increase in wage inequality across the subperiods. The change in variance due to increasing all of the search parameters simultaneously is -0.0011, as shown in Table 6. This suggests that changes in search frictions do not explain the increase in wage variance we observe. In contrast, we find that maintaining the search parameters while changing the production function gives all the increase in the wage variance.[23] Hence, we conclude from here that the key primitive which affected the increase in German wage inequality is the pro-

[23] Many labor market changes occured in West Germany during the periods we consider, including the Hartz Reforms, the development of open clauses, active labor market policies, and continued reunification. These events and others affecting job mobility and wages out-of-unemployment are captured in the production function. In this sense, the production function overstates the role of technological change in affecting wage inequality. See Card, Heining, and Kline (2013) for a discussion of labor market reforms.

duction function.[24]

Table 6: Wage Variance Counterfactuals

	Wage Variance
$f(x,y)_{1990s}$ + $Search_{1990s}$	0.1672
$f(x,y)_{1990s}$ + $Search_{2000s}$	0.1661
$f(x,y)_{2000s}$ + $Search_{1990s}$	0.2085
$f(x,y)_{2000s}$ + $Search_{2000s}$	0.2054

5.2 Separating the Effects of Changes in Production Technology and Induced Sorting Patterns on the Rise in Wage Dispersion

How does the production function affect the change in wage variance? Our next decomposition is designed to separate the channels by which the production function affects wages. Recall that the value functions (Equations 4-6 and 8) contain the equilibrium bargaining sets given by $B^w(x)$, $B^f(y)$, $B^e(x,y)$ and $B^p(y)$ as well has the production function $f(x,y)$. In equilibrium, the change in $f(x,y)$ as described in the previous decomposition induces a change in the bargaining sets as well as wages. In turn, changes to the bargaining sets induce changes in the equilibrium match density.

We consider two counter-factual experiments to tease apart the effects of the change in the production function. In each counter-factual, we simulate wages arising in a partial equilibrium, meaning wages may respond to changes in the production function or bargaining sets, but not both. To isolate the direct effect of changing match outputs without altering behavior (bargaining sets), we compute the wages which arise from the partial equilibrium of the

[24]The bargaining protocol in the model takes place at the firm level which differs from the wide-spread, sectoral-level bargaining in West Germany. Jung and Schnabel (2011) show that only 19% of firms pay at the sectoral-bargained level using a 2006 survey of over 8,000 firms. Furthermore, larger firms disporportionately deviate and pay above the sectoral agreement. Thus, the oversampling of large firms in our sample likely helps the fit of the model (at least in the later subperiod) despite the differing bargaining protocols.

estimated production function from the 2000s, while maintaining the bargaining sets and search parameters from the 1990s. The wage variance for this counter-factual equilibrium is 0.1882 which means this direct effect accounts for 55% of the change in wage variance. To isolate the indirect effect of changes in sorting behavior, we compute wages which arise from the partial equilibrium with the estimated production function and search parameters from the 1990s, but with the bargaining sets from the 2000s. The wage variance in this case is 0.1863 meaning this indirect effect accounts for 50% of the change in wage variance. Notice that both effects do not add up to the increase in wage variance from changing the production function alone due to general equilibrium responses. However, we still see that the direct and indirect effects of changes in the production function account for wage variance to a similar degree.

5.3 Measuring the Contribution of Search Frictions to Wage Dispersion

Our first three decompositions suggest important roles for the production function through its direct effect on output and its indirect effect on behavior and sorting. However, it appears that search frictions themselves do not have much of a role to play in increasing wage variance. Here, we assess the role that search frictions play in the cross section. As mentioned, the model's roots in Becker (1973) suggest a very natural way of understanding the role of search frictions – remove them completely and compute frictionless wages. In Becker's environment, firms take the wage schedule as given and maximize profits $\pi(x, y) = f(x, y) - w(x)$. This yields the first order condition

$$f_x(x, y) = w_x(x).$$

This condition must hold at the equilibrium allocation $y^* = \mu(x)$ and therefore, wages can be obtained by integrating along the equilibrium path

$$\begin{aligned} w^*(x) &= \int_0^x f_x(\tilde{x}, \mu(\tilde{x}))d\tilde{x} + w_0 \\ w_0 &\in [0, f(x_{min}, \mu(\tilde{x}_{min}))] \end{aligned}$$

where w_0, the constant of the integration, is the share of the output going to the lowest type worker.

In the production functions that Becker considers, this equilibrium path is on the main diagonal ($\mu(x) = x$) in the case of positive assortative matching (PAM), and on the off diagonal ($\mu(x) = 1-x$) in the case of negative assortative matching (NAM). The identification strategy we use does not rely on the global modularity of the production function. In fact, the production function we estimate is neither sub-nor super-modular, so we compute the optimal planner's allocation from the estimated production function and numerically compute equilibrium wages. The planner's problem is to maximize output by assigning a worker type to a firm type with the constraint that each firm type can only hire one worker type. We use an implementation of the algorithm in Munkres (1957) to obtain a solution to this linear assignment problem. These optimal allocations are displayed in Figure 4.

We find that the log wage variance is 6.1% and 9.8% higher in the 1990s and the 2000s respectively when $w_0 = f(x_{min}, \mu(\tilde{x}_{min}))$ is imposed. Furthermore, the variance in logs increases as w_0 falls, hence these numbers we report constitute a lower bound on the amount by which search frictions *reduce* log wage variance. This may appear somewhat surprising given that search frictions are often thought to increase wage dispersion. In this quantitative exercise however, wage variance is lower, because search frictions prevent workers and firms from fully exploiting the complementarities in production. Hence, search friction appear to govern the level of wage dispersion but do not explain its change over time.

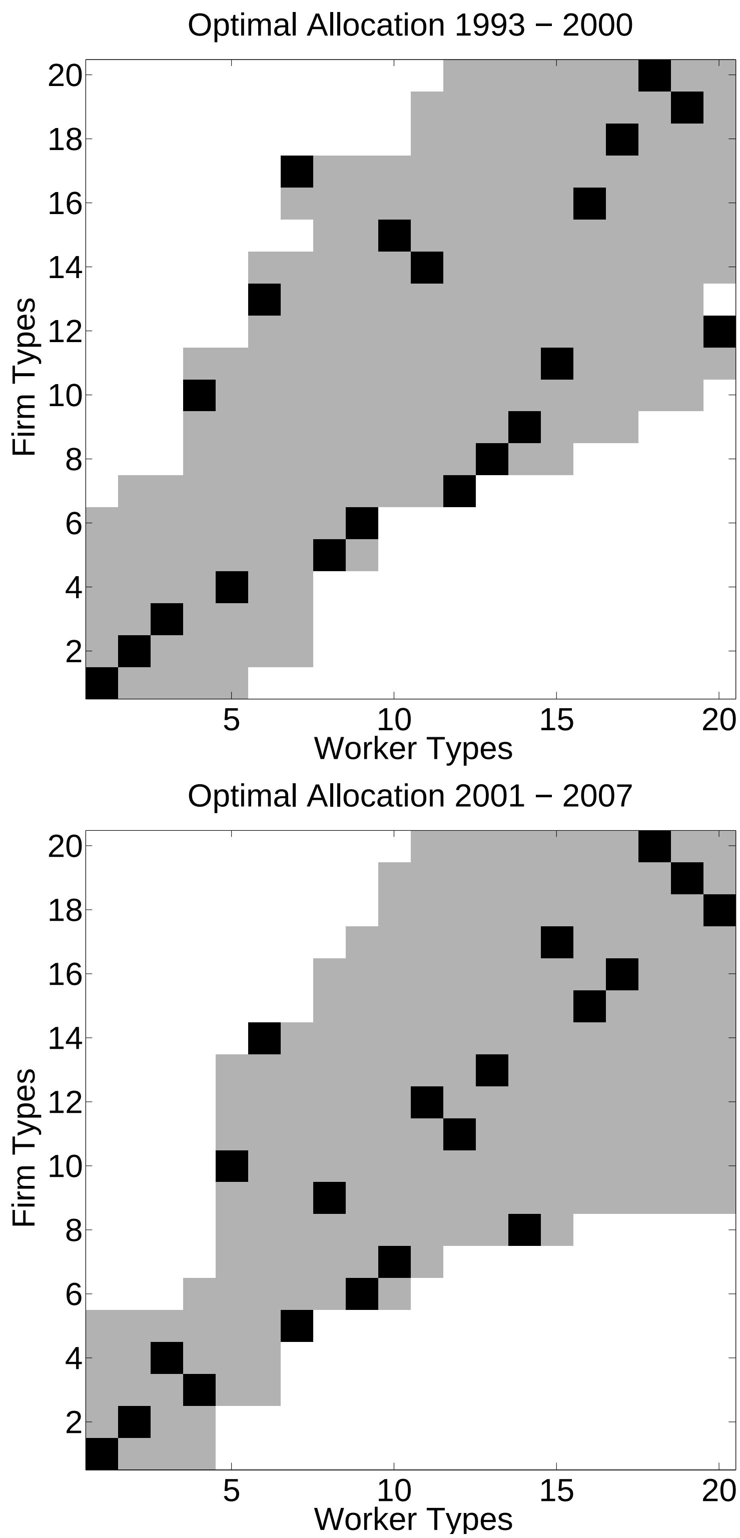


Figure 4: The grey shading is the acceptance set. The dark spots are is the output maximizing assignment of workers to firms assuming full employment.

6 Testing Additive Separability

We estimate a structural model where wages drive sorting between workers and firms and find that the model fits wages well. However, the leading non-structural wage decomposition due to Abowd, Kramarz, and Margolis (1999) (AKM) explains wage variance equally well. To fit wages, AKM specify log wages as additive in returns to observable characteristics, worker and firm fixed effects and an error term. Estimated via least squares, this specification yields R^2 statistics of around 90% across several datasets, including France, the United States, and Germany.[25] The literature has considered this high R^2 to be supportive of the additively separable specification as a good first approximation to the wage determination process.[26]

We find a high R^2 statistic insufficient to conclude that the wage structure is additively separable. AKM's log additive separability specification imposes the restriction that two workers who both move from firm j to firm j' receive the same percentage wage increase or decrease, implying that wages do not drive sorting.[27] In contrast, theoretical models of sorting, such as the one we estimate, permit different workers to experience different wage gains or losses when moving across firms, inducing sorting between workers and firms.[28] We test how well the data supports this restriction, which also indicates the importance of non-parametrically estimating wages in place of this prominent

[25]See Lopes de Melo (2013) for AKM results across various matched employer-employee datasets.

[26]For example, see Card, Heining, and Kline (2013).

[27]Residual log wages are specified as the sum of worker fixed effect (α_i), firm fixed effect (ψ_j) and error, u_{it}. Concretely, residual log wages are written as $\alpha_i + \psi_j + u_{it}$. To consistently estimate the fixed effects, workers are assumed to not make any mobility decisions based on u_{it}. Thus, the linear regression implies that sorting on residual wages are guided by solely by $(\alpha_i, \psi_j)_{j=1,\ldots,J}$. However, this specification means that workers experience identical wage gains (or losses) across firms and thus make identical accept/reject decisions based on wages alone, ruling out wages guiding sorting.

[28]Shimer (2005) provided an example of in structural model where wages decompose to separable worker and firm fixed effects, and positive assortative matching still takes place. However, sorting in this example takes place due to unemployment risk, because the highest paying jobs are hard to obtain. It is not based on comparative advantage, which is realized as nonlinearities in the wage function.

parametric specification, to capture significant deviations from log additive separability.

We take no stance on the true data generating process for residual wages when evaluating the additive separability restriction. We specify a general process (a dummy variable for each worker-firm match) for residual wages and test whether this general process satisfies the implications of constant log wage differentials for colleagues. Previous theoretical critiques argue that the fixed effects cannot be interpreted as primitives of a structural model. Gautier and Teulings (2006) argue that environments featuring comparative advantage do not have a universally "most productive" firm, so we cannot interpret fixed effects as a meaningful ordering of firms. In Eeckhout and Kircher (2011), wages are non-monotonic in firm productivity, because workers must compensate firms for the option value of forming more productive matches. In this setting, firm fixed effects cannot be interpreted structurally as a measure of productivity. Another strand of critiques takes the specification as correct but shows that the fixed effects are subject to estimation error and bias due to limited mobility. For example, Andrews, Gill, Schank, and Upward (2012) show that the firm fixed effect estimates are noisy when workers move infrequently between plants.

Card, Heining, and Kline (2013) look at average wage changes for movers going from one wage quartile to another. They find fairly symmetric wage changes, in that workers moving to a higher quartile tend to receive a wage increase similar in magnitude to the decrease that workers experience moving to a lower quartile. They also divide worker and firm fixed effects into deciles and look at the average residual within these decile cells which they find to be small. They interpret their findings, along with a high R^2, as evidence in support of the worker-firm fixed effect specification.

We do not assume the fixed effect specification to be correct but instead turn to the data for direct evidence without using a structural model for residual wages. We exploit worker mobility in the data to examine log wage differentials under the AKM specification. It restricts log wage differences between workers to be constant across firms. Rather than looking at the average wages of movers, we first estimate a less restrictive wage equation and then focus on testing the restrictions with workers who are colleagues at two firms. This method provides a direct statistical test of additive separability and directly indicates the presence of complementarities. If workers base their job mobility decisions on these complementarities, then the restrictions of the AKM log linear wage specification will fail in a more general setting. For that reason, we begin with a general specification and test whether restrictions implied by AKM hold.

We now explain our test. w_{ijt} refers to the wages that worker i earns at firm j at time t. The worker-firm fixed effects model in the empirical literature specifies log wages ($\log w_{ijt}$) as

$$\log w_{ijt} = z'_{it}\gamma + \underbrace{\sum_i \alpha_i D_i + \sum_j \psi_j D_j + u_{it}}_{\text{residual wages}} \tag{10}$$

where $\log w_{ijt}$ are log real wages that worker i earns at firm j at time t, z_{it} are observable characteristics of i at time t with return rates γ, α_i is a worker fixed effect, ψ_j is a firm or establishment fixed effect, D is an indicator variable for the observation of i or j, and u_{it} captures everything else. In this piece-rate wage structure, wage differentials come entirely from worker fixed effect differentials ($\alpha_i - \alpha'_i$) and zero-mean idiosyncratic errors after conditioning on observables like education and experience and working at the same firm. Therefore in expectation, differences in worker fixed effects account for wage differentials between workers at the same firm. If unobserved worker-firm complementarities captured in the error term play a role in workers' decisions to take jobs, then the correlation between the worker's new firm fixed effect and the error term causes worker and firm fixed effects to be inconsistently estimated.

We use a specification that puts these complementarities in the non-error component of wages and test how well additive separability fits this more general wage specification. We begin by specifying log wages more generally as

$$\log w_{ijt} = z_{it}'\gamma + \underbrace{\sum_i \sum_j \varphi_{ij} D_{ij} + u_{it}}_{\text{residual wages}}, \tag{11}$$

where φ_{ij} is the match effect on wages (not to be confused with a match quality shock which we allow for in the error process) and D_{ij} is an indicator variable for the match.[29] Under the null hypothesis of additive separability, the difference-in-difference of φ_{ij} is

$$\Delta_{ij}\varphi \equiv (\varphi_{ij} - \varphi_{ij'}) - (\varphi_{i'j} - \varphi_{i',j'}) = 0, \quad \forall (i, i', j, j') \tag{12}$$

because the two-way fixed effects model amounts to the linear restriction $\varphi_{ij} = \alpha_i + \psi_j$. We test these linear restrictions individually. We construct $\Delta_{ij}\hat{\varphi}$ by taking all possible difference-in-difference combinations (i, i', j, j') observed in the data.

Then, we construct our test statistic

$$TS^{ij} = \frac{\Delta_{ij}\hat{\varphi}}{SE(\Delta_{ij}\hat{\varphi})}. \tag{13}$$

To convey the idea of our test, we assume u_{it} is distributed *i.i.d.* normal and allow for general, persistent error processes and match quality shocks later. This simplification allows us to calculate the standard error of $\Delta_{ij}\hat{\varphi}$ as

$$SE(\Delta_{ij}\hat{\varphi}) = \sqrt{\hat{\sigma}_u^2 \left(\frac{1}{T_{ij}} + \frac{1}{T_{i'j}} + \frac{1}{T_{ij'}} + \frac{1}{T_{i'j'}} \right)} \tag{14}$$

where T_{ij} is the number of periods worker i workers at firm j and $\hat{\sigma}_u^2$ is the consistent estimator of σ_u^2 constructed from the residuals of the match effects regression. Under the null, TS^{ij} is distributed $\mathcal{N}(0, 1)$. We fail to reject the null linear restriction on (i, j) if TS^{ij} falls within an acceptance region.

In practice, we do not restrict errors in the data to be *i.i.d.* normal. We

[29] The technical identification details are relegated to Appendix C.1.

allow φ_{ij} to include a match quality shock and proceed with both parametric and subsampling inference. We parameterize the error process as a stationary AR(1) plus a lognormal match quality and perform our test.[30] We also do two forms of subsampling inference. First, we assume that u_{it} is an arbitrary stationary process and we make asymptotic inference. Second, we make inference based on an approximate finite sample distribution of the test statistic to relax the stationarity assumption. We rely on standard subsampling techniques to make robust inference.[31] Table 7 contains results for our alternative inference methods under the null hypothesis.

Every pair of workers i and i' who are coworkers at firm j and j' provides direct evidence on whether wages are additively separable. If wages were indeed additively separable as assumed under the null, then we would sometimes falsely reject additive separability purely due to error. In the data, we find that the null hypothesis is rejected for a large number of the additive separability restrictions when match quality shocks are less than 15-20% of wage variance. For example, we reject at least three to four times as many restrictions using a 5% test than expected if match quality shocks that make up 5% of wage variance.[32] We present our main results in 7. Table 7 shows that additive separability fails more often than expected given match quality shocks that make up 5% of wage variance. Our prior on the wage variance due to match quality shocks in our dataset is around 2%.[33] Hence, we find match quality shocks to be too small to explain the number of deviations from additive separability that we observe in the data. Additional results allowing for larger and smaller match quality shocks and various error processes can be found in Tables 12 to 16.[34]

[30]The degree of persistence makes no notable difference in our results as shown in Tables 12 to 16 in the Appendix.

[31]We relegate the technical details for these procedures to Appendix C.2 and C.3.

[32]Match quality shocks we consider are defined to be 1) fixed over a job spell and 2) orthogonal to observable characteristics, worker and firm fixed effects, and the error process.

[33]Details on the origins of our prior can be found in Appendix C.2.

[34]In a previous version of this paper, we estimated match quality shocks in the error process. Here, we let match quality shocks vary in the share of wage variance they make up

Table 7: Failed Additive Separability Restrictions

Rejection Region	Parametric Error		Stationary Process		Finite Sample	
	Data	Model	Data	Model	Data	Model
	(1)	(2)	(3)	(4)	(5)	(6)
1%	8.14%	8.70%	10.51%	28.36%	9.53%	28.36%
5%	16.67%	13.90%	19.76%	35.44%	19.80%	35.44%
10%	23.17%	17.62%	26.37%	42.10%	26.96%	42.10%

Notes: The columns labeled "Data" are produced from data itself. The columns labeled "Model" are results using the model-simulated data as described in Section 4. The columns for "Data" represent 1,225,892 unique cases of two coworkers moving between two firms. The number of unique individuals and firms are 11,120 and 793 respectively. The columns for "Model" represent 145,471 unique cases of two coworkers moving between two firms. The number of unique individuals and firms are 9,382 and 120 respectively. If all linear restrictions held, a rejection region of $X\%$ is expected to contain $X\%$ of realizations of $\Delta_{ij}\varphi$. For columns (1) and (2), the parametric error we specify is a stationary AR(1) process with persistence equal to 0.65. We make asymptotic inference with an arbitrary stationary error process in columns (3) and (4). Columns (5) and (6) make finite sample inference using an empirical approximation of the distribution of TS^{ij}. All data cases allow for lognormal zero-mean match quality shocks with a variance equal to 5% of wage variance. Results allowing for various match quality shock variance contribution and persistence are shown in Tables 12 to 16.

We interpret our results as evidence for the presence of nonlinearities in wages as predicted by theory. In particular, structural models of search and matching give rise to these non-separabilities through production complementarities. The model we use replicates log wage non-separabilities found in the data as shown in Table 7.[35] Estimating AKM on model-generated wages which contain these nonlinearities yields R^2 in excess of 0.95. We view our results as support that wages drive sorting between workers and firms in the data as

in order to show that our results are robust to our way of estimating match quality shocks by clustering workers.

[35]The wage error process simulated in the model consists only of $i.i.d.$ measurement error (ϵ_{it}). We allow for match quality shocks in the data, but the model does not contains match quality shocks. Hence, we test whether $\log w_{ijt} = \alpha_i + \psi_j + \epsilon_{it}$. We reject additively separability restrictions even when the error process is misspecified as an AR(1) with $\rho > 0$, thereby upwardly biasing the standard errors and thus making it more difficult to reject additive separability.

they do in the model.

7 Conclusion

We estimate a standard search model described and identified in Hagedorn, Law, and Manovskii (2014) which features sorting between heterogeneous workers and firms. Wages, and only wages, guide the sorting of workers to firms in the model. This is consistent with models encountered in much of the theoretical literature on worker assignment and sorting. We find that the model fits the data well along key dimensions as it replicates wage means, variances, mobility rates, and sorting between workers and firms. Residual wages predicted by the model, together with observable characteristics generate R^2 statistics that are comparable to that of standard two-way fixed effects linear decompositions. These decompositions of log wages use many more degrees of freedom to obtain the same order of fit that we achieve with a parsimonious structural model. The use of this model permits a counter-factual analysis to disentangle the importance of production and search technology on wage dispersion.

We apply this method to examine the rise in German wage inequality in the 1990s and the 2000s and find that changes in production complementarities are responsible for the rise in residual wage variance. A significant channel through which production technology affects the increase in wage variance is through the reallocation of workers to firms induced by changes in wages. Search technology also plays an important role in determining wage variance through the allocation of workers, despite having little impact over the periods we consider.

Overall, we find the data to be consistent with theory in which wages guide the sorting of workers to firms. This finding might appear surprising in light of the well known fact that two-way fixed effects regressions fit the data extremely well, and these wage specifications limit the role wages play in sorting. The fact that our model and fixed effect log wage regressions account equally well for wages begs the question of which approach is more consistent with the data. The key difference is that log wages are assumed to be linear in worker and firm fixed effects in these regressions, while they are nonlinear in the structural model we use. We design and implement a test to directly detect the presence of nonlinearities in the wage data. In particular, we compare wage differentials of two workers observed working at two different firms. We find that the variability of these wage differentials across firms in the data is consistent with the structural model but not with the log-linear additive specification.

References

ABOWD, J. M., R. H. CREECY, AND F. KRAMARZ (2002): "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," LEHD program technical paper, U.S. Census Bureau.

ABOWD, J. M., AND F. KRAMARZ (1999): "Chapter 40 The analysis of labor markets using matched employer-employee data," vol. 3, Part 2 of Handbook of Labor Economics, pp. 2629 – 2710. Elsevier.

ABOWD, J. M., F. KRAMARZ, AND D. N. MARGOLIS (1999): "High Wage Workers and High Wage Firms," Econometrica, 67(2), 251–334.

ABOWD, J. M., F. KRAMARZ, S. PÉREZ-DUARTE, AND I. M. SCHMUTTE (2014): "Sorting Between and Within Industries: A Testable Model of Assortative Matching," NBER Working Papers 20472, National Bureau of Economic Research, Inc.

ANDREWS, M., 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.

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, 171(3), 673–697.

Bagger, J., and R. Lentz (2014): "An Empirical Model of Wage Dispersion with Sorting," mimeo, Univerity of Wisconsin-Madison.

Barth, E., A. Bryson, J. C. Davis, and R. Freeman (2014): "It's Where You Work: Increases in Earnings Dispersion across Establishments and Individuals in the U.S," NBER Working Papers 20447, National Bureau of Economic Research, Inc.

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

Burdett, K., and D. Mortensen (1998): "Wage Differentials, Employer Size, and Unemployment," International Economic Review, 39(2), 257–273.

Cahuc, P., F. Postel-Vinay, and J.-M. Robin (2006): "Wage Bargaining with On-the-Job Search: Theory and Evidence," Econometrica, 74(2), 323–364.

Card, D., J. Heining, and P. Kline (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality," The Quarterly Journal of Economics, 128(3), 967–1015.

Dustmann, C., J. Ludsteck, and U. Schönberg (2009): "Revisiting the German Wage Structure," Quarterly Journal of Economics, 124(2), 363–376.

Eeckhout, J., and P. Kircher (2011): "Identifying Sorting – In Theory," The Review of Economic Studies, 78(3), 872–906.

Fitzenberger, B., A. Osikominu, and R. Völter (2006): "Imputation Rules to Improve the Education Variable in the IAB Employment Subsample," Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift fr Wirtschafts- und Sozialwissenschaften, Duncker & Humblot, Berlin, 126(3), 405–436.

Gautier, P. A., and C. N. Teulings (2006): "How Large Are Search Frictions?," Journal of the European Economic Association, 4(6), 1193–1225.

——— (2012): "Sorting and the Output Loss due to Search Frictions," Discussion Paper TI 2011-010/3, Tinbergen Institute.

Hagedorn, M., T.-H. Law, and I. Manovskii (2014): "Identifying Equilibrium Models of Labor Market Sorting," NBER Working Papers 18661, National Bureau of Economic Research, Inc.

Horowitz, J. L. (2001): The Bootstrap, vol. 5 of Handbook of Econometrics. Elsevier.

Jung, S., and C. Schnabel (2011): "Paying More than Necessary? The Wage Cushion in Germany," LABOUR, 25(2), 182–197.

Klosterhuber, W., J. Heining, and S. Seth (2013): "Linked Employer-Employee Data from the IAB: LIAB Longitudinal Model 1993-2010 (LIAB LM 9310)," Discussion paper, Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Lamadon, T., J. Lise, C. Meghir, and J.-M. Robin (2014): "Matching, Sorting, and Wages," Working paper, University College London.

Lise, J., C. Meghir, and J.-M. Robin (2011): "Matching, Sorting, and Wages," Working paper, University College London.

Lopes de Melo, R. (2013): "Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence," Working paper, University of Chicago.

Mittag, N. (2015): "A Simple Method to Estimate Large Fixed Effects Models Applied to Wage Determinants and Matching," CERGE-EI Working Papers wp532, The Center for Economic Research and Graduate Education - Economics Institute, Prague.

Munkres, J. (1957): "Algorithms for the Assignment and Transportation Problems," Journal of the Society of Industrial and Applied Mathematics, 5(1), 32–38.

Nickell, S. J. (1981): "Biases in Dynamic Models with Fixed Effects," Econometrica, 49(6), 1417–26.

Politis, D. N., and J. P. Romano (1994): "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions," The Annals of Statistics, 22(4), pp. 2031–2050.

Shimer, R. (2005): "The Assignment of Workers to Jobs in an Economy with Coordination Frictions," Journal of Political Economy, 113(5), 996–1025.

Shimer, R., and L. Smith (2000): "Assortative Matching and Search," Econometrica, 68(2), 343–370.

Woodcock, S. D. (2015): "Match effects," Research in Economics, 69(1), 100–121.

Appendix (For Online Publication)

A Worker-Firm Rankings with Fixed Effects

We show in this Appendix that worker and firm fixed effects can identify the productivity ranks of workers and firms when the wage function is increasing (but not necessarily additive separable) in worker and firm productivity types. This is guaranteed when the underlying match density is uniform. However, the identification of ranks is not guaranteed when the match density is not uniform.

A.1 Fixed Effects Identify Productivity Ranks when Match Density is Uniform

Some context helps. Let i represent the worker identifier that, without loss of generality, is also the worker's rank. j is the firm identifier/rank.

Represent log wages as $w(i,j)$ as a draw from a joint probability mass function with support $S = \{(i,j)|i = 1,...,I; j = 1,...,J\}$ and $\alpha_i + \psi_j$ as a numerical approximation to $w(i,j)$ where $\{\alpha_i\}_{i=1,...,I}$ and $\{\psi_j\}_{j=1,...,J}$ are real numbers. Denote the number of observations of $w(i,j)$ with n_{ij}. Consequently, the match density at (i,j) is $\theta_{ij} = \frac{n_{ij}}{\sum_{ij} n_{ij}}$. The total squared approximation error from least squares estimation of α and ψ is

$$\epsilon^* = \min_{\{\alpha,\psi\}} \sum_i \sum_j \theta_{ij} \left(w(i,j) - \alpha(i) - \psi(j)\right)^2$$

$$s.t.$$

$$\sum_{j=1}^{J} \psi(j) = 0$$

where the last constraint serves to eliminate trivial multiplicity. Note that $\theta_{ij} = \frac{1}{I \cdot J}$ $\forall$ i,j corresponds to a uniform joint probability mass function for

$w(i,j)$.

Lemma 1

Consider four real numbers. $\{w_L, w_H\}$ represent wages where $w_L < w_H$. $\{\alpha_L, \alpha_H\}$ represents fixed effects. $\alpha_L < \alpha_H$, if and only if

$$(w_L - \alpha_L)^2 + (w_H - \alpha_H)^2 < (w_L - \alpha_H)^2 + (w_H - \alpha_L)^2.$$

Proof. Expanding and canceling terms in the expression above yields

$$0 < (\alpha_H - \alpha_L)(w_H - w_L) \Leftrightarrow \alpha_L < \alpha_H \blacksquare$$

Proposition 1: Ranks are identified when the match density is uniform.

Suppose $w(i,j)$ is strictly increasing in i and j, but not necessarily additive separable. If $\theta_{ij} = \frac{1}{I \cdot J}$ $\forall$ i, j, then least squares estimates $\{\alpha_i^*\}_{i=1,..,I}$ are strictly increasing in i.

Proof. Suppose the fixed effects are not increasing in i. Then, there exists some k such that $\alpha_k^* > \alpha_{k+1}^*$. Let $\tilde{w}(i,j)$ denote $w(i,j) - \psi_j^*$. Now

$$\begin{aligned}
\epsilon^* \cdot I \cdot J = & \sum_j \sum_i \left(\tilde{w}(i,j) - \alpha_i^*\right)^2 \\
= & \sum_j \Bigg[\Big(\tilde{w}(1,j) - \alpha_1^*\Big)^2 + \Big(\tilde{w}(2,j) - \alpha_2^*\Big)^2 + ... \\
& + \underbrace{\Big(\tilde{w}(k,j) - \alpha_k^*\Big)^2 + \Big(\tilde{w}(k+1,j) - \alpha_{k+1}^*\Big)^2}_{\tilde{w}(k,j) < \tilde{w}(k+1,j),\ \alpha_k^* > \alpha_{k+1}^*} + ... \\
& + \Big(\tilde{w}(I-1,j) - \alpha_{I-1}^*\Big)^2 + \Big(\tilde{w}(I,j) - \alpha_I^*\Big)^2 \Bigg]
\end{aligned}$$

by Lemma 1

$$
\begin{aligned}
> \sum_j \Bigg[& \Big(\tilde{w}(1,j) - \alpha_1^*\Big)^2 + \Big(\tilde{w}(2,j) - \alpha_2^*\Big)^2 + ... \\
& + \underbrace{\Big(\tilde{w}(k,j) - \alpha_{k+1}^*\Big)^2 + \Big(\tilde{w}(k+1,j) - \alpha_k^*\Big)^2}_{\alpha_k^* \text{ and } \alpha_{k+1}^* \text{ are swapped}} + ... \\
& + \Big(\tilde{w}(I-1,j) - \alpha_{I-1}^*\Big)^2 + \Big(\tilde{w}(I,j) - \alpha_I^*\Big)^2 \Bigg]
\end{aligned}
$$

which is a contradiction to the assumption that α^* and ψ^* being the least squares solution. The case for j follows immediately.

A.2 Identification Failure under Nonuniform Match Density

Identification of ranks is not guaranteed when the match density is not uniform, i.e. when there is sorting. We provide a very simple example where fixed effects do not identify ranks.

Counterexample. Suppose log wages are $w(i,j) = i \cdot j$ where $i \in \{1,2,3\}$ and $j \in \{1,2,3\}$. Suppose the observed distribution of wage is given by $m(i,j)$ which is

$$
m(i,j) = \begin{cases}
0.5 & i=1, j=1 \\
0 & i=1, j=2 \\
0.5 & i=1, j=3 \\
0.1 & i=2, j=1 \\
0.5 & i=2, j=2 \\
0.4 & i=2, j=3 \\
1 & i=3, j=1 \\
0 & i=3, j=2 \\
0 & i=3, j=2
\end{cases}
$$

Then the least squares estimates of the workers fixed effects are

$$p(i) = \begin{cases} 3.242 & i = 1 \\ 5.697 & i = 2 \\ 5.484 & i = 3 \end{cases},$$

which ranks worker Type 2 as Type 3, and worker Type 3 as Type 2 based on their fixed effects from smallest to largest. Workers are ranked incorrectly, because the fixed effects are not increasing in underlying type (i) under this nonuniform density for $w(i, j)$. The proof which guaranteed identification of ranks under the uniform match density fails because attaching nonuniform weights to the objective invalidates Lemma 1.

B Data

This section provides further details on our data and imputation procedures. We describe the wage trends we observe in our data and the distribution of firm sizes. We also describe the imputation methods for education and censored wages.

B.1 Wage Dispersion in Germany

Table 8 shows the rise in wage inequality in West Germany from 1993 to 2007 by the percentile ratios. Conditioning on age, year, and education (residual wages), we find increasing dispersion in wages as shown in Table 9 within these age and education groups. Despite being a subset of German wages, our data exhibits similar wage dispersion patterns seen in Dustmann, Ludsteck, and Schönberg (2009) and Card, Heining, and Kline (2013). Card, Heining, and Kline (2013) attribute rising residual wage inequality to increasing worker and firm heterogeneity from the covariance structure of wages. Dustmann, Ludsteck, and Schönberg (2009) decompose the rise in inequality due to observable changes in the workforce composition and the market prices on these observables. These observables account for a significant portion of rising wage dispersion, but still much of it is due to residual wage inequality rising. Table 9 shows this rise. The wage gap between the 90^{th} and 10^{th} and the 50^{th} and 10^{th} percentiles grew over our observation period using worker and firm fixed effects to model wage residuals. Our sample generally exhibits these same trends found in Dustmann, Ludsteck, and Schönberg (2009). Overall, log wage variance grew from 0.196 to 0.312 from the 1990s to 2000s in our dataset. Residual log wage variance increased from 0.169 to 0.205 over the same period.

Table 8: Percentile Ratio of Real Daily Logwages

	90-10 pctile	90-50 pctile	50-10 pctile
1990s	1.285	1.115	1.152
2000s	1.308	1.119	1.169

Table 9: Percentiles of Residual Daily Logwages

	90-10 pctile	90-50 pctile	50-10 pctile
1990s	1.286	1.112	1.156
2000s	1.346	1.110	1.212

Note: These tables illustrate the ratios of the 90-10, 50-10, and 90-50 imputed real daily log wages (see Appendix B.2) and residual log wage percentiles in 1993 and 2007. The base year is 1995. Regressions control for age-squared, age-cubed and year all interacted with education. Residuals refer to $\log wage - z'\hat{\gamma}$ where $z'\hat{\gamma}$ are estimated returns to the control variables.

B.2 Education Imputation

The education variable in the LIAB data comes from establishment reports to the Social Security Administration. It contains missing entries and inconsistencies. For example, education may drop from university to vocational schooling in a job spell and go back to university. We impute missing education variable using the IP1 imputation procedure developed by Fitzenberger, Osikominu, and Völter (2006). This procedure assumes establishments never over-report a worker's education and thereby forces education to weakly increase over time. This assumption makes use of the fact that the German social security office requires employers to report the highest education obtained by a worker. Hence, the education record should increase weakly as workers may acquire more education. IP1 education contains four main categories: 1) less than secondary education; 2) less than secondary education with a vocational qualification; 3) secondary education with/without vocation training; and 4) university or technical degree. We find education missing 10.73% of the observations pre-imputation and 0.01% post-imputation. As in Dustmann, Ludsteck, and Schönberg (2009), we record missings as the lowest education level after imputation.

B.3 Wage Imputation

The LIAB only reports wages up to earnings limit for social security contributions (Klosterhuber, Heining, and Seth, 2013). We find a censoring rate of 9% among all wage observations in our sample and impute the censored values. Censoring occurs evenly across the years. Our method is less sensitive to censored wages because we estimate the model using from out-of-unemployment wages which exhibit only a 2% censoring rate. First, we convert daily wages to real daily log wages using the CPI with base year 1995. Second, we fit tobit models on age-education-year cells to impute the censored upper tail following Dustmann, Ludsteck, and Schönberg (2009) and Card, Heining, and Kline (2013). We include age, job tenure, the fraction of individual wages censored at all jobs, the mean individual wage, the fraction of censored wages of lifetime coworkers, the mean of wages of lifetime coworkers, and the fraction of lifetime coworkers with some college or university education in our censored regression on log wages. We cannot observe all coworkers at non-survey establishments, so we instead use the characteristics of all coworkers observed in a worker's lifetime. These variables reflect characteristics of the worker over their lifetime rather than the establishment at a point in time. Third, we add a normal error term scaled to the variance of the age-education-year cell from the fitted value of log wages. This forms the imputed wage. We leave the wage at the real wage censored limit whenever the imputed value falls below the censored value. The imputation yields the log wages over the observation period shown in Table 10.

Table 10: Daily Log Wages (1993−2007)

	Mean	Variance	Min	Max
Censored	4.274	0.198	2.302	4.909
Fitted	4.283	0.213	2.302	6.276
Imputed	4.290	0.222	2.302	6.152

Note: Daily log wages are real daily wages computed from the CPI with the base year 1995. The sample contains 383,772 establishments (LIAB and non-LIAB), 889,307 workers, and 6,254,298 observations. Imputed wages add a draw from a normal distribution (centered at zero with variance equal to the estimated variance of age-education-year cell) to the fitted log wage.

B.4 Firm Sizes

In Table 11, we show the establishment size distribution and its relation to establishment productivity type.[36] We observe that higher productivity type establishments tend to be larger.

[36] These establishments hire at least 10 workers.

Table 11: Firm Size Distribution by Type (1993−2007)

Firm Types 1 − 10

Firm Size	Mean Numbers of Workers	Number of Firms
1 − 49	15	693
50 − 99	73	90
100 − 199	141	59
200+	392	23

Firm Types 11 − 20

Firm Size	Mean Numbers of Workers	Number of Firms
1 − 49	24	291
50 − 99	72	150
100 − 199	140	151
200+	976	274

Note: This sample includes surveyed LIAB establishments over the 1993-2007 sample period.

C Testing Additive Separability

In this appendix, we continue from Section 6 and discuss the econometric issues for identifying match effects. We then explain our parametric inference and subsampling methods for making asymptotic and finite sample inference on whether additive separability restrictions hold.

C.1 Identification of Match Effects

The consistent estimation of the match effect (φ) in Equation (15) requires the strict exogeneity assumptions shown in (16).

$$\log w_{ijt} = z_{it}'\gamma + \sum_i \sum_j \varphi_{ij} D_{ij} + u_{it} \tag{15}$$

$$\mathbb{E}[z_{it}u_{it}] = 0,\ \mathbb{E}[\varphi_{ij}u_{it}] = 0\ \forall i,j,t \tag{16}$$

Strict exogeneity requires that the regressor be uncorrelated with current, past and future values of the error term. We make the standard assumption on the orthogonality of the observable regressors (z_{it}) and the error term (u_{it}). Similar to Card, Heining, and Kline (2013), we assume a sufficient condition on the assignment of workers to jobs to ensure orthogonality between the match effects and error term. Assuming the assignment to a job defined by (i, j) does not depend on the error term is a sufficient condition for (16) to hold. This condition is known in the literature as exogenous mobility. φ_{ij} encompasses any relationship described by (i, j), so workers may sort into jobs based on anything in this match component. However, they cannot sort into the job (i, j) based on components in u. The match component may consist of worker effects (α_i), firm effects (ψ_j), and a match quality shock (η_{ij}) for example. A match quality shock is an idiosyncratic wage shock realized for a particular (i, j) match. This shock is often assumed to be orthogonal to person and firm fixed effects. This condition is not necessary to consistently estimate γ using an AKM regression nor a match effects regression as Woodcock (2015) notes. The exogenous mobility condition for the match effect regression is weaker than the exogenous mobility condition required for AKM, because workers may sort based on some match quality shock that enters the match effect in addition to separable worker and firm effects. Note that exogenous mobility is only a sufficient condition. If exogenous mobility on worker and firm fixed effects holds, then match effect identification additionally requires the match quality shock to be uncorrelated with the rest of the error component. Here, we understand match quality shocks as idiosyncratic, match-dependent wage shocks that are orthogonal to other idiosyncratic shocks (u_{it}) like productivity for example.

Within the additively separable fixed effect regression framework, recent evidence from Woodcock suggests that omitting match effects biases the estimate of returns to observable characteristics (γ). Woodcock adds a match effect (i.e. a time invariant match quality shock in our terminology) to the specification with worker and firm fixed effects and shows this bias using US match employer-employee data. Mittag (2015) finds similar evidence of bias in γ for the German LIAB dataset we employ. Hence, estimating γ consistently is a clear advantage of constructing our test based on the match effect rather than residuals from the AKM regression.

Using the match effects regression raises estimation concerns similar to AKM. First, idiosyncratic shocks (u_{it}) may induce correlation between the regressors and past values of the error term. A potential violation to (16) occurs when u_{it} predicts job transitions or observables like education. For example, a persistent positive shock to earnings may yield transitions to higher earning jobs if the shock occurs early in life, allowing a worker to invest in more education. An education decision based on a past u_{it} induces correlation between the regressors and past values of the error term, which biases the match effect and γ estimates. We will also have this same bias spread among the worker and firm fixed effects in an AKM regression. As Card, Heining, and Kline (2013) note, decisions that determine current observables based on past values of the *fixed effects* will not violate the exogeneity assumption. If such shocks to earnings (inducing more education) are due to match effects components, then the identification assumption will not be violated. However, if workers move to firm on the basis of unobserved, idiosyncratic productivity shocks, for example, then this assumption will surely be violated as the fixed effect(s) will be correlated with the error (i.e. $\mathbb{E}[\varphi_{ij}u_{it}] \neq 0 \ \forall i, j, t$).

Second, the match effect regression introduces more multicollinearity than the more parsimonious worker-firm fixed effect regression. This multicollinearity may make estimation of γ less precise. When a worker moves to a new firm but acquires more education in between jobs, then the wage increase will be attributed to both the match effect and the new education value. This occurs because match effects saturate the regression, making estimating parameters on observables less precise. Assuming (16), we still lose efficiency in estimating the coefficients on observables like education when moving from an AKM regression to a match effect regression. However, we do not consider efficiency loss in estimating γ to be a concern, because our test relies solely on the consistency of $\hat{\gamma}$. Despite the potential for the match effect to absorb most of the effects on education and experience, consistency of $\hat{\gamma}$ in the worker-firm dimension. Assuming we consistently estimate AKM and the match effect regression, we find that the estimates on returns to education and experience to be highly correlated (0.94) across the two regressions. In short, the bias-variance tradeoff between the match effect and AKM estimators for γ appears to be relatively moderate.

C.2 Parametric Inference

We specify the composite error process (u_{it}) and derive the resulting standard errors for $\Delta_{ij}\hat{\varphi}$ to do parametric inference. Our parametric model for u_{it} consists of an orthogonal match quality shock $\eta_{i,\mathcal{J}(i,t)}$ and an exogenous AR(1) process (ϵ_{it}). $\mathcal{J}(i,t)$ denotes the firm of worker i at time t and ρ is the degree of persistent in the AR process.

$$\begin{aligned} u_{it} &= \eta_{i,\mathcal{J}(i,t)} + \epsilon_{it} \\ \epsilon_{it} &= \rho\epsilon_{i,t-1} + \nu_{it} \\ \eta_{i,\mathcal{J}(i,t)} &\sim i.i.d.N(0,\sigma_\eta^2)\ \forall i,j,t \\ \nu_{it} &\sim i.i.d.N(0,\sigma_\nu^2)\ \forall i,t \end{aligned}$$

We impose the restriction that $\rho < 1$, which allows for an arbitrarily persistence process but not an exact unit root process. It can be shown that the test statistic under H_0 ($\Delta_{ij}\varphi = 0$) is

$$\Delta_{ij}\hat{\varphi} = \Delta_{ij}\bar{x}'(\gamma - \hat{\gamma}) + \Delta_{ij}\eta + \frac{1}{\underbrace{\Delta t_{ij}}_{=T_{ij}-t_{ij}} + 1} \sum_{s=t_{ij}}^{T_{ij}} \sum_{k=0}^{s} \rho^k \nu_{s-k} + ...$$

$$+ \frac{1}{\Delta t_{i'j'} + 1} \sum_{s=t_{i'j'}}^{T_{i'j'}} \sum_{k=0}^{s} \rho^k \nu_{s-k},$$

dropping subscripts for η and x where $\bar{x}$ is the within match average. Under our parametric assumptions, we have the following distributions for the components of $\Delta_{ij}\hat{\varphi}$

$$\Delta_{ij}\bar{x}'(\gamma - \hat{\gamma}) \sim N(0, \Delta_{ij}\bar{x}'\mathbb{V}(\hat{\gamma})\Delta_{ij}\bar{x}) \text{ (Estimation Error)}$$

$$\Delta_{ij}\eta \sim N(0, 4\sigma_\eta^2) \text{ (Match Quality Shock)}$$

$$\sum_i \sum_j \frac{1}{\Delta t_{ij} + 1} \sum_{s=t_{ij}}^{T_{ij}} \sum_{k=0}^{s} \rho^k \nu_{s-k} \sim N(0, \Omega) \text{ (AR(1) Error Process)}$$

where $\Omega = \mathbb{V}\left[\frac{1}{\Delta t_{ij}+1} \sum_{s=t_{ij}}^{T_{ij}} \sum_{k=0}^{s} \rho^k \nu_{s-k} + ...\right]$.

It can also be shown that

$$\mathbb{V}\left[\sum_{s=t}^{T} \sum_{k=0}^{s} \rho^k \nu_{s-k}\right] = \sigma_\nu^2 \left(\left[\frac{1-\rho^{\Delta t+1}}{1-\rho}\right]^2 \left(\frac{\rho^2 - \rho^{2(t+1)}}{1-\rho^2}\right) + \left(\frac{1}{1-\rho}\right)^2 \left(1 + \Delta t - 2\rho \cdot \frac{1-\rho^{\Delta t+1}}{1-\rho} + \rho^2 \cdot \frac{1-\rho^{2(\Delta t+1)}}{1-\rho^2}\right)\right)$$

and

$$
cov\left(\sum_{s=t_1}^{T_1}\sum_{k=0}^{s}\rho^k\nu_{s-k},\sum_{s=t_2}^{T_2}\sum_{k=0}^{s}\rho^k\nu_{s-k}\right)=
$$

$$
\sigma_\nu^2\cdot\left[\left(\frac{1-\rho^{\Delta t_1+1}}{1-\rho}\right)\cdot\left(\frac{1-\rho^{\Delta t_2+1}}{1-\rho}\right)\cdot\rho^2\cdot\frac{\rho^{t_1+t_2}-\rho^{t_2-t_1}}{1-\rho^2}+\right.
$$

$$
\left.\left(\frac{1-\rho^{\Delta t_2+1}}{1-\rho}\right)\cdot\left(\frac{1}{1-\rho}\right)\cdot\left(\frac{\rho^{t_2-T_1}-\rho^{t_2-t_1+1}}{1-\rho}-\frac{\rho^{t_2-T_1+1}-\rho^{\Delta t_1+(t_2-t_1)+3}}{1-\rho^2}\right)\right].
$$

Hence, we have the variance and covariance terms to construct Ω. $\hat{\gamma}\to_p\gamma$ as $n\to\infty$, thus we obtain the following approximate distribution

$$
\Delta_{ij}\hat{\varphi}|x\approx N\left(0,4\sigma_\eta^2+\Omega\right)
$$

under H_0. Ω depends on (ρ,σ_v^2) and the start and end dates of the matches in the quartet (i,i',j,j'), so we need $(\rho,\sigma_v^2,\sigma_\eta^2)$ to compute the standard error $\left(\sqrt{4\sigma_\eta^2+\Omega}\right)$. In practice, we discretize a grid over ρ and σ_η^2 over which we conduct our test, because of the difficulty of obtaining a consistent estimates of (ρ,σ_η^2).[37] We present results for persistency ranging from 0 (*i.i.d.* errors) to 0.65. The rejection rate does not vary greatly in the degree of persistence, so our parametric results are robust to persistence in the AR error process. We discretize the variance of match quality shocks (σ_η^2) using a grid of the share of variance due to match quality shocks. The grid ranges from 0 to 30% of wage variance. Tables 12, 13, and 14 show that the orthogonal match quality shocks need to be 15 to 20% of wage variance to not reject the null of additive separability under our parametric specification. This range exceeds our prior on the share of variance attributable to orthogonal match quality shocks in our dataset by an order of at least 5.[38]

[37]See Nickell (1981) for an explanation of the bias in estimating persistency (ρ) by standard least squares methods. See Woodcock (2015) for an argument on the bias in estimating σ_η^2 by standard least squares methods.

[38]Woodcock (2015) provides estimates of the variance of match quality for the US, which

Table 12: $\Delta_{ij}\varphi$ Parametric Rejection Rate at 10% level

$\rho\backslash\eta$	0.00	0.05	0.10	0.15	0.20	0.25	0.30
0.00	43.67	23.62	14.72	9.64	6.54	4.56	3.26
0.10	43.03	23.43	14.61	9.57	6.50	4.54	3.24
0.20	42.53	23.28	14.52	9.52	6.46	4.52	3.22
0.30	42.15	23.16	14.45	9.48	6.44	4.50	3.21
0.40	41.88	23.08	14.41	9.45	6.42	4.49	3.21
0.50	41.76	23.06	14.39	9.43	6.41	4.48	3.20
0.65	41.96	23.17	14.46	9.48	6.44	4.50	3.22

Notes: The sample size is 1,225,892 observations, 11,120 workers, and 793 firms. The rows show rejection rates for different values of ρ. The columns show reject rates for different values of $\sigma^2_\eta/Var(\log w_{ijt})$. The match quality shock variance is discretized as a share of wage variance. The the grid for σ^2_η in levels is {0.00, 0.01, 0.02, 0.04, 0.05, 0.06, 0.07}.

Table 13: $\Delta_{ij}\varphi$ Parametric Rejection Rate at 5% level

$\rho\backslash\eta$	0.00	0.05	0.10	0.15	0.20	0.25	0.30
0.00	36.96	17.14	9.29	5.41	3.32	2.09	1.34
0.10	36.30	16.95	9.20	5.37	3.30	2.08	1.34
0.20	35.78	16.80	9.12	5.34	3.28	2.07	1.33
0.30	35.38	16.67	9.07	5.30	3.26	2.06	1.33
0.40	35.10	16.58	9.03	5.29	3.25	2.05	1.32
0.50	34.97	16.55	9.02	5.28	3.25	2.05	1.32
0.65	35.21	16.67	9.07	5.31	3.27	2.06	1.33

Notes: See Table 12.

ranges from 2% in the case of orthogonal match quality shocks to 18% of wage variance using a mixed effects estimator. The mixed effects estimator assumes worker effects, firm effects and match quality shocks are random effects and have zero covariance conditional on the error term and $\hat{\gamma}$. No similar estimates exist to our knowledge for Germany on the LIAB M3 panel, however our estimate in the case of orthogonal match quality shocks is around 2% of wage variance. This estimate is also subject to upward bias as Woodcock shows.

Table 14: $\Delta_{ij}\varphi$ Parametric Rejection Rate at 1% level

$\rho\backslash\eta$	0.00	0.05	0.10	0.15	0.20	0.25	0.30
0.00	26.23	8.50	3.49	1.57	0.76	0.41	0.25
0.10	25.54	8.36	3.45	1.56	0.75	0.41	0.25
0.20	24.99	8.24	3.40	1.54	0.75	0.41	0.25
0.30	24.58	8.15	3.37	1.53	0.75	0.40	0.25
0.40	24.31	8.09	3.35	1.52	0.74	0.40	0.25
0.50	24.20	8.06	3.35	1.52	0.74	0.40	0.25
0.65	24.45	8.14	3.38	1.54	0.75	0.41	0.25

Notes: See Table 12.

C.3 Subsampling Asymptotic and Finite Sample Inference

Assuming a stationary error process yields the asymptotically pivotal test statistic

$$TS^{ij} = \frac{\Delta_{ij}\hat{\varphi}}{SE(\Delta_{ij}\hat{\varphi})} \rightarrow_d N(0,1)$$

by the Lindenberg-Lévy Central Limit Theorem. This standardized test statistic converges to a standard normal distribution as $T \rightarrow \infty$. Therefore, we can conduct inference using the asymptotic critical values. This inference requires knowing the error process in order to calculate exact standard errors as was the case in our parametric inference. We can also make allowances for error processes more general than the AR(1). We use block subsampling to obtain a finite sample approximation of the standard errors, which is identical to block bootstrapping using a sample smaller than the original sample. We also obtain consistent finite sample approximations to the critical values of the test-statistic (TS^{ij}) using block subsampling with replacement.

The subsampling technique mirrors block bootstrapping on a subsample (Horowitz, 2001). Politis and Romano (1994) provide weak conditions under which subsampling yields consistent estimates of aspects of the cumulative distribution function like critical values. These conditions amount to the existence of a limiting distribution for the appropriately normalized test statistic under the true model. Consistency requires that $T \to \infty$, $B \to \infty$, and $\frac{B}{T} \to 0$ where B is the number of random subsamples. In our case, a random subsample is a subset of individual histories. The Politis and Romano (1994) consistency theorem holds for stationary data. Hence, we can consistently estimate finite distribution critical values for TS^{ij}, and we can consistently estimate the asymptotic standard error of $\Delta_{ij}\varphi$ assuming only an arbitrary stationary error process. We use a parametric version of random subsampling with replacement where we block resample the residual to preserve the correlation structure of the errors.

Our random subsampling procedures draws a subset of individual histories and resamples residuals using the specification in Equation (15) to obtain approximate finite sample distributions of $\{\hat{\varphi}_{ij}\}$ for all available (i, i', j, j') quartets in addition to an estimate of $SE(\Delta_{ij}\hat{\varphi})$. We resample a normal (lognormal in levels) match quality shock over different variance parameterizations and resample the stationary error from the residuals of equation (15). We resample so that each quartet (i, i', j, j') appears at least 100 times. The full random subsampling with replacement procedure goes as described in Politis and Romano (1994) and Horowitz (2001). In practice, we set B to be 500 worker histories. In simulation, we find that the rejection region bounds converge to their true finite sample rejection bounds within 200 draws for log wages generated from a model with additively separable worker and fixed fixed effects, match quality shocks, and a normal *i.i.d.* error. We use the estimates from each subsample $\{\{\hat{\varphi}^b_{ij}\}^B_{b=1}\}_{ij}$ to obtain finite sample approximations to the asymptotic standard error and the distribution of each TS^{ij} for every quartet. We report the main results of this procedure in Section 6, and Tables 15 and 16 report full results. Our subsampling inference yields similar conclusions as the parametric inference. Orthogonal match quality shocks need to be 15 to 20% of wage variance to not reject the null of additive separability for an arbitrary error process and lognormal match quality shocks.

Table 15: Rejection Rates using Bootstrapped Standard Errors

η	10%	5%	1%
0.00	41.10	34.27	23.52
0.01	37.03	30.18	19.67
0.02	33.80	26.95	16.69
0.05	26.37	19.76	10.51
0.10	18.45	12.44	5.34
0.15	13.46	8.33	3.02
0.20	9.96	5.66	1.69
0.25	7.47	3.94	0.99
0.30	5.72	2.80	0.60

Notes: The sample size is 1,225,892 observations, 11,120 workers, and 793 firms. The rows show reject rates for different values of $\sigma_\eta^2/Var(\log w_{ijt}$. The match quality shock variance is discretized on a grid as a share of wage variance.

Table 16: Rejection Rates using Bootstrapped Empirical Distribution

η	10%	5%	1%
0.00	42.60	33.72	17.94
0.01	38.30	29.98	15.69
0.02	34.80	26.81	13.83
0.05	26.96	19.80	9.53
0.10	18.79	12.52	5.00
0.15	13.69	8.41	2.99
0.20	10.09	5.72	1.73
0.25	7.63	4.02	1.05
0.30	5.85	2.89	0.67

Notes: See Table 15.