High Wage Workers Work for High Wage Firms*

Katarína Borovičková
New York University

Robert Shimer
University of Chicago

February 13, 2020

Abstract

We propose a new measure of the correlation between the types of matched workers and firms and show that this captures sorting in a variety of structural models. We also propose an estimator of the correlation and prove that the estimator is consistent when the number of workers and firms grows to infinity even if each worker only has a small number of jobs and each firm only employs a small number of workers. Model simulations also confirm that our estimator is accurate in small data sets. Using administrative data from Austria, we find that the correlation between worker and firm types lies between 0.4 and 0.6. In contrast, the Abowd, Kramarz, and Margolis (1999) fixed effects estimator suggests a near-zero correlation in our data set. This reflects a combination of biases in the AKM correlation estimator and limitations of the AKM correlation as a measure of sorting.

*We are grateful for comments from John Abowd, Fernando Alvarez, Stéphane Bonhomme, Jaroslav Borovička, Thibaut Lamadon, Rasmus Lentz, Ilse Lindenlaub, Elena Manresa, Magne Mogstad, Derek Neal, Martin Rotemberg, and Christopher Taber, as well as participants in various seminars. Any remaining errors are our own. This material is based in part on work supported by the National Science Foundation under grant numbers SES-1559225 and SES-1559459.

1 Introduction

There is sorting everywhere in the economy. Wealthier, more educated, more attractive men on average marry wealthier, more educated, more attractive women (Becker, 1973). Higher income households reside in distinct neighborhoods and send their children to different schools than low income households (Tiebout, 1956). Elite universities enroll the most qualified undergraduates (Solomon, 1975). The one place where it has been hard to find evidence of sorting is in the labor market. A fair summary of an extensive literature following Abowd, Kramarz, and Margolis (1999) (hereafter AKM) is that the correlation between the fixed characteristics of workers and their employers is close to zero and possibly negative.[1] This is often interpreted as saying that there is no evidence that high wage workers work for high wage firms and is used to justify theoretical models in which there is no sorting between workers and firms (Postel-Vinay and Robin, 2002; Christensen, Lentz, Mortensen, Neumann, and Werwatz, 2005).

This paper proposes a new measure of sorting and revisits this conclusion through the lens of that measure. Our measure of sorting is the correlation between a worker's type and her employer's type. We define a worker's type to be the expected log wage she receives in an employment relationship conditional on taking the job. That is, if we could observe a worker for a long time, her type would be the average log wage she receives. Similarly, we define a firm's type to be the expected log wage that it pays to an employee conditional on hiring the worker, or equivalently the average log wage paid in a long time series.[2]

Beyond proposing this measure of sorting, our paper makes three main contributions. First, we develop simple structural models to use as laboratories for measuring sorting. In particular, we ask whether our measure of the correlation between types captures an empirically infeasible but intuitive notion of sorting in those models. We show that in a particular statistical model, our measure of the correlation between types is identical to AKM's measure of correlation. We then develop a search model with both ex ante heterogeneity and idiosyncratic productivity shocks. We show that the two measures of correlation are quantitatively similar, although there are limited situations where one performs better than the other. Fi-

[1] In addition to the original study on French data by AKM, see Abowd, Creecy, and Kramarz (2002) for Washington State, Iranzo, Schivardi, and Tosetti (2008) for Italy, Gruetter and Lalive (2009) for Austria, Card, Heining, and Kline (2013) for Germany, Bagger, Sørensen, and Vejlin (2013) and Bagger, Fontaine, Postel-Vinay, and Robin (2014) for Denmark, and Lopes de Melo, 2018 for Brazil, among others. Recent papers emphasize bias in the OLS estimates of the AKM model and find a larger correlation; see especially Kline, Saggio, and Sølvsten (2019) and Bonhomme, Lamadon, and Manresa (2019).

[2] Our definition of firm type is close to Christensen, Lentz, Mortensen, Neumann, and Werwatz (2005), who define a firm's type to be equal to the average wage (in levels rather than logs) it pays. It is worth noting that both AKM's and our definition of firm type is consistent with high type firms being either high or low productivity firms, for the reasons discussed in Eeckhout and Kircher (2011).

nally, we develop a discrete choice model with an idiosyncratic amenity shock and find that our measure of the correlation between types captures an intuitive notion of sorting, while the AKM correlation delivers nonsense. We conclude that the correlation between types is a promising reduced-form measure of sorting in a variety of standard economic environments.

Our second contribution is to develop an estimator of the correlation between types. Intuitively, our approach is to directly estimate the variance-covariance matrix of matched types using moment conditions that do not impose any functional form on the underlying type distributions. Our estimator is motivated by the properties of real-world data sets that have a large number of workers and firms but few independent observations for each worker and firm. Our key identifying assumption is that for each worker, we have two or more observations of the actual wage received and that these observations are independently and identically distributed conditional on the worker's type, and symmetrically for firms. Our estimated correlation then pertains to the sample of workers and firms for which we have these two observations. We prove that our estimator is consistent in the limit as the number of workers and firms goes to infinity, even if we only observe a small number of independent observations for each worker and firm. We also use model-generated data to evaluate the behavior of the estimator in small data sets. We find that both the bias and variance of the estimator are generally small even with only 2,500 workers and 500 firms.

Our third contribution is to estimate the correlation in real-world data. Our primary data set captures the universe of private sector workers in Austria from 1986 to 2018. We first measure the correlation between types using annual wage data and find it is about 0.67 for men and 0.62 for women. However, we recognize that annual wage observations might not be independent conditional on type, particularly for workers who do not switch employers. This means these estimates may be inconsistent. To construct conditionally independent observations, we rely on economic theory. First, we average all our wage data to the worker-firm match level. In simple search models without on-the-job search, such as Shimer and Smith (2000), wages in any two employment relationships are independent conditional on the worker's type. This suggests that we can use match-level data on all workers who have at least two jobs and all firms that have at least two employees in our data set. Second, in a more realistic search model with on-the-job search, such as Burdett and Mortensen (1998) and Postel-Vinay and Robin (2002), the wages in any two jobs that are separated by an unemployment spell are independent conditional on the worker's type. We define the time between registered unemployment spells as an employment spell and further trim the data to keep only the longest job during each employment spell for each worker. Our numerical results depend on which data set we use, and our preferred estimates use the last approach, with one observation per employment spell per worker. Using this data set, we estimate that the correlation between worker and firm types is 0.47 for men and 0.43 for women.

We view these as lower bounds on the true correlation because our approach ignores the fact that workers and firms change over time and firms are collections of heterogeneous jobs. We illustrate this by incorporate time-varying observable characteristics for both workers and firms. For example, we let firms have different types when matched with workers with different education levels. This raises the estimated correlation to 0.52 for men and 0.51 for women. We get similar results when we allow for variation in both workers' and firms' types depending on whether the job is blue or white collar, and when we allow for variation in workers' types depending on the firm's industry.

Our approach is also amenable to estimation using short time series. In particular, we can estimate the correlation between worker and firm types using only a single year's data, which further mitigates issues of time-varying unobserved heterogeneity. Using only data on workers who lose a job, become unemployed, and find a new job within a year, we find a steady increase in sorting among men since 1986 but a decrease in sorting among women, particularly during the eight years after 2001. At the end of our sample, the correlation between men's types and their employer's types was 0.55, while it was only 0.38 for women.

The last section of our paper explores the difference between our results and estimates of the AKM correlation. We verify that OLS estimates of the AKM correlation are small in our data, 0.12 for men and 0.07 for women. There are two possible explanations for this difference. Either our correlation is much bigger than the true AKM correlation or the true AKM correlation is much bigger than the OLS estimate of it. Our structural models establish the first possibility and indicate that in these cases, our correlation is preferred. On the other hand, recent papers propose bias corrections and alternative estimators of the AKM correlation (e.g., Andrews, Gill, Schank, and Upward, 2008; Kline, Saggio, and Sølvsten, 2019; Bonhomme, Lamadon, and Manresa, 2019).[3] We obtain very different estimates of the AKM correlation depending on which estimator we use. In some cases, the bias-corrected AKM estimates are very similar to the OLS estimates, while in other cases they are much closer to our correlation. In our view, this uncertainty about the true AKM correlation further supports our measure of correlation and our estimator.

The goal of this paper is to develop and estimate a measure of sorting that is sensible in a variety of structural models. An alternative approach would be to estimate a structural model of sorting, as in Lise, Meghir, and Robin (2016), Hagedorn, Law, and Manovskii (2017), Lopes de Melo (2018), or Bagger and Lentz (forthcoming). There are advantages to

[3]Jochmans and Weidner (2019) show that the bias depends on the worker-firm network structure and offer bounds on the bias as a function of the degree of connectivity of the network. Since Andrews, Gill, Schank, and Upward (2008) and Kline, Saggio, and Sølvsten (2019) offer exact formulas, we use their approach to the bias correction.

each approach. On the one hand, the estimates in those papers impose more structure on the data, which may drive the results. On the other hand, those papers are able to address issues that go beyond the scope of this paper, where we focus exclusively on measuring sorting.

The remainder of the paper proceeds as follows. Section 2 defines our measure of sorting, the correlation between worker and firm types. In Section 3, we use several models as laboratories to study how our measure of sorting compares to the AKM measure of correlation as well as to an intuitive measure of sorting. We propose an estimator in Section 4 and implement it on an Austrian data set, which we describe in Section 5. Section 6 gives our main empirical results, showing that the correlation between worker and firm types lies between 0.4 and 0.6. Section 7 compares our results to estimates of the AKM correlation using OLS and other approaches. Section 8 concludes.

2 Measuring Sorting in Theory

2.1 The Economy

We consider the cross-section of an economy with a fixed measure of workers employed by a fixed measure of firms. Workers and firms are distinguished by their characteristics, $x \in X$ and $y \in Y$, respectively. Let $F(x)$ denote the distribution of workers' characteristics. Let $\Phi_x(y)$ denote the distribution of the employer's characteristics conditional on the worker's characteristics. For now, we treat F and Φ_x as primitives; however, we think of these distributions as being generated by a structural dynamic model such as Burdett and Mortensen (1998), Shimer and Smith (2000), or Postel-Vinay and Robin (2002). In such a model, differences in Φ across x might reflect the fact that different workers find or accept different jobs with different probabilities or that they have different patterns of job-to-job mobility.

Define $G(y) \equiv \int_X \Phi_x(y) dF(x)$ to be the unconditional distribution of the characteristics of *jobs* in the economy. This is distinct from the distribution of the characteristics of firms to the extent that firms with different characteristics employ different numbers of workers. We also define $\Psi_y(x)$ to be the conditional distribution of the worker's characteristics given the firm's characteristics. By Bayes rule, we have $\Phi_x(y)F(x) \equiv \Psi_y(x)G(y)$ for all x and y.

A worker with characteristics x matched to a firm with characteristics y earns a wage that depends on both characteristics and on a shock. Let $w(x, y, z)$ denote the z^{th} quantile of the log wage distribution in an (x, y) match.[4] In competitive environments, the wage depends only on x, but the presence of search frictions, compensating differentials, or measurement

[4]This is the distribution of log wages in matches that actually occur. If x and y reject some wage draws or turnover is higher following some wage draws, that is reflected in the matching distributions Φ and Ψ, not in the log wage distribution.

error in x all imply that the wage may be correlated with y and other features (such as alternative job opportunities) captured by z.

2.2 A New Measure of Sorting

We are interested in measuring the correlation between matched workers and firms in an employment relationship. To do this, we need a cardinal, unidimensional measure of workers' and firms' types. Workers' and firms' characteristics x and y may be vector-valued and in any case do not have even an ordinal interpretation.[5] We therefore propose measuring the correlation between the expected log wage received by a worker conditional on her characteristics and the expected log wage paid by her employer conditional on its characteristics. That is, we are interested in understanding whether high wage workers typically work in high wage firms.

For now we assume that we know the distributions F, Φ, G, and Ψ, as well as the wage function w, and define our measure of sorting. In Section 3 we use structural models to show that our proposed measure captures the extent of sorting in model economies. Of course, in real world data sets we do not observe F, Φ, G, and Ψ or the wage function w, and so we explain in Section 4 how to estimate the correlation between types using the limited data that are available. Let

$$\lambda(x) \equiv \int_Y \int_0^1 w(x,y,z)\, dz\, d\Phi_x(y) \tag{1}$$

$$\text{and } \mu(y) \equiv \int_X \int_0^1 w(x,y,z)\, dz\, d\Psi_y(x) \tag{2}$$

denote the expected log wage received by a worker with characteristics x and the expected log wage paid by a firm with characteristics y, respectively. From now on, we identify a worker by her expected log wage and call $\lambda(x)$ her type. Symmetrically, we identify a firm by the expected log wage it pays and call $\mu(y)$ its type.

Our object of interest is the correlation between the type of a worker and the type of her job in the cross-section of matches at a point in time,

$$\rho \equiv \frac{c}{\sigma_\lambda \sigma_\mu}, \tag{3}$$

[5] Lindenlaub and Postel-Vinay (2017) study a model with multidimensional characteristics and examine the conditions under which there is positively assortative matching dimension-by-dimension. It is impossible to measure this stronger notion of sorting using wage data alone.

where

$$\bar{w} \equiv \int_X \int_Y \int_0^1 w(x, y, z)\, dz\, d\Phi_x(y)\, dF(x) = \int_X \lambda(x) dF(x) = \int_Y \mu(y) dG(y) \tag{4}$$

is the mean log wage, also equal to both the mean worker type and the mean job type;

$$\sigma_\lambda \equiv \sqrt{\int_X (\lambda(x) - \bar{w})^2\, dF(x)} \;\text{ and }\; \sigma_\mu \equiv \sqrt{\int_Y (\mu(y) - \bar{w})^2\, dG(y)} \tag{5}$$

are the cross-sectional standard deviations of worker types and job types; and

$$c \equiv \int_X \int_Y (\lambda(x) - \bar{w})(\mu(y) - \bar{w})\, d\Phi_x(y)\, dF(x) \tag{6}$$

is the covariance between worker and job types in an employment relationship.[6] We assume throughout that all of these first and second moments are finite.

We highlight the special case where $\Phi_x(y) = G(y)$ for all x and y. This means that each worker is equally likely to work in every job. In this case, we can rewrite the covariance as

$$c \equiv \int_X (\lambda(x) - \bar{w}) \left(\int_Y \mu(y)\, dG(y) - \bar{w} \right) dF(x).$$

The term in the inner parenthesis is zero by the definition of $\bar{w}$, hence the covariance is zero. Since the variance of worker and firm types is still generally positive, the correlation between types is zero. More generally, the sign of the correlation depends on whether high wage workers are particularly likely to work at high wage firms.

2.3 The AKM Measure of Sorting

We contrast our measure of sorting with a common alternative due to Abowd, Kramarz, and Margolis (1999) (hereafter AKM). The authors' starting point is the assumption that the log wage in a match between worker i with characteristics x_i and firm j with characteristics y_j is linear in the worker's and firm's fixed effects,

$$w(x_i, y_j, z) = \alpha_i + \psi_j + \eta \tag{7}$$

[6] Lopes de Melo (2018) shows that the correlation between a worker's wage and the wage of her coworkers is a useful moment in estimating his structural model. The corresponding covariance is $\int_Y \int_X \int_X (\lambda(x) - \bar{w})(\lambda(x') - \bar{w})\, d\Psi_y(x')\, d\Psi_y(x)\, dG(y)$.

where $\alpha_i = \alpha(x_i)$ is the worker fixed effect, $\psi_j = \psi(y_j)$ is the firm fixed effect, and $z \equiv \zeta_{i,j}(\eta)$ is an error term where the distribution $\zeta_{i,j}$ has mean zero for all (i, j) pairs.[7] An important goal in that research agenda is measuring the correlation between α_i and ψ_j among matched worker-firm pairs (i, j), which we denote ρ_{AKM}.

If the linear wage equation (7) is correctly specified and we had infinitely much data for each pair (x, y), we could recover $\alpha(x)$ and $\psi(y)$ by integrating over the mean zero error term. This gives us a system of linear equations,

$$\int_0^1 w(x, y, z) dz = \alpha(x) + \psi(y),$$

which determine α and ψ up to an additive constant. If equation (7) is misspecified, this equation cannot hold for all types. Still, we define the fixed effects as the solution to the following moment conditions,

$$\alpha(x_i) = \int_Y \int_0^1 (w(x_i, y, z) - \psi(y))\ dz\, d\Phi_{x_i}(y) \tag{8}$$

$$\psi(y_j) = \int_X \int_0^1 (w(x, y_j, z) - \alpha(x))\ dz\, d\Psi_{y_j}(x), \tag{9}$$

which is equivalent to running OLS on data containing all matched pairs. As is well known, these moment conditions uniquely define α and ψ up to an additive constant if (and only if) there is no way to partition the workers and firms into two nonempty sets A and B such that workers and firms in set A (B) only match with firms and workers in set A (B).

We then compute the AKM correlation in the matched pairs as

$$\rho_{AKM} = \frac{c_{AKM}}{\sigma_\alpha \sigma_\psi} \tag{10}$$

where

$$\bar{\alpha} \equiv \int_X \alpha(x)\, dF(x), \quad \bar{\psi} \equiv \int_Y \psi(y)\, dG(y), \tag{11}$$

$$\sigma_\alpha \equiv \sqrt{\int_X (\alpha(x) - \bar{\alpha})^2\, dF(x)}, \quad \sigma_\psi \equiv \sqrt{\int_Y (\psi(y) - \bar{\psi})^2\, dG(y)}, \tag{12}$$

$$c_{AKM} \equiv \int_X \int_Y (\alpha(x) - \bar{\alpha})(\psi(y) - \bar{\psi})\, d\Phi_x(y)\, dF(x). \tag{13}$$

We do not focus here on how to estimate ρ_{AKM}; there are well-known statistical problem

[7] Abowd, Kramarz, and Margolis (1999) also allow for time-varying observable worker and firm characteristics. We suppress those for expositional simplicity.

with the fixed effects estimator, often called "limited mobility bias." Instead, we assume that we know the distributions F, Φ, G, and Ψ, as well as the wage function w, and are interested in how ρ_{AKM} behaves in this idealized environment. To explore this, we turn next to some structural models.

3 Models as Laboratories for Measuring Correlation

This section develops simple economic envirornments to explore how the two proposed measures of sorting, ρ and ρ_{AKM}, behave in structural models where we have a strong sense of whether there is sorting. We start with a model in which the AKM wage equation is correctly specified. We then turn to a search model based on Shimer and Smith (2000), extended to include match productivity shocks (Goussé, Jacquemet, and Robin, 2017) and finally look at a discrete choice model, as in Card, Cardoso, Heining, and Kline (2018).

3.1 AKM is Correctly Specified

We start with an important special case in which the AKM correlation and our correlation coincide. Assume the AKM wage equation (7) is correctly specified with $\alpha_i = x_i$ for all i and $\psi_j = y_j$ for all j. Also assume that the joint density of (x, y) (and hence (α, ψ)) in matched worker-firm pairs has linear conditional expectations, that is, $\mathbb{E}_x(y)$ is linear in x and $\mathbb{E}_y(x)$ is linear in y.[8] Let

$$\begin{pmatrix} \sigma_x^2 & c_{AKM} \\ c_{AKM} & \sigma_y^2 \end{pmatrix}$$

be its variance-covariance matrix so $\rho_{AKM} = \frac{c_{AKM}}{\sigma_x \sigma_y}$. In this case, ρ and ρ_{AKM} have the same magnitude:

Proposition 1 *Assume that the joint distribution of $\alpha = x$ and $\psi = y$ is such that the conditional expected values are linear, that is, $\mathbb{E}_x(y)$ is linear in x and $\mathbb{E}_y(x)$ is linear in y. Let $\rho_{\text{AKM}} \in (-1, 1)$ denote the correlation between x and y. Then λ and μ are linear transformations of x and y with standard deviations $\sigma_\lambda = |\sigma_x + c_{\text{AKM}}/\sigma_x|$ and $\sigma_\mu = |\sigma_y +$*

[8] If the joint distribution of x and y is *elliptical*, i.e. the associated density function ξ can be expressed as

$$\xi(x, y) = \tilde{\xi}\left(\frac{(x-\bar{x})^2}{\sigma_x^2} - \frac{2\rho_{AKM}(x-\bar{x})(y-\bar{y})}{\sigma_x \sigma_y} + \frac{(y-\bar{y})^2}{\sigma_y^2}\right),$$

then conditional expectations are linear. The bivariate normal and bivariate t distributions are elliptical.

$c_{\mathrm{AKM}}/\sigma_y|$. *Let* ρ *denote the correlation and* c *denote the covariance of* λ *and* μ*. Then*

$$\min\{\sigma_x^2, \sigma_y^2\} \gtreqless -c_{\mathrm{AKM}} \Rightarrow \begin{cases} \rho = \rho_{\mathrm{AKM}} \text{ and } \min\{\sigma_\lambda^2, \sigma_\mu^2\} > c \\ \rho \text{ is undefined} \\ \rho = -\rho_{\mathrm{AKM}} \text{ and } \min\{\sigma_\lambda^2, \sigma_\mu^2\} < c. \end{cases}$$

The proof in Appendix A.1 shows that with linear conditional expectations, equations (1) and (7) imply that $\lambda(x_i)$ is a linear transformation of x_i

$$\lambda(x_i) = \int_Y (x_i + y) d\Phi_{x_i}(y) = \kappa_0 + \left(1 + \frac{c_{AKM}}{\sigma_x^2}\right) x_i,$$

for some constant κ_0. Symmetrically, equations (2) and (7) imply $\mu(y_j)$ is a linear transformation of y_j,

$$\mu(y_j) = \int_X (x + y_j) d\Psi_{y_j}(x) = \theta_0 + \left(1 + \frac{c_{AKM}}{\sigma_y^2}\right) y_j,$$

for some constant θ_0. The magnitude of the correlation coefficient between two random variables is unaffected by a linear transformation, though it may change sign if one of the transformations is decreasing, i.e. either $c_{AKM} < \sigma_x^2$ or $c_{AKM} < \sigma_y^2$. The proof shows that these conditions are equivalent to $c > \sigma_\lambda^2$ and $c > \sigma_\mu^2$, respectively.

Since $\rho_{AKM} = \frac{c_{AKM}}{\sigma_x \sigma_y} > -1$, whenever we have $\sigma_x = \sigma_y$, it is the case that $\sigma_x^2 = \sigma_y^2 > -c_{AKM}$. Proposition 1 then implies $\rho = \rho_{AKM}$. This illustrates how to construct an economy where ρ takes any value in the interval $(-1, 1)$.

We view this statistical model as an important benchmark case. Our approach defines a worker's type λ to be equal to her expected log wage and a firm's type μ to be equal to the expected log wage it pays. AKM define the units of types to be that which boosts the expected log wage by a unit *holding fixed the partner's type*. While these two measures are distinct, Proposition 1 establishes conditions under the correlation between the two measures are equal. Any structural model with an equilibrium satisfying the above properties would feature the same magnitude of ρ and ρ_{AKM}.

A natural hypothesis is that by perturbing this example, either by changing the wage function or by changing the distributional assumption on matches, it is possible to construct examples where $\rho \gtrless \rho_{AKM}$. It appears, however, that this is not the case. In the structural models below, we always find that $\rho > \rho_{AKM}$. More interestingly, extensive numerical simulations suggests that $\rho \geq \rho_{AKM}$ in any economy,[9] although a proof of this conjecture

[9]The simulations assume there are m types of workers, $x = 1, \ldots, m$, and n types of jobs, $y = 1, \ldots, n$, and allow for an arbitrary joint distribution of matched workers and jobs, as well as an arbitrary average

eludes us. Once we recognize that the two correlation measures are different, the interesting question is whether either correlation captures an economically reasonable notion of sorting in realistic environments. In an effort to answer that question, we next use a variety of structural models as laboratories for evaluating the two measures of sorting.

3.2 Two-Sided Search Model with Match-Specific Shocks

We next examine a search model with two-sided heterogeneity (Shimer and Smith, 2000) and match-specific heterogeneity (Goussé, Jacquemet, and Robin, 2017). The match-specific productivity shocks ensure that any worker and firm have a positive probability of matching, but different matches use a different threshold for the idiosyncratic shock. It also implies that the wage is not pinned down by the worker and firm characteristics but depends on the idiosyncratic shock as well.

The model is formulated in continuous time. There is measure 1 of risk-neutral workers and measure 1 of risk-neutral firms. Everyone discounts the future at rate r. Each worker has characteristic x, distributed in the population according $\tilde{F}(x)$. Similarly, each firm has characteristic y, distributed according to $\tilde{G}(y)$.[10] Workers can be either unmatched or matched to one firm; likewise, firms can be either unmatched or matched to one worker, so a firm and a job are identical here.

Search is random and only unmatched firms and workers can search. Let $u(x)$ be the unemployment rate among workers with characteristic x and $v(y)$ vacancy rate among firms with characteristic y. A worker meets a vacancy at the rate θ and the firm characteristic is randomly drawn from the distribution $\tilde{G}$. Since there are equal measures of workers and firms, this means that a firm meets a worker at the same rate θ and the worker characteristic is randomly drawn from the distribution $\tilde{F}$. If the worker or firm is matched, it is as if the meeting never happened. If both are unmatched, with probability $u(x)v(y)$, the pair draws a match-specific productivity $z \geq 0$ from a cumulative distribution ζ and then decide whether to match and produce flow $zH(x, y)$. Match-specific productivity is independently and identically distributed across matches and is fixed for the duration of the match. They split the surplus according to Nash bargaining, with worker's bargaining power $\gamma \in (0, 1)$. Assume $H(x, y)$ is strictly positive for almost all x and y. Matches randomly end at rate δ, leaving the worker unemployed and the job vacant.

Let $U(x)$ and $V(y)$ denote be the value of being an unemployed worker and a vacant firm,

wage in an (x, y) match. We then use equations (1) and (2) to construct $(\lambda(x), \mu(y))$ and equations (8) and (9) to construct $(\alpha(x), \psi(y))$. We then compute ρ and ρ_{AKM} using equations (3) and (10) and compare their magnitudes.

[10] In Section 2.1, we use F and G to denote the distribution of characteristics of employed workers and filled jobs. We use tildes to distinguish the population distributions $\tilde{F}$ and $\tilde{G}$ from these.

respectively. The surplus of a match between x and y is $S(x,y,z) = zH(x,y) - rU(x) - rV(y)$. The decision to match is described by a threshold rule: a match is formed if $z \geq \bar{z}(x,y)$ where $\bar{z}(x,y)$ is such that $S(x,y,\bar{z}(x,y)) = 0$. In Appendix A.2, we develop a system of equations that fully describes the equilibrium, while here we focus on the behavior of wages. Nash bargaining implies that

$$w(x,y,z) = \gamma(zH(x,y) - rU(x) - rV(y)) + rU(x),$$

and hence the expectation of the log wage in an (x,y) match is

$$w(x,y) = \frac{1}{1-\zeta(\bar{z}(x,y))} \int_{z \geq \bar{z}(x,y)} \log\left(\gamma\big(zH(x,y) - rU(x) - rV(y)\big) + rU(x)\right) d\zeta(z).$$

If the distribution of match-specific productivity is exponential, we prove in Appendix A.2 that the expected log wage in a match (x,y) is monotone in $H(x,y)$ for given x. That is, if higher y matches are more productive, they also pay higher expected log wages conditional on matching. We obtain a similar result numerically when the match-specific productivity distribution is Pareto with a sufficiently high variance. In contrast, in Shimer and Smith (2000), a given x's wage is maximized at some value of y, typically an interior point, even if H is strictly increasing. We find numerically that is also the case for a Pareto match-specific productivity distribution when the variance is small.

Another difference from Shimer and Smith (2000) is that if match-specific productivity is unbounded above, almost all matches (x,y) are created with strictly positive probability. With enough complementarity in the production function, a worker with the lowest characteristic would never match with the highest characteristic firm in Shimer and Smith (2000). In this model, we observe such a match if the match-specific productivity is high enough. Still, high draws are rare and therefore we observe sorting based on characteristics.

We solve the model with a discrete number of characteristics n distributed uniformly on $X = Y = \{\frac{1-0.5}{n}, \frac{2-0.5}{n}, \ldots \frac{n-0.5}{n}\}$. We use the CES production function

$$H(x,y) = (ax^{\frac{\xi-1}{\xi}} + (1-a)y^{\frac{\xi-1}{\xi}})^{\frac{\xi}{\xi-1}},$$

where $\xi \geq 0$ is the elasticity of substitution and $a \in [0,1]$ is the worker's share in production. We assume that the distribution of match productivity shocks is Pareto, with some minimum value $\underline{z}$ and variance σ_z^2.[11] Our benchmark uses the following parameter values: $r = 1$, $\delta = 10$,

[11] We choose Pareto rather than the exponential distribution because it allows us to change the variance of the shocks. With the exponential distribution, doubling its parameter only doubles value functions but has no impact on the matching probabilities, unemployment rates and vacancy rates.

$\theta = 10^4$, $\gamma = 0.5$, $a = 0.5$, $\xi = 1$, $\sigma_z^2 = 0.1$, $\underline{z} = 1$, and $n = 500$.

With given model parameters, the model tells us exactly how often each worker matches with each firm as well as the average log wage in those matches. We can thus compute $\lambda(x)$ and $\mu(y)$ from equations (1) and (2), as well as $\alpha(x)$ and $\psi(y)$ from equations (8) and (9). We then find ρ and ρ_{AKM} from equations (3) and (10). Since we know the data generating process, we can and deliberately choose to ignore questions about estimation in this section.

We are interested in exploring how our measures of sorting change as we vary four key model parameters: the meeting rate θ, the bargaining power γ, the variance of the match-specific shocks σ_z^2, and the elasticity of substitution in the production function ξ. In each experiment, we compare ρ and ρ_{AKM} with the correlation between x and y, $\rho_{x,y}$, an intuitive measure of the extent of sorting. Note that since x and y are not observable in real-world data, $\rho_{x,y}$ is not something that is generally feasible to compute, but it is possible in the model. Also note that the units of x and y are only defined through the assumption that characteristics have a discrete uniform distribution. Without that, we could do a nonlinear transformation of x and y and offset this through a change in the production function that would leave the real structure of the economy, and hence ρ and ρ_{AKM} unchanged, but change $\rho_{x,y}$. This degree of freedom implies that the correlation between x and y is instructive but not necessarily what we should be targeting.

Figure 1 shows results from these experiments. In the top left panel, we vary the meeting rate θ. When θ is low, it is difficult for workers to meet vacancies and thus they tend to accept any offer that they receive, conditional on a favorable match-specific shock. As match acceptance thresholds are low for everyone, we see little sorting. As $\theta \to \infty$, workers receive offers very quickly and become selective at which offer to accept. As a result, sorting increases. The correlation $\rho_{x,y}$ (red) is increasing in θ, and ρ (blue) and ρ_{AKM} (green) capture this pattern.[12]

In the second experiment, we change worker's bargaining power. As γ converges to 0 or 1, workers (or firms) are paid only their outside option regardless of whom they match with, hence sorting weakens. The correlation between x and y is hump-shaped (top right panel in Figure 1) and again ρ and ρ_{AKM} properly capture this nonlinear pattern.

In our third comparative static exercise, we change the variance of the match productivity shocks. As the variance increases, match-specific productivity, rather than worker and firm characteristics, plays a more important role and sorting becomes weaker. The bottom left

[12]Interestingly, the correlation does not converge to 1 as $\theta \to \infty$. This is because of the match-specific productivity shocks. With minimal search frictions, the economy resembles a discrete choice model (Section 3.3) where workers see multiple offers characterized by (y, z) and choose the one with the highest value. In this limit, workers match with firms with heterogeneous characteristics and hence the correlation between x, y is high but less than 1.

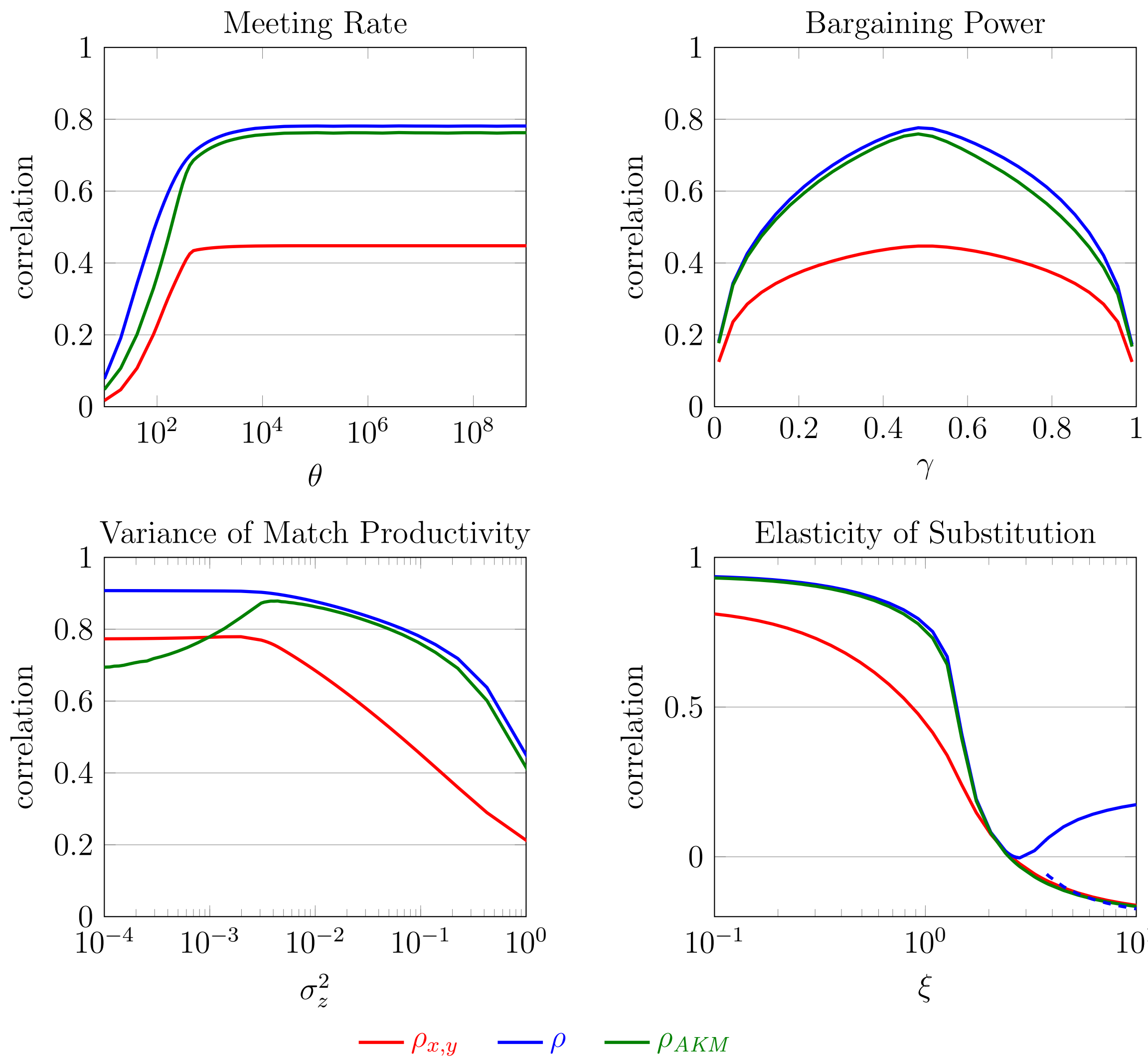

Figure 1: Comparative statics exercise in the two-sided search model extended to include match-specific productivity. The figures show the correlation between characteristics ($\rho_{x,y}$, red lines), types (ρ, blue lines), and AKM fixed effects (ρ_{AKM}, green lines) in the matched pairs for different parameter values. In each experiment, we keep all but one parameter at their benchmark values, $r = 1$, $\delta = 10$, $\theta = 10^4$, $\gamma = 0.5$, $a = 0.5$, $\xi = 1$, $\sigma_z^2 = 0.1$, $\underline{z} = 1$, $n = 500$, and indicate on the horizontal axis which parameter we are changing.

panel shows that ρ_{AKM} does not produce this pattern. Instead, when the variance is low, an increase in the variance raises the AKM correlation. This appears to reflect the fact that when the variance is low, the expected log wage is non-monotonic in characteristics and hence misspecification of the AKM wage equation is quantitatively serious; see Eeckhout and Kircher (2011), Lise, Meghir, and Robin (2016), Lopes de Melo (2018), and Bagger and Lentz (forthcoming). Thus we find a shortcoming of ρ_{AKM} when nonlinearities in the wage equation are important.

Finally, we turn to the elasticity of substitution in the production function. The production function is Leontief for $\xi = 0$ and linear for $\xi \to \infty$ and so sorting is less important as ξ increases, as shown in the bottom right panel of Figure 1. When ξ is small, there is positive sorting and all three measures perform well. But when ξ is sufficiently large, ρ remains positive even though $\rho_{x,y}$ and ρ_{AKM} are negative. This is because high x workers tend to work for low y firms, but low y firms actually pay higher wages on average. Hence, high wage workers work for high wage firms, a pattern which our measure captures. From wages alone, we are not able to say that the high paying firms are actually those with the low productivity and hence in fact there is negative sorting (Eeckhout and Kircher, 2011). Nevertheless, Proposition 1 proposes a test to detect the "sign flip": when $\min\{\sigma_\lambda^2, \sigma_\mu^2\} < c$, then ρ and ρ_{AKM} have the opposite sign in the case of linear conditional expectations. The blue dotted line shows the correlation adjusted for the flip sign, meaning that we plot $-\rho$ when this condition is satisfied.[13] The sign flip test generally picks up the negative correlation

These experiments illustrate that ρ does a good job of measuring sorting in the model economy, especially with the sign correction. Similarly, ρ_{AKM} performs well in situations where expected log wages are close to linear in worker and firm fixed effects, and hence the AKM wage equation is well-specified. When the wage equation is seriously misspecified, as is the case when σ_z^2 is small, ρ_{AKM} can be a misleading measure of the extent of sorting. Nevertheless, broadly speaking we are impressed by the similarity of ρ and ρ_{AKM} in Figure 1.

3.3 Discrete Choice Model

We next examine a discrete choice model, a partial equilibrium version of Card, Cardoso, Heining, and Kline (2018). There is a fixed number of workers indexed by i and a fixed number of firms indexed by j. Each worker is characterized by x distributed according to $F(x)$ and each firm is characterized by y distributed according to $\tilde{G}(y)$.[14] Each worker chooses which firm to work at in order to maximize his utility, the sum of the log wage

[13] This is the only case where we detect a sign flip in the two-sided search model.

[14] In equilibrium, different firms employ different numbers of workers and so the distribution of jobs G is distinct from the distribution of firms $\tilde{G}$.

$w(x, y)$ and an idiosyncratic amenity value ε. The log wage depends deterministically on the worker's and firm's characteristics, while the amenity is independently and identically distributed across workers and firms.

Worker i sees the amenity value he would get in each firm and chooses to work for the firm j that maximizes $w(x_i, y_j) + \varepsilon_{i,j}$. Each firm is willing to employ any worker who wants the job. We assume that the wage function w is bounded from above for each x and that amenities are drawn from an exponential distribution with mean (and hence standard deviation) s. This ensures that the workers' choice of y_j has a non-trivial limit when the number of firms goes to infinity (Malmberg and Hössler, 2014, Theorem 18.4). In the limit, the probability that a worker with characteristic x chooses a firm with characteristic y is

$$\Phi_x(y) \sim \exp\left(\frac{w(x, y)}{s}\right) d\tilde{G}(y).$$

Thus workers are more likely to choose high wage jobs, but the wage becomes less important when the standard deviation of the amenity shock, s, increases.

We again use this model as a laboratory to study performance of our correlation measure. We assume that the log wage is given by a quadratic function of the worker and firm characteristics:

$$w(x, y) = k_x x + k_y y - (\sqrt{k_{xx}}x - \sqrt{k_{yy}}y)^2$$

with k_{xx} and k_{yy} positive. Then the log wage of a worker with characteristic x is maximized at firm a firm with characteristic

$$y^*(x) = \frac{k_y + 2\sqrt{k_{xx}k_{yy}}x}{2k_{yy}}.$$

However, workers with characteristic x will not always choose to work at firms with characteristic $y^*(x)$ since their utility depends on the amenity value as well.

When the characteristics x and y are distributed normally, then the joint distribution of matched (x, y) pairs is normal and we obtain closed form expressions for types $\lambda(x)$ and $\mu(y)$, AKM fixed effects $\alpha(x)$ and $\psi(y)$, and the correlations ρ, ρ_{AKM}, and $\rho_{x,y}$. We present the key formulae in Appendix A.3. Here we discuss some numerical results.

For our benchmark, we parameterize the wage function with $k_x = 1$, $k_y = 0$ and $k_{xx} = k_{yy} = 1/a^2$, which implies is $w(x, y) = x - (x - y)^2/a$ and $y^*(x) = x$ for all x. We also assume that worker characteristics are distributed $N(m_x, \sigma_x)$ and firm characteristics are distributed $N(m_y, \sigma_y)$ with $m_x = m_y$ (the common mean is irrelevant) and $\sigma_x = \sigma_y = 1$. Finally, we set the standard deviation of the amenity distribution to $s = 1$ and set the wage cost of mismatch to $a = 1$ as well. We conduct several experiments by varying key model parameters, s, a, $m_x - m_y$, and σ_x, and measuring how sorting changes.[15]

Again, with given model parameters, the model tells us the probability that each worker matches with each firm and so we can compute $\rho_{x,y}$. Then using the wage equation, we find $\lambda(x)$ and $\mu(y)$ from equations (1) and (2), as well as $\alpha(x)$ and $\psi(y)$ from equations (8) and (9). We then compute ρ and ρ_{AKM} from equations (3) and (10). Again, we ignore questions about estimation in this section.

The top left panel of Figure 2 shows how the standard deviation of the amenity shock affects sorting. When the standard deviation is zero, the amenity does not play any role in workers' decisions and each worker x chooses a firm with characteristic $y^*(x)$. As a result, firms with characteristic $y^*(x)$ employ only workers with characteristic x, and the correlation $\rho_{x,y}$ is one. As the standard deviation increases, sorting weakens and $\rho_{x,y}$ declines to zero (red line). Our correlation measure ρ exhibits the same pattern, monotonically declining from perfect sorting to zero as amenities become more important (blue line), albeit a bit faster than the decline in correlation $\rho_{x,y}$.

The AKM correlation ρ_{AKM} follows a very different pattern. In the limit as s converges to zero, AKM attributes all the wage variation to the worker fixed effects, generating a zero correlation between worker and firm effects. More generally, we prove that the sign of $\lim_{s\to 0} \rho_{AKM}$ is the same as the sign of $m_x - m_y$, even though there is always perfect sorting between x and y (and between λ and μ) in this limit. As the standard deviation increases, ρ_{AKM} becomes negative but remains close to zero. It then flips signs at $s > 2$. In short, ρ_{AKM} does not capture the extent sorting in the underlying economy.

Interestingly, the "sign flip" test from Proposition 1 picks up the incorrect sign of ρ_{AKM} when $s < 2$. In this region (and only in this region), we find that both $\sigma_\lambda^2 > c > \sigma_\mu^2$ and $\sigma_\alpha^2 > -c_{AKM} > \sigma_\psi^2$. If the assumptions of Proposition 1 were satisfied, this would be consistent with $\rho = -\rho_{AKM}$. The assumptions are not satisfied, but the conclusion that the two correlations have different signs still holds. We stress that it is not the sign of ρ that is misleading in this example, but instead the sign of ρ_{AKM}.

The results are similar when we change the wage function by varying a. As $a \to 0$, the penalty from not taking the right job goes to infinity and hence we get perfect sorting with workers of type x choosing $y^*(x)$. Both ρ and $\rho_{x,y}$ are equal to 1. As a increases, workers' wages are less sensitive to their employer and so sorting weakens and eventually disappears as $a \to \infty$. Interestingly, the difference between ρ and $\rho_{x,y}$ is negligible for all values of a.

Once again, ρ_{AKM} fails to capture sorting, especially in the region where it is nearly perfect. The correlation ρ_{AKM} is again negative for $a < 2$, at which point it jumps up to

[15] Other changes in the wage function, i.e. in the parameters k_x, k_y, k_{xx}, and k_{yy}, are isomorphic to changes in the distribution of workers' and firms' characteristics, and so we omit that from our analysis.

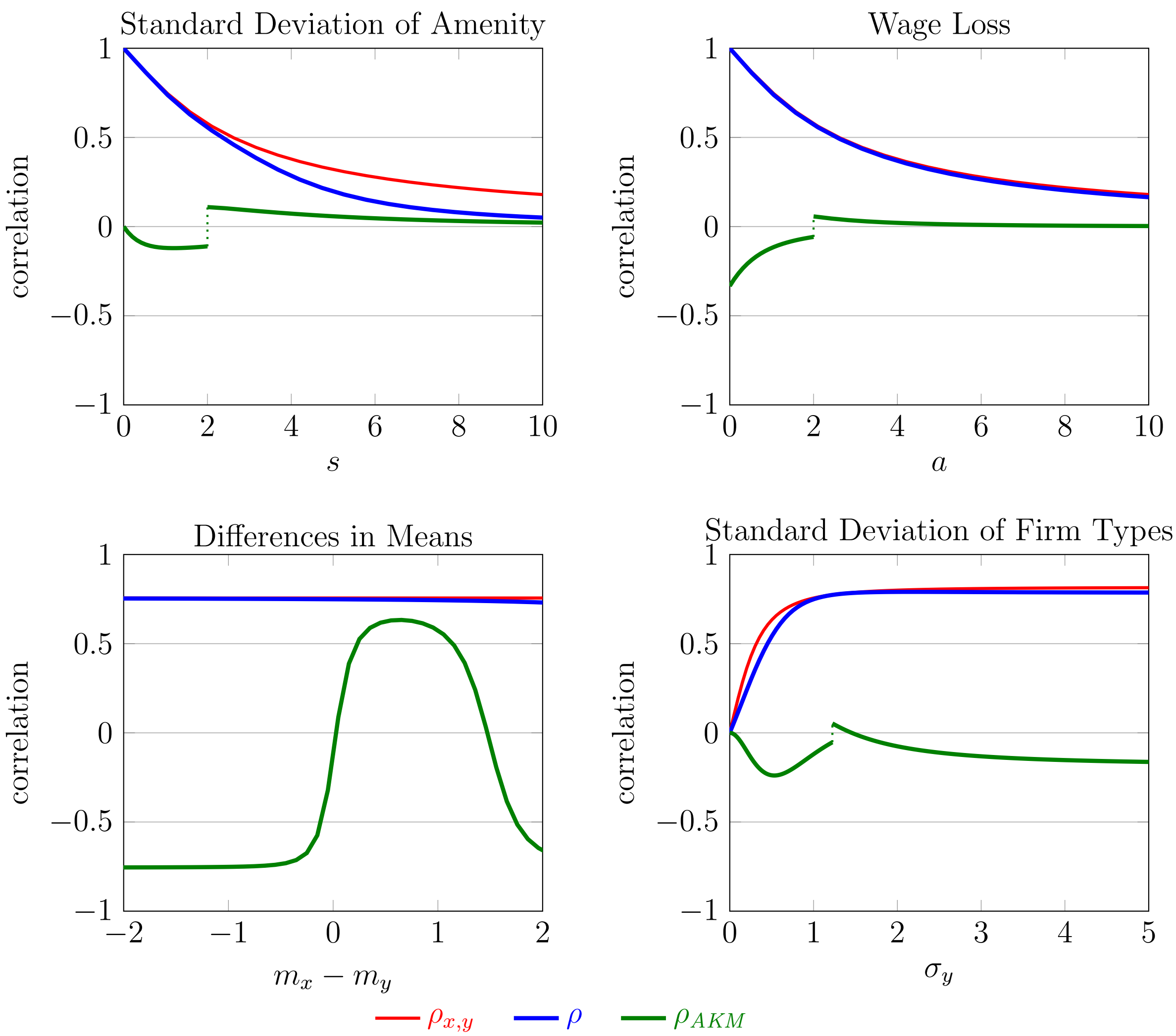

Figure 2: Comparative statics exercise in the discrete choice model. The figures show the correlation between characteristics ($\rho_{x,y}$, red lines), types (ρ, blue lines), and AKM fixed effects (ρ_{AKM}, green lines) in the matched pairs for different parameter values. In each experiment, we keep all but one parameter at their benchmark values, $a = 1$, $s = 1$, $m_x = 0$, $m_y = 0$, $\sigma_x = 1$, and $\sigma_y = 1$, and depict on the horizontal axis which parameter we are changing.

a positive number and then declines to zero. We can again prove that $\sigma_\lambda^2 > c > \sigma_\mu^2$ and $\sigma_\alpha^2 > -c_{AKM} > \sigma_\psi^2$ if and only if $a < 2$, and so the "sign flip" test indicates that ρ and ρ_{AKM} have opposite signs in this region. And again, it is the sign of ρ_{AKM} that is misleading.

In the next experiment, we vary the difference in means $m_x - m_y$. The extent of sorting as measured by $\rho_{x,y}$ does not change. This is because increasing the mean of y relative to x induces workers to choose higher y firms, but this change is the same for all x, leaving the correlation unaffected. The correlation ρ does depend on the difference in means but the variation is quantitatively negligible. On the other hand, ρ_{AKM} varies significantly with $m_x - m_y$ and hence fails as a measure of sorting.

In the last experiment, we increase the variance of firm types while keeping the variance of worker types unchanged. In the extreme case of $\sigma_y^2 = 0$, all firms have the same type and so workers only pay attention to amenities; there is no sorting in the limit without firm variance. As the variance increases, workers pay more attention to firm characteristics and so sorting increases. However, increasing the variance too much beyond the variance of worker types will not bring any additional improvement of sorting since few workers choose extreme firm types. We indeed see that $\rho_{x,y}$ starts at zero when $\sigma_y^2 = 0$, and then increases steeply until around $\sigma_y^2 = 1$, after which it flattens. The correlation ρ follows the same pattern, while ρ_{AKM} again fails to capture changes in sorting.

To summarize, in the discrete choice model, ρ properly captures sorting patterns and ρ_{AKM} does not. The most striking finding is that ρ_{AKM} is zero when there are almost no amenities ($s \to 0$) and is negative when firms are critical for wages ($a \to 0$), even though sorting is almost perfect in each of these limits. Intuitively, nonmonotonicities in the structural wage equation mean that the linear AKM wage equation (7) is misspecified. Even thought it can potentially be a useful first order approximation, our calculations reveal that in this case ρ_{AKM} is a poor measure of sorting.

4 An Estimator of the Measure of Sorting

Consider a data generating process, for example a structural model, which determines who matches with whom as well as the wage and duration of each match. If we observed many conditionally independent matches for each worker and firm, we could accurately measure λ and μ for everyone and hence directly measure ρ. Unfortunately, in practice we have very few observations for most workers and firms. This section proposes a strategy for estimating ρ in realistic data sets. We start by defining a statistical model which encompasses the structural models in Section 3. We then propose an estimator and prove it is consistent in the statistical model. Finally we examine small sample properties of the estimator by looking at artificial data sets generated by our structural models.

4.1 A Dynamic Statistical Model

Our starting point is to imagine a dynamic economy which embeds the snapshot we described in Section 2.1. To start, we suppose there is a finite set of possible characteristic X of workers and Y of firms and one or more workers and firms with each characteristic. Let I_x denote the number of workers with characteristic x and $I \equiv \sum_{x \in X} I_x$ denote the total number of workers. Similarly, let J_y denote the number of firms with characteristic y and $J \equiv \sum_{y \in Y} J_y$ the total number of firms. To establish consistency of our estimator, we later replicate this economy so there are τI_x workers with characteristic x and τJ_y firms with characteristic y for some positive integer τ. We are interested in constructing an estimator of the variance-covariance matrix of matched pairs that is consistent in the limit as τ goes to infinity.

A worker's or firm's characteristic determines the probability of matching with every other firm and worker, the wage in each match, how long each match lasts, and how long we observe the worker or firm in the data set. More precisely, a typical worker i with characteristic x_i has $M_i \in \{\underline{M}, \dots, \bar{M}\}$ matches indexed by $m = 1, \dots, M_i$. Let $w^w_{i,m}$ denote the average log wage in i's m^{th} match, $t^w_{i,m}$ denote the duration of the match, and $y_{i,m}$ denote the firm characteristic for that match. We assume that the worker's characteristic determines the distribution of M_i as well as the joint distribution of $\{w^w_{i,m}, t^w_{i,m}, y_{i,m}\}_{m=1}^{M_i}$. When a worker matches with a firm with characteristic y, there is some unspecified probability of matching with each such firm. For example, a worker may draw randomly with or without recall. We let $\mathbf{j}_{i,m}$ denote the identity of the employer. It will be convenient to define $T^w_i \equiv \sum_{m=1}^{M_i} t^w_{i,m}$, the total time that we observe worker i employed, and denote its expected value conditional on the worker's characteristic by $\bar{T}^w_{x_i}$. We assume throughout that T^w_i has a finite upper bound and $\bar{M}$, the maximum number of matches a worker can have, is also finite.

Symmetrically, a typical firm j with characteristic y_j has $N_j \in \{\underline{N}, \dots, \bar{N}\}$ matches indexed by $n = 1, \dots, N_j$. Let $w^f_{j,n}$ denote the average log wage in j's n^{th} match, $t^f_{j,n}$ denote the duration of the match, and $x_{j,n}$ denote the worker characteristic for that match. Again, the firm's characteristic determines the distribution of N_j and the joint distribution of $\{w^f_{j,n}, t^f_{j,n}, x_{j,n}\}_{n=1}^{N_j}$. Let $\mathbf{i}_{j,n}$ denote the identity of the worker in j's n^{th} match. Again, we define $T^f_j \equiv \sum_{n=1}^{N_j} t^f_{j,n}$, the total time that firm j employs workers, and denote its expected value conditional on the firm's characteristic as $\bar{T}^f_{y_j}$. We again assume T^f_j has a finite upper bound and $\bar{N}$ is finite.

Worker and firm observations are necessarily linked. Suppose firm j employs worker i in her m^{th} match, i.e. $j = \mathbf{j}_{i,m}$. We let $\mathbf{n}_{i,m}$ denote the firm's corresponding match number.

Symmetrically, $\mathbf{m}_{j,n}$ is the match number for worker $\mathbf{i}_{j,n}$ corresponding for firm j's n^{th} match. This implies $w^w_{i,m} = w^f_{\mathbf{j}_{i,m},\mathbf{n}_{i,m}}$, $w^f_{j,n} = w^w_{\mathbf{i}_{j,n},\mathbf{m}_{j,n}}$, $t^w_{i,m} = t^f_{\mathbf{j}_{i,m},\mathbf{n}_{i,m}}$, and $t^f_{j,n} = t^w_{\mathbf{i}_{j,n},\mathbf{m}_{j,n}}$. With this notation, we can equivalently think about the observations from the perspective of either the worker or the firm.

Building on this notation and using the definition of types in equations (1) and (2), the average log wage that worker i with characteristic x_i earns during his lifetime and the average log wage that firm j with characteristic y_j pays are

$$\lambda(x_i) = \frac{\mathbb{E}_{x_i} \sum_{m=1}^{M_i} t^w_{i,m} w^w_{i,m}}{\bar{T}^w_{x_i}} \text{ and } \mu(y_j) = \frac{\mathbb{E}_{y_j} \sum_{n=1}^{N_j} t^f_{j,n} w^f_{j,n}}{\bar{T}^f_{y_j}}. \tag{14}$$

Here the expectations operators $\mathbb{E}_{x_i}$ and $\mathbb{E}_{y_j}$ indicate probabilities taken with respect to the joint distribution of wages, durations, and numbers of matches conditional on characteristic x_i and y_j. Weighting by spell duration defines the types to be the expected earnings at a typical point in time.

We can also compute the population mean of λ and μ, as in equation (4):

$$\bar{w} \equiv \frac{\sum_{x \in X} I_x \bar{T}^w_x \lambda(x)}{\sum_{x \in X} I_x \bar{T}^w_x} = \frac{\sum_{y \in Y} J_y \bar{T}^f_y \mu(y)}{\sum_{y \in Y} J_y \bar{T}^f_y}. \tag{15}$$

Worker types are weighted by the population frequency I_x and the amount of time they are employed $\bar{T}^w_x$ to capture the likelihood the worker is employed in any given cross-section. Similarly firm types are weighted by the amount of time they employ a worker. It is straightforward to prove the second equality in equation (15), that the average wage that the average worker receives is equal to the average wage that the average firm pays. The variances, corresponding to equation (5), are

$$\sigma^2_\lambda \equiv \frac{\sum_{x \in X} I_x \bar{T}^w_x (\lambda(x) - \bar{w})^2}{\sum_{x \in X} I_x \bar{T}^w_x} \text{ and } \sigma^2_\mu \equiv \frac{\sum_{y \in Y} J_y \bar{T}^f_y (\mu(y) - \bar{w})^2}{\sum_{y \in Y} J_y \bar{T}^f_y}. \tag{16}$$

Again we weight observations by their likelihood in the cross-section.

Next, following equation (6) we can compute the covariance between λ and μ in matched pairs:

$$\begin{aligned} c &\equiv \frac{\sum_{x \in X} I_x \mathbb{E}_x \sum_{m=1}^{M_i} t^w_{i,m} (\lambda(x) - \bar{w})(\mu(y_{i,m}) - \bar{w})}{\sum_{x \in X} I_x \bar{T}^w_x} \\ &= \frac{\sum_{x \in X} I_x \mathbb{E}_x \sum_{m=1}^{M_i} t^w_{i,m} \lambda(x) \mu(y_{i,m})}{\sum_{x \in X} I_x \bar{T}^w_x} - \bar{w}^2. \end{aligned} \tag{17}$$

For a characteristic x worker, we compute the expected value of the weighted average product of the deviations of the worker's type from the population mean and her employer's type from the population mean. The weight attached to each match is the duration of the match, and hence the total weight attached to each characteristic x worker is $\bar{T}^w_x$. Equivalently, we can look at this from the perspective of firms and write this as

$$c = \frac{\sum_{y \in Y} J_y \mathbb{E}_y \sum_{n=1}^{N_j} t^f_{j,n} \lambda(x_{j,n}) \mu(y)}{\sum_{y \in Y} J_y \bar{T}^f_y} - \bar{w}^2.$$

This shows that the weight attached to each characteristic y firm is $\bar{T}^f_y$. Finally, the correlation ρ is defined as usual in equation (3).

4.2 Auxiliary Assumptions and an Estimator

This section proposes estimators of c, σ^2_λ, and σ^2_μ that are consistent under some reasonable assumptions in the limit as $\tau \to \infty$. The key assumption is that we have at least two independent observations for each worker and firm. Most obviously, we impose that the minimum number of observations for workers and firms are $\underline{M} \geq 2$ and $\underline{N} \geq 2$. Additionally, we impose four auxiliary assumptions which ensure that these observations are suitably independent:

1. For worker i with characteristic x_i, $w^w_{i,m} = \bar{w}^w_{x_i} + \varepsilon^w_{i,m}$ and $\varepsilon^w_{i,m}$ is independently and identically distributed across $m = \{1, \ldots, M_i\}$ with mean zero and a finite standard deviation $\sigma^w_{x_i}$. Moreover, $t^w_{i,m}$ and $\varepsilon^w_{i,m'}$ are independent for all $(m, m') \in \{1, \ldots, M_i\}^2$.

2. For firm j with characteristic y_j, $w^f_{j,n} = \bar{w}^f_{y_j} + \varepsilon^f_{j,n}$ and $\varepsilon^f_{j,n}$ is independently and identically distributed across $n = \{1, \ldots, N_j\}$ with mean zero and a finite standard deviation $\sigma^f_{y_j}$. Moreover, $t^f_{j,n}$ and $\varepsilon^f_{j,n'}$ are independent for all $(n, n') \in \{1, \ldots, N_j\}^2$.

3. For any worker i with characteristic x_i and all $m \in \{1, \ldots, M_i\}$, $\bar{w}^w_{x_i}$ and $\varepsilon^w_{i,m'}$ are independent of $\varepsilon^f_{\mathbf{j}_{i,m},n'}$ for all $m' \neq m$ and all $n' \neq \mathbf{n}_{i,m}$. Moreover, for any firm j with characteristic y_j and all $n \in \{1, \ldots, N_j\}$, $\bar{w}^f_{y_j}$ is independent of $\varepsilon^w_{\mathbf{i}_{j,n},m'}$ for all $m' \neq \mathbf{m}_{j,n}$.

4. For all $i \neq i'$, m, and m', $\varepsilon^w_{i,m}$ and $\varepsilon^w_{i',m'}$ are independent, as are $t^w_{i,m}$ and $t^w_{i',m'}$. For all $j \neq j'$, n, and n', $\varepsilon^f_{j,n}$ and $\varepsilon^f_{j',n'}$ are independent, as are $t^f_{j,n}$ and $t^f_{j',n'}$.

The two-sided matching model considered in Section 3.2 satisfies these assumptions. So does a repeated version of the discrete choice model discussed in Section 3.3, if the amenity shock is independently drawn in each period.

Auxiliary Assumption 1 consists of two pieces. First, worker i's wage is independently and identically distributed across matches with a characteristic-specific mean $\bar{w}^w_{x_i}$ and standard deviation $\sigma^w_{x_i}$. Second, wages are independent of durations both in the same match and in other matches conditional on the worker's characteristic. We recognize that this assumption is strong without other restrictions and so in our empirical analysis in Section 5.2, we construct different real-world data sets designed to ensure that this assumption is satisfied. For example, we focus only on wages in jobs that are separated by an unemployment spell. Auxiliary Assumption 2 imposes the analogous assumptions on firms.

Auxiliary Assumption 3 imposes that if a worker and a firm are matched at some point in time, the error terms in their other matches are independent of each other. It also imposes that the error in the worker's wage equation in one match is independent of the employer type in other matches and symmetrically that the error in the firm's wage equation in one match is independent of the employee type in other matches. We stress that this assumption allows the error terms to be correlated with each other and with the partner's mean wage within a match, and indeed this will typically be the case.[16]

The first three auxiliary assumptions are useful for finding individual-level unbiased estimators of worker and firm types and the covariance between them. Auxiliary Assumption 4 rules out the possibility of correlated shocks. This gives us a law of large numbers, ensuring that the average of these unbiased estimators is consistent as the economy grows large. In the data, we handle aggregate shocks by deflating wages by the economy-wide average wage, but other correlations, e.g. within region or industry, may matter in practice.

Armed with these assumptions, we relate the worker and firm type to the means in the auxiliary wage equations:

Proposition 2 *A worker with characteristic x has type $\lambda(x) = \bar{w}^w_x$. A firm with characteristic y has type $\mu(y) = \bar{w}^f_y$.*

We relegate the proof of this and all other propositions in this section to Appendix B.1.

Next we construct consistent estimators of the variance-covariance matrix of λ and μ in matched pairs, i.e. of σ^2_λ, σ^2_μ, and c defined in equations (16) and (17). Start with the variance of worker types. Define

$$\hat{\lambda}_i \equiv \frac{\sum_{m=1}^{M_i} w^w_{i,m}}{M_i} \text{ and } \widehat{\lambda^2_i} \equiv \frac{\sum_{m=1}^{M_i} \sum_{m' \neq m} w^w_{i,m} w^w_{i,m'}}{M_i(M_i - 1)}, \tag{18}$$

[16]One situation where the third auxiliary assumption would be problematic is if a worker and firm are matched together multiple times, since in this case, the errors would naturally be correlated within all matches. In our model, we can avoid this possibility if $\bar{M} \leq \min_{y \in Y} J_y$ and $\bar{N} \leq \min_{x \in X} I_x$ by assuming that workers and firms sample partners without recall. In the data, we treat multiple spells with the same employer as a single match.

and use that to find an estimator of σ_λ^2:

Proposition 3 *In the limit as $\tau \to \infty$, a consistent estimator of the mean wage $\bar{w}$ is*

$$\hat{\bar{w}} \equiv \frac{\sum_{i=1}^{\tau I} T_i^w \hat{\lambda}_i}{\sum_{i=1}^{\tau I} T_i^w}; \tag{19}$$

and a consistent estimator of the variance of worker types σ_λ^2 is

$$\widehat{\sigma_\lambda^2} \equiv \frac{\sum_{i=1}^{\tau I} T_i^w \widehat{\lambda_i^2}}{\sum_{i=1}^{\tau I} T_i^w} - \hat{\bar{w}}^2. \tag{20}$$

We show in the proof of Proposition 3 that $\hat{\lambda}_i$ and $\widehat{\lambda_i^2}$ defined in equation (18) are unbiased estimators of $\lambda(x_i)$ and $\lambda(x_i)^2$. The proof of consistency of $\hat{\bar{w}}$ and $\widehat{\sigma_\lambda^2}$ is a standard law of large numbers argument.

The logic for firms is identical. Define

$$\hat{\mu}_j \equiv \frac{\sum_{n=1}^{N_j} w_{j,n}^f}{N_j} \text{ and } \widehat{\mu_j^2} \equiv \frac{\sum_{n=1}^{N_j} \sum_{n' \neq n} w_{j,n}^f w_{j,n'}^f}{N_j(N_j - 1)}, \tag{21}$$

unbiased estimators of $\mu(y_j)$ and $\mu(y_j)^2$. Note that

$$\hat{\bar{w}} = \frac{\sum_{j=1}^{\tau J} T_j^f \hat{\mu}_j}{\sum_{j=1}^{\tau J} T_j^f},$$

since this just averages wages from the firm's instead of the worker's perspective. Then

Proposition 4 *In the limit as $\tau \to \infty$, a consistent estimator of the variance of firm types σ_μ^2 is*

$$\widehat{\sigma_\mu^2} \equiv \frac{\sum_{j=1}^{\tau J} T_j^f \widehat{\mu_j^2}}{\sum_{j=1}^{\tau J} T_j^f} - \hat{\bar{w}}^2. \tag{22}$$

We omit the proof, since it is isomorphic to the proof of Proposition 3.

Finally, we turn to an estimator of the product of worker and firm types. Let

$$\hat{c}_{i,m} \equiv \frac{\sum_{m' \neq m} w_{i,m'}^w}{M_i - 1} \frac{\sum_{n' \neq \mathbf{n}_{i,m}} w_{\mathbf{j}_{i,m},n'}^f}{N_{\mathbf{j}_{i,m}} - 1}. \tag{23}$$

Each of the $M_i - 1$ other wages $w_{i,m'}^w$ is an unbiased estimator of $\lambda(x_i)$. Each of the $N_{\mathbf{j}_{i,m}}$ other wages $w_{\mathbf{j}_{i,m},n'}^f$ is an unbiased estimator of $\mu(y_{\mathbf{j}_{i,m}})$. Moreover, the third auxiliary assumption implies the two estimators are independent and hence the product is an unbiased estimator of $\lambda(x_i)\mu(y_{\mathbf{y}_{i,m}})$. We leverage this insight to get a consistent estimator of the covariance:

Proposition 5 *In the limit as $\tau \to \infty$, a consistent estimator of the covariance between worker and firm types c is*

$$\hat{c} \equiv \frac{\sum_{i=1}^{\tau I} \sum_{m=1}^{M_i} t_{i,m}^{w} \hat{c}_{i,m}}{\sum_{i=1}^{\tau I} T_i^w} - \hat{\bar{w}}^2. \tag{24}$$

Armed with consistent estimators of the covariance and two variances, it is straightforward to construct an estimator of the correlation as

$$\hat{\rho} \equiv \frac{\hat{c}}{\sqrt{\widehat{\sigma_\lambda^2}\widehat{\sigma_\mu^2}}}. \tag{25}$$

This estimator is consistent as $\tau \to \infty$ if σ_λ and σ_μ are both positive.

We recognize that some auxiliary assumptions are restrictive, in particular the assumption that duration of a job $t_{i,m}^w$ is independent of the error in the wage equation for that job $\varepsilon_{i,m}^w$. This assumption is violated in many models of on-the-job search, where high wage jobs last longer than low wage jobs (Burdett and Mortensen, 1998). It may also be violated in a model where jobs differ in their layoff risk, in which case a high wage may serve as a compensating differential for an unstable job (Jarosch, 2015). In Appendix C, we propose an alternative estimator that is consistent under a different set of assumptions. In particular, we relax the assumption that $t_{i,m}^w$ and $\varepsilon_{i,m}^w$ are independent and replace it with a new assumption, that the mean and variance of duration $t_{i,m}^w$ are independent of $\bar{w}_{x_i}^w$. We derive an alternative estimator that is consistent under these assumptions. We also show in Appendix C that our results using Austrian data are insensitive to which estimator we use.

Formulae (18)–(25) are readily implemented in a real-world data set by setting $\tau = 1$. Note that in finite data sets, the estimator of σ_λ^2 in equation (20) can be negative. In this case, we say that the estimator $\widehat{\sigma_\lambda^2}$ is zero and the estimator of the correlation $\hat{\rho}$ has magnitude 1 and the same sign as the estimator of the covariance $\hat{c}$ in equation (24). We make analogous definitions if the estimator of σ_μ^2 in equation (22) is negative.[17] We use these formulae throughout the rest of the paper.

4.3 Small Sample Properties of the Estimators

We next examine small sample properties of our estimator $\hat{\rho}$. We create different-sized finite artificial data sets from the two-sided search and discrete choice models introduced in Section 3. Importantly, in each data set we keep number of observations per worker deliberately small, 3.8 on average. For each choice of the number of workers $I \in \{2500, 10^4, 10^5\}$, we

[17] In practice, we only encounter these issues using simulated data, and only then when the true variance is very small.

set the number of jobs at $J = I/5$, which guarantees that the number of observations per firm is also small, as in real world data.[18] We explain details of the sample construction in Appendix B.2.

For each model and each choice of I, we create $B = 500$ artificial samples. The actual number of firms and workers in each sample may be smaller because we drop firms and workers with fewer than two observations. For each sample $b \in \{1, \ldots, B\}$, we first compute the types λ and μ from the economy with infinitely many workers and firms (Sections 3.2 and 3.3). We then use the match durations and matching network realized in sample b to recover the variance-covariance matrix (equations 16 and 17) and hence the true correlation ρ_b for that sample. Alternatively, we use log wages, match durations and the matching network in sample b to find the variance-covariance matrix using formulae (20), (22), and (24), and hence recover the feasible estimate $\hat{\rho}_b$. We are interested in how these two measures of correlation compare.

We parameterize the two-sided search model using the benchmark in Section 3.2. The realized correlation ρ_b varies across samples, reflecting randomness in the matching process. It is smaller in economies with fewer workers and firms because the scarcity of matches creates randomness in who matches with whom (fifth column in Table 1). We observe that our estimator performs well even with $I \leq 2500$ workers and $J = 500$ firms, orders of magnitude smaller sample than a typical real world data set. As the number of workers and firms increases, the error becomes smaller.

The bottom panel of Table 1 summarizes the results for the discrete choice model. Again, we parameterize it with the benchmark values in Section 3.3. We treat this as a repeated static model and assume that all matches last one period before new amenity shocks are drawn independently. This means most workers switch jobs each period. As in the previous model, the correlation in each particular sample is different and typically bigger in a larger economy. We again observe that the error in the estimator is very small, even in the sample with $I \leq 2500$ workers and $J = 500$ firms.

Figures 4 and 5 in Appendix B.3 show that we get similar results with other parameter values. Even in samples that are orders of magnitude smaller than a typical real world data set, the errors in the correlation estimates are typically in the third decimal place, except when the variance of firm types is nearly equal to zero.

[18] We use the distribution of observations per worker in the Austrian data in the sample for men corresponding to column (3) of Table 2.

	sample means					distribution of $\hat{\rho}_b - \rho_b$		
	workers	firms	$\frac{\text{matches}}{\text{worker}}$	$\frac{\text{matches}}{\text{firm}}$	ρ_b	5%	mean	95%
Two-sided Search								
	2,412	500	3.9	18.7	0.757	-0.007	-0.001	0.005
	9,660	2,000	3.9	18.8	0.771	-0.003	0.000	0.003
	96,618	20,000	3.9	18.8	0.775	-0.001	0.000	0.001
Discrete Choice								
	2,496	500	4.3	21.6	0.744	-0.008	-0.001	0.006
	9,996	2,000	4.4	21.8	0.748	-0.004	0.000	0.003
	99,997	20,000	4.4	21.9	0.749	-0.001	0.000	0.001

Table 1: Monte Carlo simulations in the two-sided search model with match-specific productivity shocks and discrete choice model. For each choice of I, J, we create $B = 500$ artificial data sets as described in the main text. The first five columns show several descriptive statistics computed as means across samples – number of workers I, number of firms J, number of matches per worker, number of matches per firm and true sample correlation ρ_b. The last three columns show the mean, the 5^{th} and 95^{th} quartile of the error distribution, $\hat{\rho}_b - \rho_b$. We parameterize the two-sided search model using the benchmark in Section 3.2. We parameterize the discrete choice model using the benchmark in Section 3.3.

5 Data

5.1 Data Description

We measure the correlation between worker and firm types using two panel data sets from the Austrian social security registry (Zweimuller, Winter-Ebmer, Lalive, Kuhn, Wuellrich, Ruf, and Buchi, 2009), the Austrian Social Security Database (ASSD) and the Arbeitsmarktdatenbank (AMDB, Labor Market Database). The ASSD covers the universe of workers in the private sector from 1972 to 2007, the AMDB from 1986 to 2018.[19] For each worker, each data set contains information about every job they hold. More precisely, in every calendar year and for every worker-firm pair,[20] we observe annual earnings and days worked during the year. In the AMDB, we see two sources of earnings, regular wage payments and bonus payments, which we combine together to compute annual earnings. Earnings are top-coded at the maximum social security contribution level, which rises over time.[21]

[19]The two data sets cover the same set of workers but we are not allowed to merge them, and therefore we will treat them as two separate data sets.

[20]A firm is identified by its employer identification number (EIN). Some firms may have multiple EINs.

[21]For example, in 2018, the cap for monthly wage earnings is €5,130 and the cap for annual bonus payment is €10,260. The fraction of male worker-firm observations affected by top-coding fell from a peak of 15.2 percent in 1990 to 10.3 percent in 2018. Top-coding affects far fewer female worker-firm observations, varying from 1.7 to 4.8 percent during our sample period. We discuss the importance of top-coding for our results in Appendix D.

Each data set has its advantages. The AMDB covers later years while the ASSD provides more demographic information on workers. In particular, both data sets have some limited information on workers and firms, including workers' birth year and sex, and region and industry for firms. We observe registered unemployment spells in the AMDB and, after 1986, in the ASSD. In the ASSD, we additionally observe whether the position is blue or white collar, as well as the education of most workers who experience registered unemployment.

Following Card, Heining, and Kline (2013), we focus on workers age 20–60. We look separately at men and women, but recognize that selection into employment may be a more serious issue for women. We drop marginal jobs (less than 10 hours per week) and data that include an apprenticeship. We note that the data sets do not have an indicator of part-time jobs. While this might not be a serious concern for men, part-time work is prevalent among women. Between 1994 and 2007, on average 4.7 percent of employed men and 34.0 of employed women worked part-time.[22] Thus, caution is required when interpreting the results for women.

For each worker-firm-year, we first construct a measure of the log daily wage by taking the difference between log annual earnings, which is the sum of wages and bonus, and log days worked. We then regress this on a full set of dummies for the calendar year and age. The first set of dummies captures the effects of aggregate nominal wage growth, while the second removes a standard age-earnings profile. Our analysis focuses on these wage residuals. In most of our analysis we use AMDB data. We supplement this with ASSD data in order to examine impact of observable characteristics.

5.2 Independence Assumptions

Auxiliary Assumption 1 states that the error terms in the wage equation $\varepsilon^w_{i,m}$ are independent and identically distributed across matches m for a given worker i. Auxiliary Assumption 2 is an analogous condition on the firm side. We recognize that this might not be always satisfied in the data. We approach this in several ways, always motivated by economic theories such as Burdett and Mortensen (1998), Shimer and Smith (2000), and Postel-Vinay and Robin (2002). These theories tell us that this condition is easily satisfied for firms but not always for workers. In this section we explain how we select a sample of workers where the conditional independence assumption is likely to be satisfied.

We start by selecting all workers for whom we have at least two wage observations during the 33 years of data. This includes workers who are employed in at least two years, as well as workers who work for two different employers in the same calendar year. We treat the

[22] These statistics come from the Statistical office of Austria, https://www.statistik.at.

annual residual wage observations as independent and measure the correlation accordingly. We call this independence assumption I.

The advantage to measuring the correlation using independence assumption I is that we minimize sample selection issues, since we only drop workers with a single employer in a single year. The disadvantage is that a worker's wage at a single employer is likely to be serially correlated, a violation of the conditional independence assumption. We therefore take a weighted average of the residual wage at the level of the worker-firm match, weighting by days worked, and treat this match-level wage as a single observation.[23] We then select all workers who are employed by at least two employers and measure the correlation. We call this independence assumption II: wages are independent across matches.

We recognize that, due to job-to-job movements, residual wages might be correlated across employment relationships. To understand the problem, consider the job ladder model from Burdett and Mortensen (1998). There, an employed worker accepts a job offer from another firm if and only if it pays a higher wage. This means that the wage in jobs held before and after the job-to-job transition are correlated. According to this model, an unemployment spell breaks this correlation and so wages in two employment relationships separated by an unemployment spell are independent. Guided by these insights, we select all workers with at least two employment spells separated by a spell of registered unemployment and take the match-level wage in the longest job during each employment spell. This is independence assumption III: wages are independent across employment spells.

According to Burdett and Mortensen (1998) and Postel-Vinay and Robin (2002), the wage in any two jobs during different employment spells are conditionally independent; however, they are not necessarily identically distributed. For example, the first accepted wage out of unemployment comes from a lower distribution than subsequent wages. To address this concern, we select only workers with at least two uncensored employmenet spells (that is, workers with UEUEU transitions, where E represents an employment spell and U a registered unemployment spell). For these workers, we consider three different wage measures. First, during any employment spell that starts and ends with an unemployment spell, we measure the average wage in the longest job. Second, for any employment spell that starts out of unemployment, we measure the average wage in the first job out of unemployment. Third, for any employment spell that ends with unemployment, we measure the average wage in the last job before unemployment. In many models with on-the-job search, any such wage observations are independently and identically distributed. We call all of these independence assumption IV.

[23] Recalls are common in the Austrian labor market (Pichelmann and Riedel, 1992). We treat all instances where a worker is employed by a firm as a single observation.

We comment on how we handle multiple matches between the same worker and firm. This happens if a match between a worker and a firm lasts more than one year or if a worker is employed by the same firm at two different times of his career. We treat both these situations in the same way. Under independence assumption I, we treat each annual wage observation for a worker-firm pair as a different match and hence by construction, workers have multiple matches with the same employer. As noted, this is unlikely to be consistent with our auxiliary assumptions. Under independence assumptions II, III, and IV, we combine all wage observations of a single worker-firm match into one observation by constructing the average wage. If the wage observations are separated by an unemployment spell, we drop that unemployment spell from our analysis. Thus we are left with at most one match between a worker and firm.

Our approach requires at least two observations for each worker and each firm. After making the inial selection of workers, as described above, we trim our data set by first dropping any firm that only employs a single worker in the data set. If this leaves any of the workers with a single wage observation, we drop them from the data as well. We repeat. This process necessarily stops in a finite number of steps, either with an empty data set or with a data set containing the largest set of workers with multiple employers and employers with multiple workers in that data set. In our case the resulting data set is always non-empty.

There is a tradeoff when we impose stricter independence assumptions, i.e. go from assumption I to assumption IV. On the one hand, our auxiliary assumptions are more plausible when we impose stricter assumptions. On the other hand, the sample becomes more selected. The concern is that the distribution of worker characteristics $F(x)$, the conditional distribution of jobs given worker characteristics $\Phi_x(y)$, and the wage equation $w(x, y, z)$ change as we restrict the sample. Since our definition of the correlation ρ depends on these objects, one would expect the correlation to differ across samples, even if our auxiliary assumptions were satisfied in all these samples. We address these selection issues explicitly when we discuss our results, which we turn to next.

6 Estimated Correlation between Types

6.1 Main Results

Table 2 shows the main results for men and women. We estimate the correlation and covariance between matched worker and firm types, as well as the variance of types. Different columns correspond to different independence assumptions.

Column (1) of Table 2 uses independence assumption I to construct the correlation.

Estimated Correlation and Variances

	(1)	(2)	(3)	(4)	(5)	(6)
Men						
correlation of matched types $\hat{\rho}$	0.669	0.555	0.473	0.488	0.463	0.469
covariance of matched types $\hat{c}$	0.065	0.042	0.027	0.027	0.026	0.028
variance of worker types $\widehat{\sigma_\lambda^2}$	0.107	0.075	0.051	0.051	0.051	0.054
variance of job types $\widehat{\sigma_\mu^2}$	0.090	0.075	0.062	0.059	0.063	0.066
number of workers (thousands)	3,884	2,916	1,646	971	971	970
number of firms (thousands)	705	532	343	272	273	276
number of observations (thousands)	64,359	16,401	7,211	4,883	4,881	4,880
share of observations top-coded	0.112	0.074	0.047	0.020	0.009	0.012
Women						
correlation of matched types $\hat{\rho}$	0.615	0.440	0.423	0.459	0.432	0.443
covariance of matched types $\hat{c}$	0.090	0.042	0.031	0.033	0.030	0.033
variance of worker types $\widehat{\sigma_\lambda^2}$	0.168	0.094	0.063	0.066	0.060	0.066
variance of job types $\widehat{\sigma_\mu^2}$	0.127	0.097	0.085	0.080	0.080	0.082
number of workers (thousands)	3,340	2,532	1,494	798	798	797
number of firms (thousands)	758	530	345	249	248	250
number of observations (thousands)	52,128	12,451	5,582	3,294	3,294	3,291
share of observations top-coded	0.035	0.031	0.021	0.009	0.004	0.005
independence assumption	I	II	III	IV	IV	IV
observations included	all	all	longest	longest	first	last

Table 2: Estimates of correlations, covariances, and variances between matched workers' and firms' types using AMDB 1986–2018. All columns use residual log wages, obtained by regressing log wages on year and age dummies. Columns (2)–(6) aggregate residual wages to the worker-firm match level by taking a weighted average of wages within the match across years. Before applying our method, we iteratively drop firms and workers with a single wage observation. Each column uses a different sample to estimate the correlation. Independence assumption I includes workers with at least two firm-year wage observations and treats each year as an independent observation. Independence assumption II includes workers with at least two distinct employers and treats each employer as an independent observation. Independence assumption III includes workers with at least two employment spells and treats the longest jobs during each employment spell as independent observations. Independence assumption IV includes workers with at least two uncensored employment spells and treats either the longest (4), first (5), or last (6) job during each employment spell as independent observations.

This treats any two firm-year observations for a given worker as independent. We see that the correlation $\hat{\rho}$ is high, above 0.6 for both men and women. It seems unlikely, however, that this data set satisfies our auxiliary assumptions. First, the error term is likely to be serially correlated within a worker-firm match, violating auxiliary assumptions 1 and 2. Second, the worker and firm errors are likely correlated within a match. If a match lasts for multiple periods, this correlation lasts across multiple observations, violating the third auxiliary assumption. We therefore do not find these numbers credible.

Column (2) uses the more plausible independence assumption II to construct the correlation, aggregating wage observations to the level of the worker-firm match. Each component of the correlation drops, and so does the correlation. There are two potential explanations for this drop. On the one hand, we expect that independence assumption I is incorrect and so the resulting correlation in column (2) is biased. On the other hand, we lose some workers going from column (1) to column (2) and so the drop can reflect the changing sample. To evaluate importance of these two explanations, we take the firms and workers from column (2) and impose independence assumption I. That is, we treat any two firm-year observations for a given worker as independent. The estimated correlation is 0.644 for men and 0.585 for women, only slightly smaller than the numbers in column (1). This suggests that the difference between columns (1) and (2) is driven primarily by a violation of our auxiliary assumptions in column (1), with a bias reduction in column (2).

We next turn to independence assumption III, which treats wage observations from different employment spells as independent, as in standard theories of on-the-job search. Column (3) shows a drop in the estimated correlation both for men and women, even though the change for women is small. To evaluate the importance of selection and bias, we take the sample of workers and firms from (3) and impose independence assumption II. That is, we use all matches of these workers. The estimated correlation is 0.502 for men and 0.415 for women. In this case, selection plays a significant role for both men and women, although there is also some bias for men. Nevertheless, we view the estimates in column (3) as a reasonable baseline.

Finally, we look at independence assumption IV, which recognizes that wage observations at different points during different employment spells are independent but not identically distributed. Columns (4), (5), and (6) look at the longest, first, and last job during multiple employment spells. These estimates are remarkably similar to the correlation in column (3), both for men and women.

In summary, dropping the biased estimates in column (1), the estimated correlation between types ranges from 0.463 to 0.555 for men, and from 0.423 to 0.459 for women. The exact number depends on the independence assumption. As we move from the independence assumption I to IV, the auxiliary assumptions are more likely to be satisfied. The downside is that each concept imposes additional restrictions on the sample. We choose to focus on the results in column (3) because we believe those are likely to satisfy the independence assumption while minimizing the sample selection issues in the last three columns.

We note that in each column of Table 2, $\min\{\widehat{\sigma_\lambda^2}, \widehat{\sigma_\mu^2}\} > \hat{c}$. Viewed through the lens of Proposition 1, this suggests we are not in the case where we need to worry about our correlation giving the opposite sign as the AKM correlation.

6.2 Top Coding

In our baseline results in Table 2 column (3), top-coding affects 4.7 percent of men's observations and 2.1 percent of women.[24] To assess the importance of top-coding, we propose making the top-coding more severe in each year.[25] Figure 6 in Appendix D shows that for men, the estimated correlation is nearly independent of the share of top-coded observations. For women, the estimated correlation is a decreasing function of the share observations that are top-coded. Extrapolating to a lower share of top-coded observations suggests that in the absence of top-coding, the estimated correlation for women might be slightly higher.

6.3 Confidence Intervals

We construct confidence intervals using a bootstrap procedure. Our main approach to the bootstrap involves constructing artificial data sets that generate the moments reported in Table 2, including the variances of worker and firm types, the covariance of matched workers' and firms' types, as well as the variance of log wages, the distribution of the number of matches per worker and firm, and the joint distribution across matches of the durations of workers' jobs. The artificial data sets also target the number of workers and firms but do not hit this exactly. They do not target who matches with whom, i.e. the structure of the network of matches, or the wage paid in each match. Appendix E describes how we construct these artificial data sets.

We construct $B = 500$ artificial data sets. For each data set $b = 1, \ldots, B$, we know the data generating process and hence we know each worker's and firm's type λ and μ. We then

[24]We consider the log wage for a worker-firm pair to be top-coded if at least one annual wage or bonus observation for that worker-firm pair is top-coded, and report the share of such worker-firm pair observations. The share of top-coded observations increases to 6.7 percent for men and remains unchaged at 2.1 percent for women if we look at calculate share of the annual worker-firm observations which are top-coded.

[25]The usual approach involves imputing values to the top-coded observations (see for example, Card, Heining, and Kline, 2013). Interpreting either approach requires an assumption that the behavior of top-coded observations is similar to the behavior of other high wages. We believe our approach is more transparent and easier to implement.

compare the actual correlation between types, ρ_b, with the correlation estimated using our approach, $\hat{\rho}_b$. We construct confidence intervals using the difference $\rho_b - \hat{\rho}_b$ as described in Appendix F. We find that this difference is typically small and is centered around zero, as one would expect for a consistent and unbiased estimator. For example, in column (3) of Table 2, the estimated correlation for men is $\hat{\rho} = 0.4732$ and the 95 percent confidence interval is $[0.4719, 0.4746]$. For women, the estimated correlation is $\hat{\rho} = 0.4231$ and the 95 percent confidence interval is $[0.4214, 0.4249]$. The results in the other columns are similar.

A drawback of our artificial data sets is that the network structure in the artificial and real-world data differ in some important dimensions. For example, in the real-world data, about 3 percent of a typical worker's coworkers at one employer are also coworkers at another one of her employers. In our artificial data, this happens about 0.1 percent of the time. We propose an alternative procedure to address this concern. We hold fixed the set of matches in real-world data and draw random types for each worker and firm. We then draw wages for each match in a manner that is consistent with the definition of types. Unfortunately, generating types that are consistent with the real world correlation structure requires drawing a correlated random vector of dimension $I + J$, which is computationally infeasible. Instead, we ask what we would measure if the correlation between types were zero. If the true value of ρ were zero, 95 percent of the time our approach using the matching network in column (3) would have generated estimates of $\hat{\rho}$ between -0.0040 and 0.0034 for men between -0.0055 and 0.0050 for women. It is vanishingly unlikely that the Austrian data could have been generated from an economy without sorting.

6.4 Other Observable Characteristics

We now examine how controlling for fixed observable characteristics of workers and firms affects the estimated correlation. We use ASSD data set in this section because it provides more observable characteristics than AMDB data. This data has registered unemployment spells over a shorter time period, from 1986 to 2007, and so we first replicate column (3) of Table 2 on this shorter sample. Column (1) in Table 3 shows that the correlation is slightly lower in this earlier sample.

We now relax the assumption that a firm has the same type (expected wage) for all its employees. Effectively we break a firm into different types for employees with different skill levels and estimate the correlation between types on this adjusted data set. To start, treat a firm j as a cross between a firm identifier and a worker's education level. We use five different education categories: no completed education, middle school, technical secondary school, academic secondary school, and college. We start with the same data set as in column

Impact of Observables

	(1)	(2)	(3)	(4)
men				
correlation of matched types $\hat{\rho}$	0.439	0.521	0.525	0.580
covariance of matched types $\hat{c}$	0.019	0.023	0.024	0.028
variance of worker types $\widehat{\sigma_\lambda^2}$	0.039	0.039	0.041	0.049
variance of job types $\widehat{\sigma_\mu^2}$	0.049	0.052	0.052	0.047
number of workers (thousands)	1,101	949	1,045	917
number of firms (thousands)	234	337	247	181
number of observations (thousands)	4,376	3,895	3,975	2,706
share of observations top-coded	0.078	0.071	0.074	0.070
women				
correlation of matched types $\hat{\rho}$	0.418	0.505	0.523	0.527
covariance of matched types $\hat{c}$	0.028	0.036	0.040	0.038
variance of worker types $\widehat{\sigma_\lambda^2}$	0.061	0.061	0.066	0.072
variance of job types $\widehat{\sigma_\mu^2}$	0.075	0.083	0.088	0.072
number of workers (thousands)	951	786	895	646
number of firms (thousands)	238	315	241	163
number of observations (thousands)	3,190	2,660	2,757	1,787
share of observations top-coded	0.054	0.024	0.028	0.022
independence assumption	III	III	III	III
education	no	yes	no	no
white/blue collar	no	no	yes	no
industry	no	no	no	yes

Table 3: Results controlling for education, job classification, and industry using ASSD 1986–2007. All columns use residual log wages, aggregated to the worker-firm match level by taking a weighted average of wages within the match across years. All columns use independence assumption III, treating the longest jobs during each employment spell as independent observations. Column (1) shows results without controlling for any observables. In column (2), we treat each firm $\times$ education category as a separate firm. In column (3), we treat each worker $\times$ job position and firm $\times$ job position as different workers and firms. In column (4), we treat each worker $\times$ industry as different workers. We adjust the counts of workers and firms accordingly.

(1) of Table 3, but lose about ten percent of workers because they are missing education data.[26] We then drop firm $\times$ education observations that only appear once in the data set. This in turn forces us to drop some workers, etc. Finally, we measure the correlation between the remaining worker and firm $\times$ education types. Table 3 column (2) shows that allowing firm types to differ by educational category raises the correlation between matched types from 0.439 to 0.521 for men and from 0.418 to 0.505 for women. This is consistent with the view that firms are a collection of heterogeneous jobs and so ignoring that heterogeneity causes us to underestimate the true correlation.

We proceed in a similar way with the type of position, treating a firm identifier as distinct for white and blue collar jobs. Even though the type of position is a permanent characteristic for the majority of workers, some do hold both blue and white collar jobs, and thus we treat a worker at different positions as a different worker type as well. This again substantially raises the estimated correlation to 0.525 for men and 0.523 for women, see Table 3 column (3). Again, we interpret this as evidence that firms are collections of heterogeneous jobs and sorting occurs both across firms and across job categories within firms.

Finally, we investigate the role of industry. We use ten one-digit SIC industry categories. These are fixed at the firm level but workers switch industries, in which case we treat them as if they were different individuals. Even though we start from the same set of workers and firms, we lose observations when the worker does not hold two jobs in the same industry, about 38 percent of the observations for men and 37 percent for women. The correlation between the remaining matched workers and jobs is the highest yet, 0.580 for men and 0.527 for women, reported in Table 3 column (4). In summary, the headline numbers in Table 2 significantly understate the amount of sorting in the economy because they ignore the fact that firms are collections of heterogeneous jobs and workers' type changes over time. It seems plausible that if we could account for unobserved time-varying heterogeneity, the measured correlation would increase further.

6.5 Time Series

Our approach is amenable to time series analysis. To see this, we redo all of our analysis using only a single year's data at a time. That is, we measure the average log wage for a worker-firm pair using only wage information from the considered year, even if the match exists in other years. We focus throughout on independence assumption III, selecting the

[26]Missing education data is not random, even conditional on unemployment. Those men (women) without education data earn a residual log wage that is 0.19 (0.16) standard deviations higher than the average residual log wage of workers with recorded education. Furthermore, workers with missing education have fewer employment spells on average, 2.4 compared to 4.1 for men, and 2.3 compared to 3.4 for women.

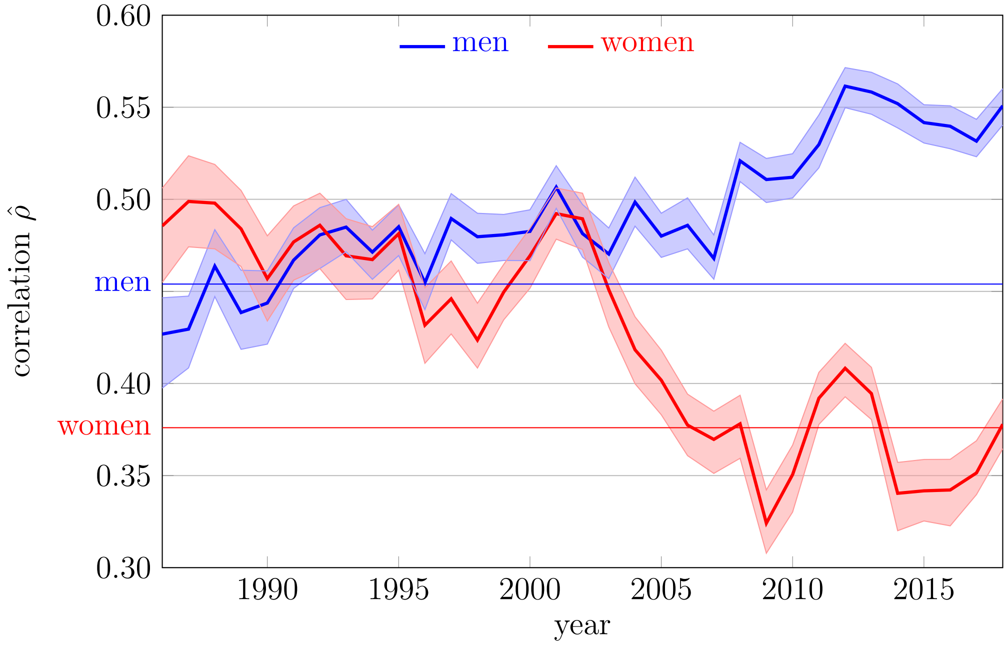

Figure 3: Year-by-year correlation between worker and firm types under independence assumption III, using AMDB 1986–2018. Thick solid lines are computed year-by-year and shaded areas are bootstrapped 95 percent confidence intervals. For each year, the sample considers all workers who switched employers after an unemployment spell within that year, and includes one job for each employment spell of these workers. The sample only includes the wage observations for that year, even if the match continued in other years. The thin horizontal lines are correlations computed in samples constructed by pooling all year-by-year samples together.

last job before the unemployment spell and the first job after the unemployment spell.

Using only those workers who switch employers after an unemployment spell within a year reduces our sample size from 1.6 million workers to an average of 62,000 workers per year for men, and from 1.5 million to 36,000 for women. This is still sufficiently large to estimate the annual correlation between worker and firm types. Figure 3 shows that the correlation between worker and firm types increased slightly for men, from an initial 0.427 in 1986 to around 0.489 in 1997, where it stayed until 2007. We see another increase in the correlation in the last ten years, reaching 0.551 in year 2018. The figure also shows a general downward trend in the correlation for women. In both cases, the bootstrapped 95 percent confidence intervals are small in every year. The stability of these estimates from year-to-year provides additional support for our methodology.

The annual correlations average 0.493 for men and 0.425 for women, very similar to the correlations of 0.473 and 0.423 reported in column (3) of Table 2 using the full sample. This reflects a combination of two different forces. First, the sample of workers is different, since for the time series analysis we use workers who have multiple employment spells within a year, while some workers may have multiple spells, but only in different years. To assess the importance of this, we pool the samples from the time series analysis and estimate a single correlation, 0.454 for men and 0.376 for women.[27] Since these are smaller than the numbers reported in column (3) of Table 2, we conclude that sample selection reduces the measured correlation.

Second, in the time series analysis, workers and firms are not linked across years and hence a worker in one year is treated independently of the same worker in a different year. If types change over time in ways that are not captured by age dummies in a wage regression, the correlation in the pooled sample will understate the correlation at any point in time since the time-average types λ_i and μ_j are noisy measures of the time-varying types $\lambda_{i,t}$ and $\mu_{j,t}$. In Appendix G, we establish this formally when conditional expectations are linear. This suggests yet another reason why the estimates in Table 2 understate the amount of sorting between workers and firms.

One possible concern with our results in this section is that we focus on the last job in one employment spell and the first job in the next spell. Although the wages in these jobs may be independent, theory tells us that they are not drawn from the same distribution. Indeed, in our data there are level differences in wages within a spell: the mean log wage in the first job after unemployment is lower than the mean log wage in the second job, which is lower than in the third job, etc.. There are two reasons why we believe that this is not a major issue. First, the estimated correlation using only first jobs or only last jobs in each employment spell is very similar; see columns (5) and (6) in Table 2. Second, we have regressed log wages on the job's order within a spell, in addition to age and year dummies, before constructing wage residuals. This additional control has little quantitative impact on the correlations in Figure 3.

7 Comparison with the AKM Correlation

The standard method of measuring whether high wage workers take high wage jobs is due to Abowd, Kramarz, and Margolis (1999). The authors propose running a linear regression of log wages against a worker fixed effect α and a firm fixed effect ψ, as in equation (7). This gives them estimates of each fixed effect, $\hat{\alpha}_i$ for all i and $\hat{\psi}_j$ for all j. They then compute the

[27] In this pooled sample, we aggregate all worker-firm-year residual wages back to the worker-firm level by computing an average log wage over years. We then keep only the longest match in each employment spell. The sample contains 949,508 men and 688,427 women.

Comparison with AKM Correlation

	(1)	(2)	(3)	(4)	(5)	(6)
Men						
$\hat{\rho}$	0.662	0.555	0.473	0.488	0.463	0.468
$\hat{\rho}_{AKM}$	0.086	0.202	0.122	0.053	0.018	0.029
number of workers (thousands)	3,862	2,915	1,646	971	971	970
number of firms (thousands)	682	532	343	272	273	276
number of observations (thousands)	64,193	16,400	7,211	4,883	4,881	4,880
share of observations top-coded	0.112	0.074	0.047	0.020	0.009	0.012
Women						
$\hat{\rho}$	0.606	0.440	0.423	0.459	0.432	0.443
$\hat{\rho}_{AKM}$	0.026	0.093	0.067	0.061	0.047	0.057
number of workers (thousands)	3,305	2,532	1,494	798	798	797
number of firms (thousands)	720	530	345	249	248	250
number of observations (thousands)	51,813	12,450	5,582	3,294	3,294	3,291
share of observations top-coded	0.035	0.031	0.021	0.009	0.004	0.005
independence assumption	I	II	III	IV	IV	IV
observations included	all	all	longest	longest	first	last

Table 4: Comparison of our estimates of correlation and AKM fixed effects estimates using AMDB 1986–2018. Both ρ and ρ_{AKM} are estimated on the largest connected set. Otherwise the data construction is identical to Table 2. See the notes to that table for details.

correlation between $\hat{\alpha}_i$ and $\hat{\psi}_j$ in matched pairs. As we mentioned in the introduction, a fair summary of the extensive literature that follows that paper is that the estimated correlation is close to zero and sometimes negative.

Table 4 verifies that this is true with our approach to the data. We use the same approach as in Table 2, with one difference: the AKM correlation is only identified on the largest connected set of workers and firms and so we work with this set. Comparing Tables 2 and 4, we see that very few workers and firms are not part of the largest connected set and so unsurprisingly this has little effect on our estimate $\hat{\rho}$. We then estimate the AKM correlation. The exact formulae for the duration-weighted estimator $\hat{\rho}_{AKM}$ are in Appendix H.1. Our estimates lie between 0.018 and 0.202 for men and 0.026 and 0.093 for women. Across the six columns, the fixed effects correlation is on average 0.433 below our estimate of the correlation for men and 0.409 below our estimate of the correlation for women.

Why is the estimated correlation between the AKM fixed effects so much smaller than the estimated correlation between our measure of types? There are two possible reasons. First, Section 3 shows that ρ is typically, possibly always, larger than ρ_{AKM}, and in structural models the differences may be substantial. We also argued that our correlation better captures the extent of sorting in the real economy when these two measures are different.

Second, a number of authors have noted that the estimator of the AKM correlation using an OLS regression, $\hat{\rho}_{AKM}$, does not generally converge to ρ_{AKM} in a large data set (Postel-Vinay and Robin, 2006; Barth and Dale-Olsen, 2003; Abowd, Kramarz, Lengermann, and Pérez-Duarte, 2004). Instead, consistency requires that the number of independent observations per worker and firm goes to infinity holding fixed the number of workers and firms. This is not a natural feature of real-world data sets. For example, even using 33 years of Austrian data, the median worker has two employers and the median firm has three employees. The literature has called this incidental parameters problem "limited mobility bias" and proved that it leads to an underestimate of the AKM covariance and an overestimate of the standard deviation of worker and firm fixed effects. Together these imply $\rho_{AKM} > \hat{\rho}_{AKM}$ when the former is positive.

The literature has offered some solutions to the limited mobility bias problem; see Appendix H.2 for details. We find that the bias corrections proposed in Andrews, Gill, Schank, and Upward (2008) and Kline, Saggio, and Sølvsten (2019) give us estimates of ρ_{AKM} that are much smaller than our estimate of ρ. For example, under independence assumption III, our estimate of the ρ is 0.46 for men and 0.39 for women, while the OLS estimate of ρ_{AKM} is -0.02 for men and -0.05 for women.[28] The Andrews, Gill, Schank, and Upward (2008) bias correction turns the estimate of ρ_{AKM} positive but still small, 0.10 for both men and women. The correction in Kline, Saggio, and Sølvsten (2019) increases this a bit more, to 0.11 for men and 0.10 for women. On the other hand, the approach in Bonhomme, Lamadon, and Manresa (2019) has a much bigger effect on the results, increasing the correlation to 0.30 for men and 0.29 for women, although this is still one-third smaller than our estimates of the correlation.

Table 5 shows that the results are qualitatively similar in other samples, with our measure of correlation always somewhat larger than the estimate of the AKM correlation from Bonhomme, Lamadon, and Manresa (2019), which in turn is bigger than the bias-corrected estimates in 2008 and Kline, Saggio, and Sølvsten (2019), which are again bigger than the OLS estimates of the AKM correlation. Our reading of the literature is that there is no agreement yet about which (if any) of these methods reliably estimates ρ_{AKM} in finite samples. For this reason, we do not know whether the difference between our estimates of ρ and existing estimates of ρ_{AKM} reflect conceptual differences between the two concepts or biases in estimating ρ_{AKM}. In the former case, we argued in Section 3 that there are theoretical

[28] These are different than the numbers in Table 4 because the bias corrections impose some restrictions on how we look at the data. The two most important ones are that we use exactly two matches for each worker and that we cannot weight any of the estimates by duration. Again, see Appendix H.2 for details.

Comparison with Bias-Corrected Estimates of AKM Correlation

	(1)	(2)	(3)	(4)	(5)	(6)
Men						
$\hat{\rho}$	0.679	0.582	0.460	0.436	0.426	0.433
$\hat{\rho}_{AKM}$, OLS	-0.098	0.042	-0.020	-0.056	-0.080	-0.080
$\hat{\rho}_{AKM}$, AGSU	-0.043	0.152	0.098	0.069	0.061	0.074
$\hat{\rho}_{AKM}$, KSS	0.048	0.165	0.106	0.082	0.073	0.088
$\hat{\rho}_{AKM}$, BLM	0.259	0.350	0.300	0.309	0.267	0.291
number of workers (thousands)	3,459	2,800	1,577	908	907	906
number of firms (thousands)	186	302	229	166	166	169
number of observations (thousands)	6,918	5,601	3,154	1,816	1,815	1,811
share of observations top-coded	0.0550	0.086	0.068	0.027	0.010	0.014
Women						
$\hat{\rho}$	0.595	0.378	0.386	0.434	0.426	0.438
$\hat{\rho}_{AKM}$, OLS	-0.149	-0.053	-0.055	-0.048	-0.074	-0.070
$\hat{\rho}_{AKM}$, AGSU	-0.104	0.046	0.101	0.162	0.153	0.172
$\hat{\rho}_{AKM}$, KSS	-0.022	0.048	0.103	0.173	0.165	0.187
$\hat{\rho}_{AKM}$, BLM	0.122	0.254	0.292	0.364	0.362	0.341
number of workers (thousands)	2,816	2,409	1,425	738	738	736
number of firms (thousands)	172	311	241	158	157	159
number of observations (thousands)	5,631	4,818	2,849	1,476	1,476	1,473
share of observations top-coded	0.015	0.030	0.025	0.010	0.004	0.005
independence assumption	I	II	III	IV	IV	IV
observations included	all	all	longest	longest	first	last

Table 5: Estimates of the correlation ρ and the AKM correlation using different methods, using AMDB 1986–2018. The initial data construction is identical to Table 2 but we further restrict the sample to the largest leave-one-out connected set where every worker has exactly two distinct employers, and every firm employs at least two workers. The estimates are not weighted by duration. $\hat{\rho}$ is our measure of sorting. $\hat{\rho}_{AKM}$, OLS is the OLS estimate of the AKM correlation. $\hat{\rho}_{AKM}$, AGSU uses the Andrews, Gill, Schank, and Upward (2008) bias correction. $\hat{\rho}_{AKM}$, KSS uses the Kline, Saggio, and Sølvsten (2019) correction. $\hat{\rho}_{AKM}$, BLM uses the mixture model in Bonhomme, Lamadon, and Manresa (2019).

reasons to prefer ρ to ρ_{AKM}. In the latter case, there are pragmatic reasons for estimating ρ, since we know how to do it.

8 Conclusion

This paper proposes a new measure of sorting, the correlation between a worker's average wage and her employer's average wage. We find that this measure performs at least as well as the AKM correlation in a variety of structural models. We then propose an estimator of the correlation that is consistent in data sets with many workers and firms but few observations for each worker and firm. Using Austrian data, we find strong evidence for sorting between high wage workers and high wage firms. The correlation between a worker's type and her employer's type exceeds 0.4 and has been growing over time for men but shrinking for women. This contrasts with the previous literature, which has used the AKM approach and concluded that there is little sorting of high wage workers into high wage jobs. Whether this reflects theoretical limitations of the AKM measure of sorting or biases in estimating the AKM measure of sorting remains an open question.

We have focused in this paper on measuring sorting between workers and firms, but our empirical approach may be useful in a variety of other settings. For example, our approach can measure sorting between firms and banks in the corporate loan market. It can measure sorting of innovators into teams, as well as sorting between innovators and innovating firms. Our approach can detect whether schools assign better students to better teachers. In short, our methodology applies to any setting where we know the identity of both parties in a match and can observe an outcome for the pair.

Is our measured correlation large? This is a quantitative question that goes beyond the scope of this paper. Still, there are reasons to think that our approach understates the true extent of sorting. We have already mentioned three such reasons: we focus only on workers who experience unemployment, while those who are continuously employed appear to have a higher correlation; workers' types change over time, arguably more dramatically during a spell of registered unemployment (Ljungqvist and Sargent, 1998); and firms are collections of heterogeneous jobs at a point in time and so there is not really a single firm type that is applicable to all workers. Even in a frictionless environment, one would not expect to see many firms that only hire high wage workers, since real-world production processes and hierarchies utilize a mix of skills (Garicano, 2000).

In closing, we mention one more reason why we may be understating the extent of sorting: our approach focuses only on vertical sorting as captured through wages. Horizontal sorting of similarly-paid workers with different skills is likely also important (Lindenlaub, 2017;

Lindenlaub and Postel-Vinay, 2017). Our estimated correlations therefore suggest that the labor market effectively gets the highest wage workers together at the highest wage firms.

References

Abowd, John M., Robert H. Creecy, and Francis Kramarz (2002). "Computing person and firm effects using linked longitudinal employer-employee data". Center for Economic Studies, US Census Bureau.

Abowd, John M., Francis Kramarz, Paul Lengermann, and Sébastien Pérez-Duarte (2004). "Are Good Workers Employed by Good Firms? A Test of a Simple Assortative Matching Model for France and the United States". Mimeo.

Abowd, John M., Francis Kramarz, and David N. Margolis (1999). "High Wage Workers and High Wage Firms". *Econometrica* 67.2, pp. 251–333.

Andrews, Martyn J., Leonard Gill, Thorsten Schank, and Richard Upward (2008). "High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias?" *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171.3, pp. 673–697.

Bagger, Jesper, François Fontaine, Fabien Postel-Vinay, and Jean-Marc Robin (2014). "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics". *American Economic Review* 104.6, pp. 1551–1596.

Bagger, Jesper and Rasmus Lentz (forthcoming). "An Equilibrium Model of Wage Dispersion and Sorting". *Review of Economic Studies*.

Bagger, Jesper, Kenneth L Sørensen, and Rune Vejlin (2013). "Wage Sorting Trends". *Economics Letters* 118.1, pp. 63–67.

Barth, Erling and Harald Dale-Olsen (2003). "Assortative Matching in the Labour Market? Stylised Facts about Workers and Plants". Mimeo.

Becker, Gary S. (1973). "A Theory of Marriage: Part I". *Journal of Political Economy* 81.4, pp. 813–846.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa (2019). "A Distributional Framework for Matched Employer Employee Data". *Econometrica* 87.3, pp. 699–738.

Burdett, Kenneth and Dale Mortensen (1998). "Wage Differentials, Employer Size, and Unemployment". *International Economic Review* 39.2, pp. 257–73.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline (2018). "Firms and Labor Market Inequality: Evidence and Some Theory". *Journal of Labor Economics* 36.S1, S13–S70.

Card, David, Jörg Heining, and Patrick Kline (2013). "Workplace Heterogeneity and the Rise of West German Wage Inequality". *Quarterly Journal of Economics* 128.3, pp. 967–1015.

Christensen, Bent Jesper, Rasmus Lentz, Dale T. Mortensen, George R. Neumann, and Axel Werwatz (2005). "On-the-Job Search and the Wage Distribution". *Journal of Labor Economics* 23.1, pp. 31–58.

Eeckhout, Jan and Philipp Kircher (2011). "Identifying Sorting—In Theory". *Review of Economic Studies* 78.3, pp. 872–906.

Garicano, Luis (2000). "Hierarchies and the Organization of Knowledge in Production". *Journal of Political Economy* 108.5, pp. 874–904.

Goussé, Marion, Nicolas Jacquemet, and Jean-Marc Robin (2017). "Marriage, Labor Supply, and Home Production". *Econometrica* 85.6, pp. 1873–1919.

Gruetter, Max and Rafael Lalive (2009). "The Importance of Firms in Wage Determination". *Labour Economics* 16.2, pp. 149–160.

Hagedorn, Marcus, Tzuo Hann Law, and Iourii Manovskii (2017). "Identifying Equilibrium Models of Labor Market Sorting". *Econometrica* 85.1, pp. 29–65.

Iranzo, Susana, Fabiano Schivardi, and Elisa Tosetti (2008). "Skill Dispersion and Firm Productivity: An Analysis with Employer-Employee Matched Data". *Journal of Labor Economics* 26.2, pp. 247–285.

Jarosch, Gregor (2015). "Searching for Job Security and the Consequences of Job Loss". Mimeo.

Jochmans, Koen and Martin Weidner (2019). "Fixed-Effect Regressions on Network Data". *Econometrica* 87.5, pp. 1543–1560.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten (2019). *Leave-out Estimation of Variance Components*. NBER Working Paper No. 26244.

Lindenlaub, Ilse (2017). "Sorting Multidimensional Types: Theory and Application". *Review of Economic Studies* 84.2, pp. 718–789.

Lindenlaub, Ilse and Fabien Postel-Vinay (2017). "Multidimensional Sorting Under Random Search". mimeo.

Lise, Jeremy, Costas Meghir, and Jean-Marc Robin (2016). "Matching, Sorting and Wages". *Review of Economic Dynamics* 19, pp. 63–87.

Ljungqvist, Lars and Thomas J. Sargent (1998). "The European Unemployment Dilemma". *Journal of Political Economy* 106.3, pp. 514–550.

Lopes de Melo, Rafael (2018). "Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence". *Journal of Political Economy* 126.1, pp. 313–346.

Malmberg, Hannes and Ola Hössler (2014). "Probabilistic Choice with an Infinite Set of Options: An Approach Based on Random Sup Measures". *Modern Problems in Insurance Mathematics*. Ed. by Dmitrii Silvestrov and Anders Martin-Löf. Springer. Chap. 18, pp. 291–312.

Pichelmann, Karl and Monika Riedel (1992). "New Jobs or Recalls?" *Empirica* 19.2, pp. 259–274.

Postel-Vinay, Fabien and Jean-Marc Robin (2002). "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity". *Econometrica* 70.6, pp. 2295–2350.

— (2006). "Microeconometric Search-Matching Models and Matched Employer-Employee Data". *The Proceedings of the 9th World Congress of the Econometric Society*, pp. 877–907.

Shimer, Robert and Lones Smith (2000). "Assortative Matching and Search". *Econometrica* 68.2, pp. 343–369.

Solomon, Lewis C. (1975). "The Definition of College Quality and Its Impact on Earnings". *Explorations in Economic Research, volume 2, number 4*. NBER, pp. 537–587.

Tiebout, Charles M. (1956). "A Pure Theory of Local Expenditures". *Journal of Political Economy* 64.5, pp. 416–424.

Zweimuller, Josef, Rudolf Winter-Ebmer, Rafael Lalive, Andreas Kuhn, Jean-Philipe Wuellrich, Oliver Ruf, and Simon Buchi (2009). *Austrian Social Security Database*. Mimeo.

Appendix

A Details of Structural Models

A.1 A Statistical Model

We provide the proof of Proposition 1 omitted from the main text.

Proof of Proposition 1. First recall that one can obtain the best linear prediction of the expected value of y conditional on x by running a least square regression of y on x,

$$\mathbb{E}_x(y) = \bar{y} + \frac{c_{AKM}}{\sigma_x^2}(x - \bar{x}),$$

where $\bar{x}$ and $\bar{y}$ denote the expected values of x and y. Since by assumption the expectation of x conditional on y is linear in y, this relationship is exact. We can then use the definition of λ in equation (1) directly to write

$$\lambda_i = \mathbb{E}_{x_i} w(x_i, y) = x_i + \mathbb{E}_{x_i}(y) = \bar{y} - \frac{c_{AKM}}{\sigma_x^2}\bar{x} + \left(1 + \frac{c_{AKM}}{\sigma_x^2}\right) x_i, \tag{26}$$

and so λ_i is an affine function of x_i. A symmetric argument establishes that $\mathbb{E}_y(x) = \bar{x} + \frac{c_{AKM}}{\sigma_y^2}(y - \bar{y})$ and

$$\mu_j = \bar{x} - \frac{c_{AKM}}{\sigma_y^2}\bar{y} + \left(1 + \frac{c_{AKM}}{\sigma_y^2}\right) y_j. \tag{27}$$

This verifies that μ_j is an affine function of y_j.

Next, assume that $\min\{\sigma_x^2, \sigma_y^2\} > -c_{AKM}$. Equations (26) and (27) imply λ_i is an increasing function of x_i and μ_j is an increasing function of y_j. Therefore the correlation between λ and μ is the same as the correlation between x and y, $\rho = \rho_{AKM}$. Moreover, equations (26) and (27) imply that the standard deviations of λ and y are $\sigma_\lambda = \sigma_x + \rho_{AKM}\sigma_y$ and $\sigma_\mu = \sigma_y + \rho_{AKM}\sigma_x$, both positive by the assumption at the start of this paragraph. Using this and $\rho = \rho_{AKM}$ gives us $\sigma_\lambda^2 - c = \sigma_\lambda^2(1 - \rho\sigma_\mu/\sigma_\lambda) = \sigma_x\sigma_\lambda(1 - \rho_{AKM}^2) > 0$, and symmetrically $\sigma_\mu^2 - c = \sigma_y\sigma_\mu(1 - \rho_{AKM}^2) > 0$. Hence $\min\{\sigma_\lambda^2, \sigma_\mu^2\} > c$.

Alternatively, suppose that $\sigma_x^2 > -c_{AKM} > \sigma_y^2$. Then λ_i is an increasing function of x_i and μ_j is a decreasing function of y_j. Therefore $\rho = -\rho_{AKM}$. Equations (26) and (27) imply that the standard deviations of λ and μ are $\sigma_\lambda = \sigma_x + \rho_{AKM}\sigma_y$ and $\sigma_\mu = -(\sigma_y + \rho_{AKM}\sigma_x)$. Using this and $\rho = -\rho_{AKM}$ gives us $\sigma_\lambda^2 - c = \sigma_\lambda^2(1 - \rho\sigma_\mu/\sigma_\lambda) = \sigma_x\sigma_\lambda(1 - \rho_{AKM}^2) > 0$ and $\sigma_\mu^2 - c = -\sigma_y\sigma_\mu(1 - \rho_{AKM}^2) < 0$. Hence $\min\{\sigma_\lambda^2, \sigma_\mu^2\} < c$. The case with $\sigma_y^2 > -c_{AKM} > \sigma_x^2$ is analogous. It cannot be the case that $-c_{AKM} > \max\{\sigma_x^2, \sigma_y^2\}$. Thus whenever $\min\{\sigma_x^2, \sigma_y^2\} <$

$-c_{AKM}$, $\min\{\sigma_\lambda^2, \sigma_\mu^2\} < c$.

Finally, if $\sigma_x^2 + c_{AKM} = 0$, equations (26) and (27) imply $\sigma_\lambda = 0$. If $\sigma_y^2 + c_{AKM} = 0$, then $\sigma_\mu = 0$. In these cases, the correlation between λ and μ is undefined. ■

A.2 Two-Sided Search Model with Match-Specific Shocks

We formulate equations for the value functions $U(x)$ and $V(y)$ and steady state conditions for $u(x)$ and $v(y)$. It will be useful to define the conditional expected value ω and survivor function p as

$$\omega(k) = \frac{\int_k^\infty z d\zeta(z)}{1-\zeta(k)} \text{ if } \zeta(k) < 1,\ \omega(k) = k \text{ otherwise}$$
$$p(k) = 1 - \zeta(k).$$

The value of being unemployed is then

$$\begin{aligned} rU(x) &= \theta \int_Y \left(\int_{z \geq \bar{z}(x,y)} \frac{\gamma}{r+\delta} \big(zH(x,y) - rU(x) - rV(y) \big) d\zeta(z) \right) v(y) d\tilde{G}(y) \\ &= \frac{\theta\gamma}{r+\delta} \int_Y p(\bar{z}(x,y)) \big(\omega(\bar{z}(x,y)) H(x,y) - rU(x) - rV(y) \big) v(y) d\tilde{G}(y). \end{aligned} \tag{28}$$

Similarly, the value of a vacant firm is

$$rV(y) = \frac{\theta(1-\gamma)}{r+\delta} \int_X p(\bar{z}(x,y)) \big(\omega(\bar{z}(x,y)) H(x,y) - rU(x) - rV(y) \big) u(x) d\tilde{F}(x). \tag{29}$$

Finally, the steady state conditions for unemployment and vacancy rates are

$$\delta(1-u(x)) = \theta u(x) \int_Y p(\bar{z}(x,y)) v(y) d\tilde{G}(y), \tag{30}$$
$$\delta(1-v(y)) = \theta v(y) \int_X p(\bar{z}(x,y)) u(x) d\tilde{F}(x). \tag{31}$$

Using this, we can find the conditional distribution of jobs for each worker:

$$\Phi_x(y) = \frac{p(\bar{z}(x,y)) v(y) d\tilde{G}(y)}{\int_Y p(\bar{z}(x,y')) v(y') d\tilde{G}(y')}. \tag{32}$$

The log wage in an (x, y) match with productivity shock z is

$$w(x,y,z) = \log \Big(\gamma \big(zH(x,y) - rU(x) - rV(y) \big) + rU(x) \Big), \tag{33}$$

and hence its expectation is

$$w^e(x,y) \equiv \mathbb{E}w(x,y,z) = \int_{z \geq \bar{z}(x,y)} w(x,y,z) d\zeta(z)/p(\bar{z}(x,y)). \tag{34}$$

This is enough information to compute ρ and ρ_{AKM} from the model.

Finally, note that if ζ has an exponential distribution, the expected log wage is

$$w^e(x,y) = e^{\frac{rU(x)}{\gamma s H(x,y)}} \int_{\frac{rU(x)}{\gamma s H(x,y)}}^{\infty} \frac{1}{te^t} dt + \log(rU(x)),$$

which is increasing in $H(x,y)$. Thus if the production technology is monotonic in y, the expected log wage is also monotonic in y for fixed x.[29]

A.3 Discrete Choice Model

We have closed-form formulas for all objects of interest when the distributions of worker and firm characteristics are normal, $x \sim N(m_x, \sigma_x^2)$ and $y \sim N(m_y, \sigma_y^2)$. In the interest of space, we show formulae for the standard normal case, $m_x = m_y = 0$ and $\sigma_x = \sigma_y = 1$.

Theorem 18.4 in Malmberg and Hössler (2014) implies that the distribution of firm types y conditional on worker's type x is

$$\Phi_x(y) = \frac{\exp\left(\frac{w(x,y)}{s}\right) d\tilde{G}(y)}{\int_{-\infty}^{\infty} \exp\left(\frac{w(x,y')}{s}\right) d\tilde{G}(y')}. \tag{35}$$

Under the assumption that y has a standard normal distribution and $w(x,y) = x - (x-y)^2/a$, we get that $\Phi_x(y)$ is also normal with mean $\frac{2x}{2+as}$ and variance $\frac{as}{2+as}$. The distribution of x and the distribution of y conditional on x allow us to compute the joint distribution of x and y. Importantly, the correlation is $\frac{2}{\sqrt{4+2as+a^2s^2}}$.

Next, understanding the wage and the joint distribution of x and y, we can compute the types from equations (1) and (2):

$$\lambda(x) = -\frac{as^2x^2}{(2+as)^2} + x - \frac{s}{2+as}, \tag{36}$$

$$\mu(y) = -\frac{a^3s^4y^2}{(4+2as+a^2s^2)^2} + \frac{2(2+as)y}{4+2as+a^2s^2} - \frac{s(2+as)}{4+2as+a^2s^2}, \tag{37}$$

[29]The same is true for the expected wage.

as well as the AKM fixed effects from equations (8) and (9):

$$\alpha(x) = -\frac{s(4+a^2s^2)x^2}{(2+as)(8+2as+a^2s^2)} + x + \alpha_0, \tag{38}$$

$$\psi(y) = -\frac{s(2-as)y^2}{8+2as+a^2s^2} - \frac{4s}{8+2as+a^2s^2} - \alpha_0. \tag{39}$$

Note that λ, μ, and α are all concave quadratic functions of the characteristic, while ψ is a concave quadratic if $as < 2$, a convex quadratic if $as > 2$, and a constant if $as = 2$.

In the last step, we can use the covariance matrix of x and y as well as these functional forms to find the covariance matrix of λ and μ as well as α and ψ. The exact formulae are messy even in the standard normal case and so we omit them.

B Properties of Estimator

B.1 Consistency Proofs

Proof of Proposition 2. Take worker i with characteristics x_i:

$$\lambda(x_i) = \frac{\mathbb{E}_{x_i}\sum_{m=1}^{M_i} t_{i,m}^w w_{i,m}^w}{\bar{T}_x^w} = \bar{w}_{x_i}^w + \frac{\mathbb{E}_x \sum_{m=1}^{M_i} t_{i,m}^w \varepsilon_{i,m}^w}{\mathbb{E}_{x_i} T_i^w} = \bar{w}_{x_i}^w.$$

The first equation is the definition of λ as the expected daily log earnings, equation (14). The second uses the auxiliary assumption that $w_{i,m}^w = \bar{w}_{x_i}^w + \varepsilon_{i,m}^w$. The third uses the auxiliary assumption that the expected value of $t_{i,m}^w \varepsilon_{i,m}^w$ is zero. The proof for firms is identical. ■

Proof of Proposition 3. We start by proving that $\frac{1}{\tau I}\sum_{i=1}^{\tau I} T_i^w \hat{\lambda}_i$ is a consistent estimator of $\frac{1}{I}\sum_{x\in X} I_x \bar{T}_x^w \lambda(x)$, i.e. the product of worker type and time spent matched. We do this in two steps. First, for any given worker i with characteristic x_i,

$$T_i^w \hat{\lambda}_i = \frac{T_i^w \sum_{m=1}^{M_i} w_{i,m}^w}{M_i} = T_i^w \bar{w}_{x_i}^w + \frac{\sum_{m=1}^{M_i}\sum_{m'=1}^{M_i} t_{i,m'}^w \varepsilon_{i,m}^w}{M_i} = T_i^w \lambda(x_i) + \upsilon_{1,i}$$

where $\upsilon_{1,i} \equiv \frac{1}{M_i}\sum_{m=1}^{M_i}\sum_{m'=1}^{M_i} t_{i,m'}^w \varepsilon_{i,m}^w$. The first equation uses the definition of $\hat{\lambda}_i$ from equation (18). The second uses the auxiliary assumption that $w_{i,m}^w = \bar{w}_{x_i}^w + \varepsilon_{i,m}^w$ and also writes $T_i^w = \sum_{m'=1}^{M_i} t_{i,m'}^w$. The third uses $\lambda(x_i) = \bar{w}_{x_i}^w$ (Proposition 2) and defines the error term $\upsilon_{1,i}$. Since $t_{i,m}^w$ and $\varepsilon_{i,m'}^w$ are independent for all m and m' and $\varepsilon_{i,m'}$ has mean zero, $\upsilon_{1,i}$ also has mean zero for each i. It also has a finite characteristic-dependent variance, say $\sigma_{\upsilon_1,x_i}^2 < \infty$, since $\sigma_{x_i}^w$ is finite and durations are bounded.

Summing these up implies that the expected value of $\frac{1}{\tau I}\sum_{i=1}^{\tau I} T_i^w \hat{\lambda}_i$ is $\frac{1}{I}\sum_{x\in X} I_x \bar{T}_x^w \lambda(x)$. Next, the fourth auxiliary assumption implies that the error terms $\upsilon_{1,i}$ are independent. Thus the variance of $\frac{1}{\tau I}\sum_{i=1}^{\tau I} T_i^w \hat{\lambda}_i$ is $\frac{1}{\tau I^2}\sum_{x\in X} I_x \sigma_{\upsilon,x}^2$. This converges to zero when τ goes to infinity and so consistency follows from Chebyshev's inequality, a law of large numbers.

A similar argument implies that $\frac{1}{\tau I}\sum_{i=1}^{\tau I} T_i^w$ is a consistent estimator of $\frac{1}{I}\sum_{x\in X} I_x \bar{T}_x^w$ since T_i^w is an unbiased estimator of $\bar{T}_x^w$ with a finite variance and durations are independent across workers conditional on type. Since the ratio of two consistent estimators is consistent, it follows that $\hat{\bar{w}} = \frac{\sum_{i=1}^{\tau I} T_i^w \hat{\lambda}_i}{\sum_{i=1}^{\tau I} T_i^w}$ is a consistent estimator of $\bar{w} = \frac{\sum_{x\in X} I_x \bar{T}_x^w \lambda(x)}{\sum_{x\in X} I_x \bar{T}_x^w}$.

Turn next to the second moment. As above, for worker i with characteristic x_i,

$$
\begin{aligned}
T_i^w \widehat{\lambda_i^2} &= \frac{T_i^w \sum_{m=1}^{M_i} \sum_{m'\neq m} w_{i,m}^w w_{i,m'}^w}{M_i(M_i-1)} \\
&= T_i^w (\bar{w}_{x_i}^w)^2 + \frac{2\bar{w}_{x_i}^w \sum_{m=1}^{M_i}\sum_{m'=1}^{M_i} t_{i,m'}^w \varepsilon_{i,m}^w}{M_i} + \frac{\sum_{m=1}^{M_i}\sum_{m'\neq m}\sum_{m''=1}^{M_i} t_{i,m''}^w \varepsilon_{i,m}^w \varepsilon_{i,m'}^w}{M_i(M_i-1)} \\
&= T_i^w \lambda(x_i)^2 + \upsilon_{2,i}
\end{aligned}
$$

where $\upsilon_{2,i}$ is the sum of the last two terms on the previous line. The logic is very similar to the first moment. The first equation uses the definition of $\widehat{\lambda_i^2}$ in equation (18), the second uses the auxiliary assumption that $w_{i,m}^w = \bar{w}_{x_i}^w + \varepsilon_{i,m}^w$ and also writes $T_i^w = \sum_{m'=1}^{M_i} t_{i,m'}^w$. The third uses $\lambda(x_i) = \bar{w}_{x_i}^w$ and defines another error term for each worker. For each worker, the expected value of $\upsilon_{2,i}$ is zero because of the same assumptions as for the first moment, as well as the assumption that $\varepsilon_{i,m}^w$ and $\varepsilon_{i,m'}^w$ are independent for $m \neq m'$. Moreover, the variance of the error term is characteristic dependent but finite, $\sigma_{\upsilon_2,x_i}^2 < \infty$, since $\sigma_{x_i}^w$ is finite and durations are bounded.

We can then sum up these objects, getting that the expected value of $\frac{1}{\tau I}\sum_{i=1}^{\tau I} T_i^w \widehat{\lambda_i^2}$ is $\frac{1}{I}\sum_{x\in X} I_x \bar{T}_x^w \lambda(x)^2$. Consistency again follows from the fourth auxiliary assumption, since this ensures that the error terms $\upsilon_{2,i}$ are independent across workers. Again, since the ratio of two consistent estimators is consistent, $\frac{\sum_{i=1}^{\tau I} T_i^w \widehat{\lambda_i^2}}{\sum_{i=1}^{\tau I} T_i^w}$ is a consistent estimator of $\frac{\sum_{x\in X} I_x \bar{T}_x^w \lambda(x)^2}{\sum_{x\in X} I_x \bar{T}_x^w}$, the second moment of λ.

Finally, the difference between a consistent estimator of the second moment and the square of a consistent estimator of the first moment is a consistent estimator of the variance σ_λ^2, defined in equation (16). ■

Proof of Proposition 5. We start by expanding the definition (23) of $\hat{c}_{i,m}$ using $w_{i,m'}^w =$

$\bar{w}^w_{x_i} + \varepsilon^w_{i,m'}$ and $w^f_{j,n'} = \bar{w}^f_{y_j} + \varepsilon^f_{j,n'}$ with $j = \mathbf{j}_{i,m}$:

$$\hat{c}_{i,m} = \bar{w}^w_{x_i}\bar{w}^f_{y_{i,m}} + \bar{w}^f_{y_{i,m}}\frac{\sum_{m'\neq m}\varepsilon^w_{i,m'}}{M_i-1} + \bar{w}^w_{x_i}\frac{\sum_{n'\neq \mathbf{n}_{i,m}}\varepsilon^f_{\mathbf{j}_{i,m},n'}}{N_{\mathbf{j}_{i,m}}-1} + \frac{\sum_{m'\neq m}\varepsilon^w_{i,m'}}{M_i-1}\frac{\sum_{n'\neq \mathbf{n}_{i,m}}\varepsilon^f_{\mathbf{j}_{i,m},n'}}{N_{\mathbf{j}_{i,m}}-1}.$$

Now compute the average of this across all matches:

$$\begin{aligned}
\frac{1}{\tau I}\sum_{i=1}^{\tau I}\sum_{m=1}^{M_i} t^w_{i,m}\hat{c}_{i,m} &= \frac{1}{\tau I}\sum_{i=1}^{\tau I}\Bigg(\bar{w}^w_{x_i}\sum_{m=1}^{M_i} t^w_{i,m}\bar{w}^f_{y_{i,m}} + \sum_{m=1}^{M_i} t^w_{i,m}\bar{w}^f_{y_{i,m}}\frac{\sum_{m'\neq m}\varepsilon^w_{i,m'}}{M_i-1}\\
&\quad + \sum_{m=1}^{M_i} t^w_{i,m}\bar{w}^w_{x_i}\frac{\sum_{n'\neq \mathbf{n}_{i,m}}\varepsilon^f_{\mathbf{j}_{i,m},n'}}{N_{\mathbf{j}_{i,m}}-1} + \sum_{m=1}^{M_i} t^w_{i,m}\frac{\sum_{m'\neq m}\varepsilon^w_{i,m'}}{M_i-1}\frac{\sum_{n'\neq \mathbf{n}_{i,m}}\varepsilon^f_{\mathbf{j}_{i,m},n'}}{N_{\mathbf{j}_{i,m}}-1}\Bigg)\\
&= \frac{1}{\tau I}\Bigg(\sum_{i=1}^{\tau I}\bar{w}^w_{x_i}\sum_{m=1}^{M_i} t^w_{i,m}\bar{w}^f_{y_{i,m}} + \sum_{i=1}^{\tau I}\sum_{m=1}^{M_i}\varepsilon^w_{i,m}\frac{\sum_{m'\neq m} t^w_{i,m'}\bar{w}^f_{y_{i,m'}}}{M_i-1}\\
&\quad + \sum_{j=1}^{\tau J}\sum_{n=1}^{N_j}\varepsilon^f_{j,n}\frac{\sum_{n'\neq \mathbf{n}_{i,m}} t^f_{j,n'}\bar{w}^w_{x_{j,n'}}}{N_j-1} + \sum_{i=1}^{\tau I}\sum_{m=1}^{M_i} t^w_{i,m}\frac{\sum_{m'\neq m}\varepsilon^w_{i,m'}}{M_i-1}\frac{\sum_{n'\neq \mathbf{n}_{i,m}}\varepsilon^f_{j,n'}}{N_{\mathbf{j}_{i,m}}-1}\Bigg)
\end{aligned}$$

The first equation uses the definition above of $\hat{c}_{i,m}$, while the second regroups terms. In particular, in the second term, we switch the order of summation, while in the third term we first view objects from the perspective of the firm and then switch the order of the summations. The first three auxiliary assumptions imply that the last three terms all have zero expected value and so the expected value of this expression is $\frac{1}{I}\sum_{x\in X} I_x\mathbb{E}_x\sum_{i=1}^{M_i} t^w_{i,m}\lambda(x)\mu(y_{i,m})$.

To compute the variance of the estimator, we leverage the fourth auxiliary assumption, which implies that when we square the last three terms, the only parts with a non-zero expected value are the direct squares within each term and within each worker or firm. That is, the variance of $\frac{1}{\tau I}\sum_{i=1}^{\tau I}\sum_{m=1}^{M_i} t^w_{i,m}\hat{c}_{i,m}$ is

$$\begin{aligned}
\frac{1}{\tau I^2}\Bigg(\sum_{x\in X} I_x\mathbb{E}_x\Bigg(\Bigg(\sum_{m=1}^{M_i}\varepsilon^w_{i,m}\frac{\sum_{m'\neq m} t^w_{i,m'}\bar{w}^f_{y_{i,m'}}}{M_i-1}\Bigg)^2 + \Bigg(\sum_{m=1}^{M_i} t^w_{i,m}\frac{\sum_{m'\neq m}\varepsilon^w_{i,m'}}{M_i-1}\frac{\sum_{n'\neq \mathbf{n}_{i,m}}\varepsilon^f_{j,n'}}{N_{\mathbf{j}_{i,m}}-1}\Bigg)^2\Bigg)\\
+ \sum_{y\in Y} J_y\mathbb{E}_y\Bigg(\sum_{n=1}^{N_j}\varepsilon^f_{j,n}\frac{\sum_{n'\neq \mathbf{n}_{i,m}} t^f_{j,n'}\bar{w}^w_{x_{j,n'}}}{N_j-1}\Bigg)^2\Bigg)
\end{aligned}$$

This is inversely proportional to τ and so the variance of the estimator converges to zero when τ goes to infinity, i.e. the estimator is consistent.

To finish the proof, we use the fact that $\frac{1}{\tau I}\sum_{i=1}^{\tau I} T^w_i$ is a consistent estimator of $\frac{1}{I}\sum_{x\in X} I_x\bar{T}^w_x$ (see the proof of Proposition 3) and take ratios to prove that $\frac{\sum_{i=1}^{\tau I}\sum_{m=1}^{M_i} t^w_{i,m}\hat{c}_{i,m}}{\sum_{i=1}^{\tau I} T^w_i}$ is a consistent estimator of $\frac{\sum_{x\in X} I_x \mathbb{E}_x \sum_{m=1}^{M_i} t_{i,m}^w \lambda(x)\mu(y_{i,m})}{\sum_{x\in X} I_x T_x^w}$. Finally, we have already shown in the proof of Proposition 3 that $\frac{\sum_{i=1}^{\tau I} T_i^w \hat{\lambda}_i}{\sum_{i=1}^{\tau I} T_i^w}$ is a consistent estimator of $\bar{w}$, the common mean of λ and μ. The difference between a consistent estimator of $\frac{\sum_{x\in X} I_x \mathbb{E}_x \sum_{m=1}^{M_i} t_{i,m}^w \lambda(x)\mu(y_{i,m})}{\sum_{x\in X} I_x T_x^w}$ and the square of a consistent estimator of $\bar{w}$ is a consistent estimator of the covariance. ■

B.2 Construction of Artificial Data

We generate artificial data sets from two structural models, the two-sided search model with match-specific shocks and the discrete choice model.

In the search model, we proceed as follows. For given parameter values, we solve for the steady state of the economy with a continuum of workers and firms and compute the steady state decision rules, value functions, and unemployment and vacancy rates for each worker and firm characteristic x and y. This determines $\lambda(x)$ and $\mu(y)$ for all x and y.[30] To create the sample, we choose the number of workers and firms, I and J, and assign each worker and firm its characteristic x and y according to the distributions $\tilde{F}(x)$ and $\tilde{G}(y)$.[31]

We start the search economy with some workers employed and some unemployed, respecting their characteristic-specific unemployment rates $u(x)$. For a worker i with type x_i we construct an employment history, consisting of alternating spells of employment and unemployment. An unemployment spell is characterized only by its duration, which we determine by a draw from an exponential distribution with parameter $1/\left(\theta \int_Y p(\bar{z}(x,y))v(y)d\tilde{G}(y)\right)$, the reciprocal of the job finding rate of worker i. An employment spell is characterized by four objects: the firm's characteristic y, its identity j, the log wage w, and the match duration τ. We draw the firm's characteristic y from the equilibrium distribution of matches conditional on the worker type $\Phi_{x_i}(y)$, equation (32). Next, we draw the firm's identity at random from the set of firms with that characteristic y. We then draw the match-specific shock z from the distribution $d\zeta(z)$ conditional on $z \geq \bar{z}(x_i, y)$, which guarantees that a worker with x_i and a firm with y indeed want to form a match. Finally, we construct the log wage $w(x_i, y, z)$ using equation (33). The duration of the match is determined by a draw from the exponential distribution with parameter $1/\delta$. We assume that we observe each worker for $\bar{T}$ periods and keep creating spells of employment and unemployment until the sum of all spell durations

[30]We use λ and μ from an economy with a continuum of workers and firms. This is because we do not know how to solve a two-sided matching model with a finite number of workers and firms. In particular, the state variable in the finite econmy is the distribution of all matches and the finite economy does not have a steady state.

[31]We choose I and J to be multiples of 500, which is the number of characteristics in our numerical solution of the model. We assign equal number of firms and workers to each characteristic. Thus, we do not have randomness at this stage.

reaches $\bar{T}$. We choose $\bar{T}$ such that the median worker holds 4.9 jobs. If a worker has multiple jobs with the same firm, we keep one at random. We drop all workers and firms with only one observation.

We proceed similarly in the discrete choice model. For given parameter values, we first compute $\lambda(x)$ and $\mu(y)$ under the assumption that there is a continuum of workers and firms. That is, we use a generalization of equation (36) which relaxes the assumptions that $m_x = m_y = 0$ and $\sigma_x = \sigma_y = 1$. To create the sample, we choose the number of workers and firms, I and J, and draw characteristics for each from the distributions $F(x)$ and $\tilde{G}(y)$. For worker i with characteristic x_i, we draw the number of jobs M_i using the actual distribution of jobs per worker, corresponding to column (3) for men in Table 2. For each job, we assume that the matching probabilities solve a version of equation (35).[32] In the last step, we assign each match the log wage $w(x_i, y_j)$ and a duration of 1 period. Again, if a worker has multiple jobs with the same firm, we keep one at random. We drop all workers and firms with only one observation.

B.3 Monte Carlo Confidence Intervals

We construct B data samples for each model and set of parameters. For each sample $b \in \{1, \dots, B\}$, we first use the types λ and μ and the match durations and matching network realized in sample b to recover the variance-covariance matrix (equations 16 and 17) and hence the true correlation ρ_b for that sample. Alternatively, we use log wages, match durations and the matching network in sample b to find the variance-covariance matrix using formulae (20), (22), and (24), and hence recover the feasible estimate $\hat{\rho}_b$. Let $e_b = \hat{\rho}_b - \rho_b$ be the estimation error in sample b. We find values $\underline{e}$ and $\bar{e}$ such that

$$P(e_b \leq \underline{e}) = 0.025 \text{ and } P(e_b > \bar{e}) = 0.025.$$

The 95 percent confidence interval for ρ is $[\rho + \underline{e}, \rho + \bar{e}]$, where ρ is the correlation in an infinite sample. Note that the interval does not have to be centered.

In addition to the numerical results in Table 1, here we show results with other parameter values. Throughout we assume there are $I = 10,000$ workers and $J = 2,000$ firms, although the number of workers and firms in the final sample is lower because we only keep workers and firms with at least two observations. This is a rather conservative choice of the sample size, orders of magnitude lower than a typical real-world data. Figure 4 shows the confidence

[32] In contrast to the search model, we can solve the discrete choice model with a finite number of workers and firms. There is no analytical solution for the matching probabilities, but we can find them using Monte Carlo. In practice, this makes little quantitative difference for our results. For example, the correlation between the continuous version of λ and its finite counterpart typically exceeds 0.999.

intervals in the search model with match-specific productivity shocks. We see that intervals are very tight across the entire range of parameter values we consider. The only exception is the case when sorting between x and y characteristics in the model changes from positive to negative (ξ is close to 2.8). In this case, all firms pay similar wages and hence the variance of firm types is very small, $\sigma_\mu^2/(\sigma_\lambda^2 + \sigma_\mu^2) \approx 10^{-3}$. In some bootstrap samples, the estimate of the variance of firm types and the covariance are both negative and hence we record a correlation of $\hat{\rho}_b = -1$.

Figure 5 shows the confidence intervals in the discrete choice model. The confidence intervals are again very tight in most cases. The exception is the bottom left panel when the difference in means $m_x - m_y$ exceeds 1.5. The reason is a combination of a smaller sample size and selection of firms. Even though we start off simulations with $I = 10,000$ workers and $J = 2,000$ firms, the types of workers and firms are so different that many firms end up with zero or one worker, and many workers hold all their jobs at the same firm. To satisfy the sample restrictions for our estimator, we drop these firms and workers from the sample and as a result, we end up with a smaller sample. Moreover, firms which remain in the sample tend to be those with high value of y, hence they are selected and do not reflect properly the distribution of jobs.

C Alternative Identifying Assumptions and Estimator

One of our identifying assumptions states that, conditional on characteristics, the error in the wage equation and the duration of the match are independent. Several structural models of sorting would suggest that this assumption is violated since jobs which pay higher wages tend to last longer. We can relax this assumption but only if we impose an additional assumption that the average duration of the match $\bar{t}_x^w$ is independent of worker's type $\lambda(x)$, and symmetrically, average duration of a job $\bar{t}_y^f$ is independent of firm's type $\mu(y)$. This means that higher-wage jobs can last longer but on average the expected duration of a job does not depend on worker's type. We construct a consistent estimator under these assumptions and report results in Table 6. We note that the assumption that type and average duration are independent cannot be tested.

We first introduce auxiliary assumptions and then propose an estimator. The assumptions are similar to those in Section 4.2 but are modified to reflect the changes discussed above. It will be useful to introduce the following notation: let $\bar{t}_x^w$ and $\bar{t}_x^{w,2}$ denote the first and second moment of match duration of a worker with characteristic x, and symmetrically, $\bar{t}_y^f$ and $\bar{t}_y^{f,2}$ the first and second moment of match duration for a firm with characteristic y.

1′. For worker i with characteristic x_i, $w_{i,m}^w = \bar{w}_{x_i}^w + \varepsilon_{i,m}^w$ and $\varepsilon_{i,m}^w$ is independently and

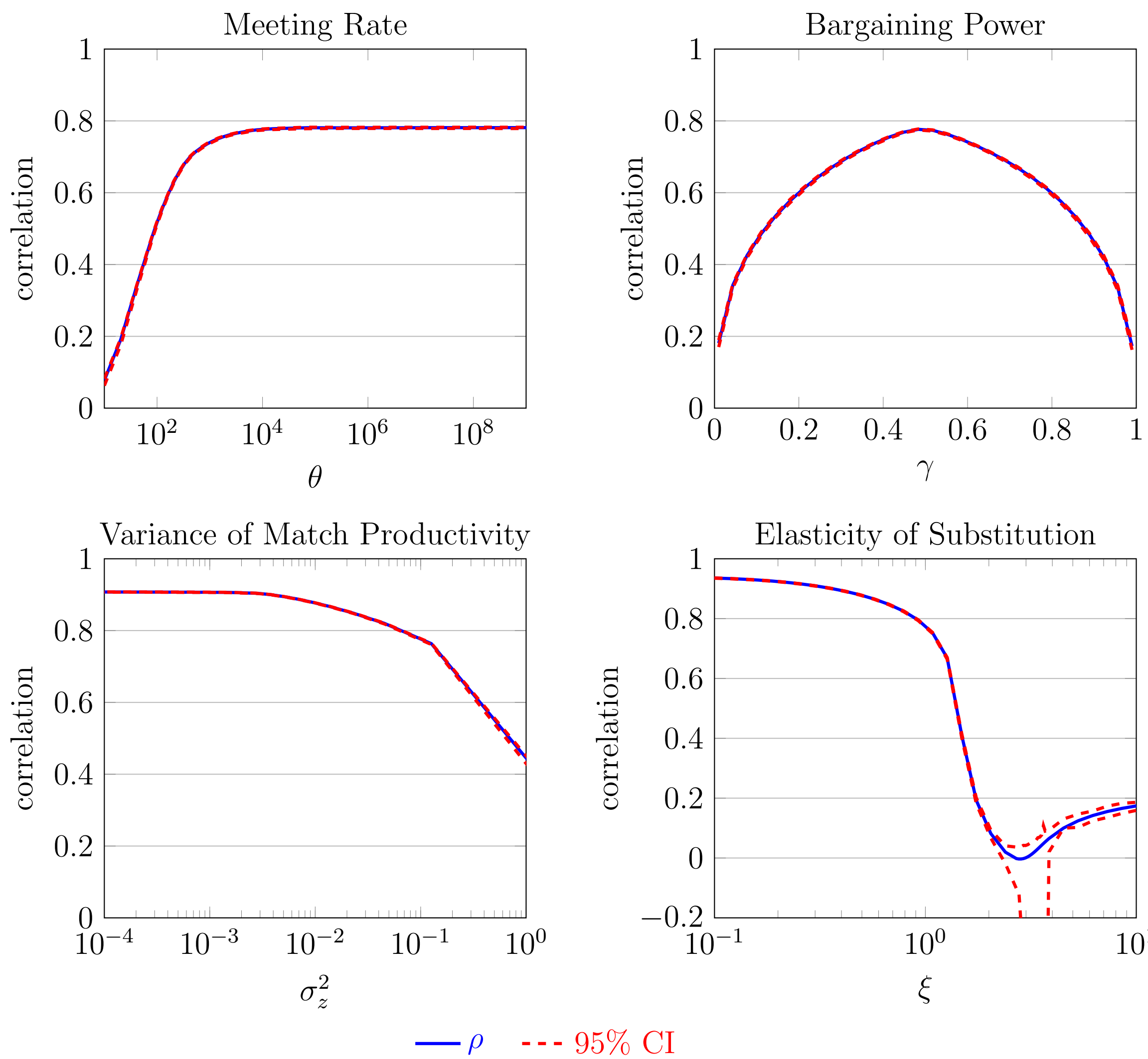

Figure 4: Confidence intervals in the two-sided search model with match-specific shocks. We plot the correlation ρ (blue line) and the bootstrapped confidence intervals (dashed red lines). In each panel, we keep all but one parameter at their benchmark values, $r = 1$, $\delta = 10$, $\theta = 10^4$, $\gamma = 0.5$, $a = 0.5$, $\xi = 1$, $\sigma_z^2 = 0.1$, $\underline{z} = 1$, $n = 500$, and depict on the horizontal axis which parameter we are changing. For the given set of parameter values, we create $B = 100$ artificial samples starting with $I = 10,000$ workers and $J = 2,000$ firms. We drop all workers and firms with only one observation. In each sample we compute the estimation error as the difference between the estimated correlation and the sample correlation, and use the 2.5% and 97.5% quartile of the error distribution to construct the 5%-confidence interval.

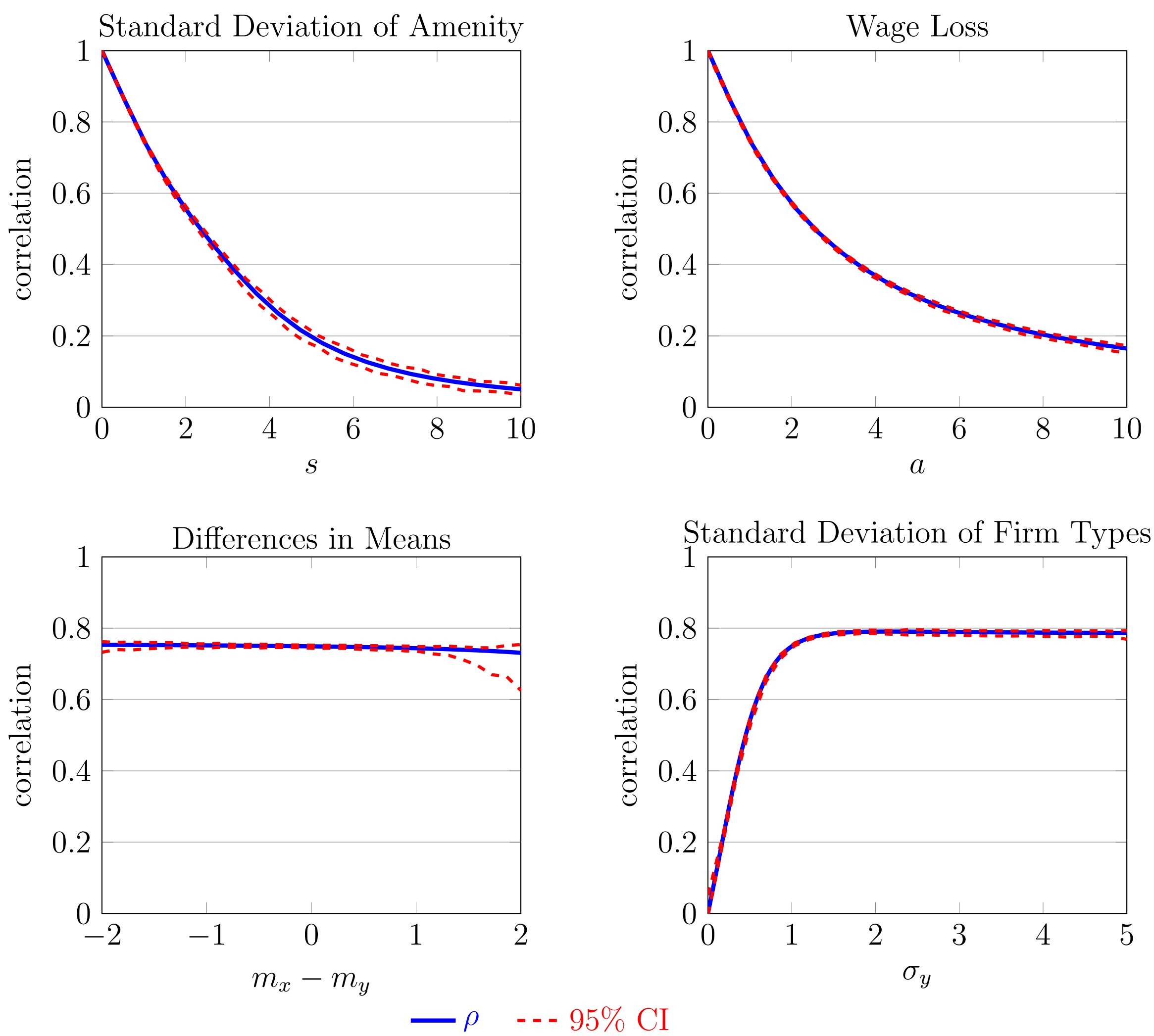

Figure 5: Confidence intervals in the discrete choice model. We plot the correlation ρ in an infinite sample (blue line) and the bootstrapped confidence intervals. In each panel, we keep all but one parameter at their benchmark values, $a = 1$, $s = 1$, $m_x = 0$, $m_y = 0$, $\sigma_x = 1$, and $\sigma_y = 1$, and depict on the horizontal axis which parameter we are changing. For the given set of parameter values, we create $B = 100$ artificial samples starting with $I = 10,000$ workers and $J = 2,000$ firms. We drop all workers and firms with only one observation. In each sample we compute the estimation error as the difference between the estimated correlation and the sample correlation, and use the 2.5% and 97.5% quartile of the error distribution to construct the 5%-confidence interval.

identically distributed across $m = \{1, \ldots, M_i\}$ with a finite standard deviation $\sigma^w_{x_i}$. Additionally, $\bar{t}^w_{x_i}$ and $\bar{t}^{w,2}_{x_i}$ are finite and independent of $\bar{w}^w_{x_i}$. Moreover, $t^w_{i,m}$ is independent of $\varepsilon^w_{i,m'}$ and $t^w_{i,m'}$ for all $m \neq m'$, but $t^w_{i,m}$ and $\varepsilon^w_{i,m}$ can be correlated.

2′. For firm j with characteristic y_j, $w^f_{j,n} = \bar{w}^f_{y_j} + \varepsilon^f_{j,n}$ and $\varepsilon^f_{j,n}$ is independently and identically distributed across $n = \{1, \ldots, N_j\}$ with a finite standard deviation $\sigma^f_{y_j}$. Additionally, $\bar{t}^f_{y_j}$ and $\bar{t}^{f,2}_{y_j}$ are finite and independent of $\bar{w}^f_{y_j}$. Moreover, $t^f_{j,n}$ is independent of $\varepsilon^f_{j,n'}$ and $t^f_{j,n'}$ for all $n \neq n'$, but $t^f_{j,n}$ and $\varepsilon^f_{j,n}$ can be correlated.

3′. For any worker i with characteristic x_i and all $m \in \{1, \ldots, M_i\}$, $\bar{w}^w_{x_i}$ and $\varepsilon^w_{i,m'}$ are independent of $\varepsilon^f_{\mathbf{j}_{i,m},n'}$ for all $m' \neq m$ and all $n' \neq \mathbf{n}_{i,m}$. Moreover, for any firm j with characteristic y_j and all $n \in \{1, \ldots, N_j\}$, $\bar{w}^f_{y_j}$ is independent of $\varepsilon^w_{\mathbf{i}_{j,n},m'}$ for all $m' \neq \mathbf{m}_{j,n}$.

4′. For all $i \neq i'$, m, and m', $\varepsilon^w_{i,m}$ and $\varepsilon^w_{i',m'}$ are independent, as are $t^w_{i,m}$ and $t^w_{i',m'}$. For all $j \neq j'$, n, and n', $\varepsilon^f_{j,n}$ and $\varepsilon^f_{j,n'}$ are independent, as are $t^f_{j,n}$ and $t^f_{j,n'}$

The first set of assumptions has been modified to relax the assumption that $\varepsilon^w_{i,m}$ and $t^w_{i,m}$ are independent, but we introduce a new assumption that first two moments of match duration are independent of worker's mean wage. The second set of assumptions introduces a symmetric change on the firm's side. The third and fourth set of assumptions remain unchanged.

Note that we do not assume that $\varepsilon^w_{i,m}$ and $\varepsilon^f_{j,n}$ have mean zero. A convenient normalization is to impose that $\mathbb{E}_{x_i}(\varepsilon^w_{i,m} t^w_{i,m}) = \mathbb{E}_{y_j}(\varepsilon^f_{j,n} t^f_{j,n}) = 0$, leaving the mean of the error terms to pick up the correlation between duration and wages. With this normalization, a worker with characteristic x has type $\lambda(x) = \bar{w}^w_x$, and a firm with characteristic y has type $\mu(y) = \bar{w}^f_y$. However, since errors in the wage equation $\varepsilon^w_{i,m}$ and $\varepsilon^f_{j,n}$ are not mean zero, $\hat{\lambda}_i$, $\hat{\mu}_j$ are no longer unbiased estimators of $\lambda(x_i)$ and $\mu(y_j)$. We therefore proceed differently, and define the following estimators:

$$\widehat{\bar{t}^w_i \lambda_i} \equiv \frac{1}{M_i} \sum_{m=1}^{M_i} t^w_{i,m} w^w_{i,m}, \qquad \widehat{\bar{t}^{w,2}_i \lambda^2_i} \equiv \frac{1}{M_i(M_i-1)} \sum_{m=1}^{M_i} \sum_{m' \neq m} (t^w_{i,m} w^w_{i,m})(t^w_{i,m'} w^w_{i,m'}),$$

$$\widehat{\bar{t}^w_i} \equiv \frac{1}{M_i} \sum_{m=1}^{M_i} t^w_{i,m}, \qquad \widehat{\bar{t}^{w,2}_i} \equiv \frac{1}{M_i(M_i-1)} \sum_{m=1}^{M_i} \sum_{m' \neq m} t^w_{i,m} t^w_{i,m'}, \tag{40}$$

and similarly for firms,

$$\widehat{\bar{t}_j^f \mu_j} \equiv \frac{1}{N_j} \sum_{n=1}^{N_j} t_{j,n}^f w_{j,n}^f, \qquad \widehat{\bar{t}_j^{f,2} \mu_j^2} \equiv \frac{1}{N_j(N_j-1)} \sum_{n=1}^{N_j} \sum_{n' \neq n} (t_{j,n}^f w_{j,n}^f)(t_{j,n'}^f w_{j,n'}^f),$$
$$\widehat{\bar{t}_j^f} \equiv \frac{1}{N_j} \sum_{n=1}^{N_j} t_{j,n}^f, \qquad \widehat{\bar{t}_j^{f,2}} \equiv \frac{1}{N_j(N_j-1)} \sum_{n=1}^{N_j} \sum_{n' \neq n} t_{j,n}^f t_{j,n'}^f. \tag{41}$$

It is straightforward to show that under auxiliary assumptions $1'$ and $2'$, these are unbiased estimators of the corresponding moments:

$$\mathbb{E}_{x_i} \widehat{\bar{t}_i^w \lambda_i} = \bar{t}_{x_i}^w \bar{w}_{x_i}^w \qquad \mathbb{E}_{x_i} \widehat{\bar{t}_i^w} = \bar{t}_{x_i}^w$$
$$\mathbb{E}_{x_i} \widehat{\bar{t}_i^{w,2} \lambda_i^2} = \bar{t}_{x_i}^{w,2} \left(\bar{w}_{x_i}^w\right)^2 \qquad \mathbb{E}_{x_i} \widehat{\bar{t}_i^{w,2}} = \bar{t}_{x_i}^{w,2}.$$

A symmetric argument holds on the firm side.

We use cross-sectional averages of the above estimators to estimate the variance of worker types, weighted again by workers' total employment duration T_i^w:

$$\widehat{\sigma_\lambda^2}^{alt} = \frac{\sum_{i=1}^{\tau I} T_i^w \widehat{\bar{t}_i^{w,2} \lambda_i^2}}{\sum_{i=1}^{\tau I} T_i^w \widehat{\bar{t}_i^{w,2}}} - \left(\frac{\sum_{i=1}^{\tau I} T_i^w \widehat{\bar{t}_i^w \lambda_i}}{\sum_{i=1}^{\tau I} T_i^w \widehat{\bar{t}_i^w}} \right)^2. \tag{42}$$

The assumption that $\bar{w}_{x_i}^w$ and $\bar{t}_{x_i}^w$ are independent, as are $\bar{t}_{x_i}^{w,2}$ and $\left(\bar{w}_{x_i}^w\right)^2$, is important here. If it is violated, then $\widehat{\sigma_\lambda^2}^{alt}$ is inconsistent due to omitted covariance terms $cov(\bar{t}_{x_i}^w, w_{x_i}^w)$ and $cov(\bar{t}_{x_i}^{w,2}, (w_{x_i}^w)^2)$. Symmetrically, an estimator of the variance of firm types is

$$\widehat{\sigma_\mu^2}^{alt} = \frac{\sum_{j=1}^{\tau J} T_j^f \widehat{\bar{t}_j^{f,2} \mu_j^2}}{\sum_{j=1}^{\tau J} T_j^f \widehat{\bar{t}_j^{f,2}}} - \left(\frac{\sum_{j=1}^{\tau J} T_j^f \widehat{\bar{t}_j^f \mu_j}}{\sum_{j=1}^{\tau J} T_j^f \widehat{\bar{t}_j^f}} \right)^2. \tag{43}$$

Finally, an estimator of the product of the firm and worker type is

$$\widehat{t_{i,m} c_{i,m}} \equiv \frac{\sum_{m' \neq m} t_{i,m'}^w w_{i,m'}^w}{M_i - 1} \frac{\sum_{n' \neq \mathbf{n}_{i,m}} t_{\mathbf{j}_{i,m},n'}^f w_{\mathbf{j}_{i,m},n'}^f}{N_{\mathbf{j}_{i,m},n'} - 1},$$
$$\widehat{t_{i,m}} \equiv \frac{\sum_{m' \neq m} t_{i,m'}^w}{M_i - 1} \frac{\sum_{n' \neq \mathbf{n}_{i,m}} t_{\mathbf{j}_{i,m},n'}^f}{N_{\mathbf{j}_{i,m},n'} - 1}, \tag{44}$$

and its cross-sectional mean is the estimator of the covariance,

$$\widehat{c}^{alt} \equiv \frac{\sum_{i=1}^{\tau I}\sum_{m=1}^{M_i} t_{i,m}^{w}\widehat{t_{i,m}c_{i,m}}}{\sum_{i=1}^{\tau I}\sum_{m=1}^{M_i} t_{i,m}^{w}\widehat{t_{i,m}}} - \left(\frac{\sum_{i=1}^{\tau I} T_i^w \widehat{\bar{t}_i^w \lambda_i}}{\sum_{i=1}^{\tau I} T_i^w \widehat{\bar{t}_i^w}}\right)\left(\frac{\sum_{j=1}^{\tau J} T_j^f \widehat{\bar{t}_j^f \mu_j}}{\sum_{j=1}^{\tau J} T_j^f \widehat{\bar{t}_j^f}}\right). \tag{45}$$

It follows that, under assumptions $1'$–$4'$, the estimators $\widehat{\sigma_\lambda^2}^{alt}, \widehat{\sigma_\mu^2}^{alt}, \widehat{c}^{alt}$ are consistent estimators of $\sigma_\lambda^2, \sigma_\mu^2, c$ as $\tau \to \infty$. Thus,

$$\widehat{\rho}^{alt} = \frac{\widehat{c}^{alt}}{\sqrt{\widehat{\sigma_\lambda^2}^{alt}\widehat{\sigma_\mu^2}^{alt}}}. \tag{46}$$

is a consistent estimator of ρ. The proof is analogous to the proof in Appendix B.1 and so we omit it.

Table 6 shows results using alternative estimators. To compute these, we use formulas (42), (43), (45) and (46) where we set $\tau = 1$. We see that results are remarkably similarly to those reported in Table 2 for all six columns; the maximum absolute difference between the estimated correlations reported in these two tables is 0.066 for men and 0.053 for women.

D Impact of Top-Coding on Estimated Correlation

We study the impact of top-coding on our estimates by varying the share of top-coded wages in the data set. We start from the wage cap in the data and then gradually decrease it by as much as fifty percent. We then censor wages at the new wage cap. Reducing the wage cap by fifty percent raises the share of top-coded observations from 4.7 percent to 46.5 percent for men and from 2.1 to 21.5 percent for women.

In Figure 6 we show the results, plotting the estimated correlation $\hat{\rho}$ for data sets with different top-coding on the vertical axis and the share of top-coded observations on the horizontal axis. For men, the estimated correlation varies very mildly, staying around 0.476 even when almost half of observations are top-coded. Top-coding matters more for women. A tighter wage cap reduces the correlation from 0.423 to 0.385 when more 20 percent of women have a top coded observation.

Our intuition is that the impact of top-coding on the estimated correlation depends on the correlation in the group affected by top-coding relative to the correlation among the rest. If the correlation is similar to the rest of the sample, then top-coding does not have a significant impact. However, if the correlation in the top-coded group is stronger, the correlation decreases after top-coding the data. It is useful to think about the components

Estimated Correlation and Variances Using Alternative Estimator

	(1)	(2)	(3)	(4)	(5)	(6)
Men						
correlation of matched types $\hat{\rho}$	0.680	0.583	0.407	0.434	0.400	0.415
covariance of matched types $\hat{c}$	0.063	0.037	0.019	0.021	0.023	0.025
variance of worker types $\widehat{\sigma_\lambda^2}$	0.104	0.063	0.039	0.043	0.050	0.054
variance of job types $\widehat{\sigma_\mu^2}$	0.081	0.065	0.054	0.053	0.066	0.067
number of workers (thousands)	3,884	2,916	1,646	971	971	970
number of firms (thousands)	705	532	343	272	273	276
number of observations (thousands)	64,359	16,401	7,211	4,883	4,881	4,880
share of observations top-coded	0.112	0.074	0.047	0.020	0.009	0.012
Women						
correlation of matched types $\hat{\rho}$	0.630	0.493	0.411	0.471	0.397	0.421
covariance of matched types $\hat{c}$	0.101	0.059	0.033	0.040	0.031	0.036
variance of worker types $\widehat{\sigma_\lambda^2}$	0.190	0.131	0.070	0.077	0.067	0.077
variance of job types $\widehat{\sigma_\mu^2}$	0.136	0.111	0.091	0.092	0.091	0.094
number of workers (thousands)	3,340	2,532	1,494	937	798	797
number of firms (thousands)	758	530	345	303	248	250
number of observations (thousands)	52,128	12,451	5,582	4,435	3,294	3,291
share of observations top-coded	0.035	0.031	0.021	0.015	0.004	0.005
independence assumption	I	II	III	IV	IV	IV
observations included	all	all	longest	longest	first	last

Table 6: Estimates of correlations, covariances, and variances between matched workers' and firms' types using AMDB 1986–2018, using alternative estimator. The data construction is identical to Table 2. See the notes to that table for details.

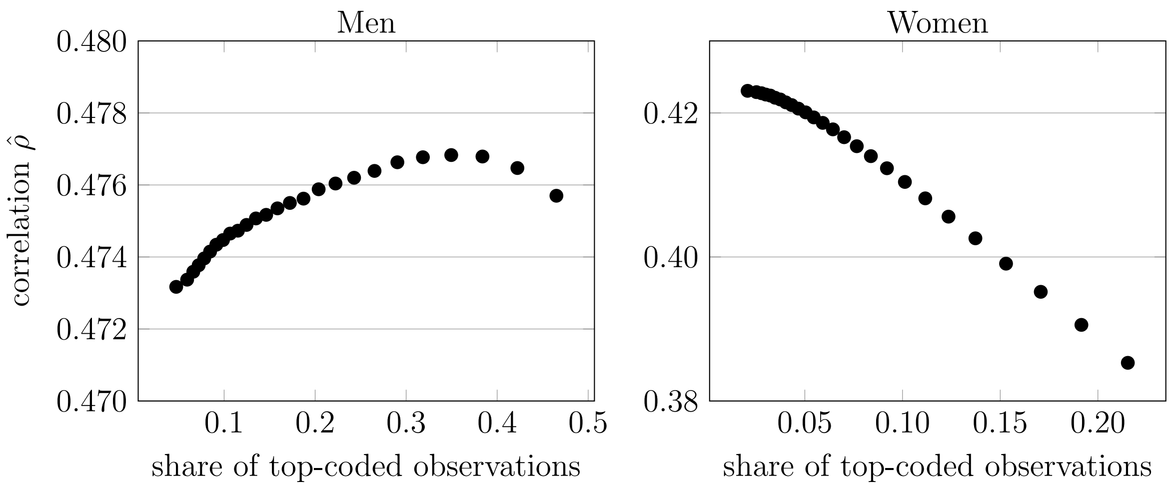

Figure 6: Impact of top-coding on the estimated correlation for men and women. Each dot corresponds to a sample where we decrease the top-code by $0, 2, 4, \ldots 50$ percent every year and truncate all wages at this new top-code. The sample of workers and firms is chosen according to independence assumption III, so the numbers are comparable to column (3) of Table 2. We plot the results as a function of the share of top-coded observations in the sample. An observation is considered top-coded if at least one wage observation of the job is top-coded.

of the correlation separately since top-coding affects the covariance as well as variances. As stricter top-coding makes wages more similar, the variance of the worker and firm types declines. The covariance can increase or decrease, as suggested earlier. We find that the covariance (not plotted) decreases with top coding from an initial 0.027 to 0.010 for men and from 0.031 to 0.018 for women when the top code is 50 percent of the top coded wage in Austria. This suggests that the covariance is larger among high-wage workers. The decline in covariance is almost exactly offset by the decline in the variances and the resulting correlation for men is barely affected. For women, the decline in covariance is stronger and hence the correlation drops modestly.

E Construction of Correlated Matching Networks

This section explains how to construct a data set that approximates a number of desired targets: the correlation between matched worker and firm types ρ, the standard deviation of worker and firm types σ_λ and σ_μ, the standard deviation of log wages σ, the number of workers and firms I and J, the distribution of the number of matches per worker M_i and per firm N_j, and the distribution of match durations.

1. We set the number of workers equal to the targets I and J.

2. For each worker $i \in \{1, \ldots, I\}$ we draw M_i and $t^w_{i,1}, \ldots t^w_{i,M_i}$, the number firms a worker works for and durations of each of his matches directly from the data. For each $j \in \{1, \ldots, J\}$, we draw the number of employees N_j. We use the distribution of N_j from the data. The model imposes the restriction that $\sum_i M_i = \sum_j N_j$. We add workers (if $\sum_i M_i < \sum_j N_j$) or firms (if $\sum_i M_i > \sum_j N_j$) until we achieve balance. We end up with $\tilde{I} \geq I$ workers and $\tilde{J} \geq J$ firms.

3. For each worker i, we choose a random λ_i from a normal distribution with mean 0 and standard deviation σ_λ.

4. For each firm j, we choose a random μ_j from a normal distribution with mean 0 and standard deviation σ_μ. We order the firms so that $\mu_1 < \mu_2 < \cdots < \mu_J$.

5. For each worker i, we choose M_i values $\chi_{i,m}$ distributed normally with mean $\lambda_i \rho \sigma_\mu / \sigma_\lambda$ and variance $\sigma_\mu^2(1 - \rho^2)$. We order these values across all i and m. The N_1 lowest values are assigned to firm 1. The next N_2 values are assigned to firm 2, etc. This gives us our matched pairs. We do not use the $\chi_{i,m}$ after this step.

6. If there are duplicate matches between i and j, we keep one at random. If this leaves us with any workers or firms with a single match, we iteratively drop those as well.

7. We measure variances, covariance and hence the correlation ρ_b using the true types λ and μ and the job durations t^w according to formulae (16), (17), and (3).

8. We construct a log wage that is consistent with worker and firm types as well as the total variance of wages. For worker i's m^{th} job, the log wage is $w^w_{i,m} = a\lambda_i + b\mu_{\mathbf{j}_{i,m}} + v_{i,m}$, where $v_{i,m}$ is an i.i.d. normal shock with mean 0 and standard deviation σ_v. The constants a and b satisfy

$$a = \frac{\sigma_\lambda - \rho\sigma_\mu}{\sigma_\lambda(1-\rho^2)} \text{ and } b = \frac{\sigma_\mu - \rho\sigma_\lambda}{\sigma_\mu(1-\rho^2)}, \tag{47}$$

and the variance of the log wage shock satisfies

$$\sigma_v^2 = \sigma^2 - \frac{\sigma_\lambda^2 + \sigma_\mu^2 - 2\rho\sigma_\lambda\sigma_\mu}{1-\rho^2}, \tag{48}$$

where σ^2 is the total variance of wages.

9. We use wage and duration data to estimate $\hat{\rho}_b$ using our estimator (25).

We claim that the data set constructed this way has the desired properties. That is, if we had an infinitely large data set, then the worker's type is indeed λ, the firm's type is μ, and the correlation between them is ρ.

To prove this, first notice that because distribution of $\chi_{i,m}$ conditional on λ_i is normal, and the distribution of λ is normal, the unconditional distribution of $\chi_{i,m}$ is also normal. The unconditional mean of $\chi_{i,m}$ is 0 by the law of iterated expectations. The expected value of $\chi_{i,m}^2$ conditional on λ_i is the conditional variance plus the square of the mean, $\sigma_\mu^2(1-\rho^2) + \frac{\lambda_i^2\rho^2\sigma_\mu^2}{\sigma_\lambda^2}$. Thus the unconditional expectation of $\chi_{i,m}^2$ is

$$\sigma_\mu^2(1-\rho^2) + \rho^2\sigma_\mu^2 = \sigma_\mu^2.$$

In short, the (unconditional) distribution of $\chi_{i,m}$ is normal with mean 0 and variance σ_μ^2. Recall that distribution of μ is normal with mean 0 and variance σ_μ^2, and therefore (in an infinitely large data set) $\chi_{i,m} = \mu_{\mathbf{j}_{i,m}}$, the type of the firm that employs i in her m^{th} match.

We next show that the λ is indeed worker's type, that is, the expected log wage of a worker with λ_i is λ_i:

$$\begin{aligned}\mathbb{E}_{\lambda_i}(w_{i,m}^w) &= \mathbb{E}_{\lambda_i}(a\lambda_i + b\mu_{\mathbf{j}_{i,m}} + v_{i,m}) = a\lambda_i + \mathbb{E}_{\lambda_i}(b\mu_{\mathbf{j}_{i,m}}) \\ &= a\lambda_i + b\frac{\lambda_i\rho\sigma_\mu}{\sigma_\lambda} = \frac{\sigma_\lambda - \rho\sigma_\mu}{\sigma_\lambda(1-\rho^2)}\lambda_i + \frac{\sigma_\mu - \rho\sigma_\lambda}{\sigma_\mu(1-\rho^2)}\frac{\lambda_i\rho\sigma_\mu}{\sigma_\lambda} = \lambda_i.\end{aligned}$$

A similar proof establishes that μ is indeed a firm type. First prove that the distribution of λ conditional on μ is normal with mean $\mu\rho\sigma_\lambda/\sigma_\mu$, and then the steps are symmetric to the steps in the argument that λ is worker's type.

Finally, we prove that the correlation in the matched pairs is indeed ρ. The expected value of $\lambda\mu$ conditional on λ is $\lambda^2\rho\sigma_\mu/\sigma_\lambda$, and thus the unconditional expected value is $\rho\sigma_\mu\sigma_\lambda$. This is the covariance between λ and μ. It then follows that the correlation is ρ. Thus our data set has all the desired properties.

F Confidence Intervals

We use a bootstrap procedure to construct standard errors. In each iteration of the bootstrap $b \in \{1, \ldots, B\}$, we follow the procedure in Appendix E to construct artificial data sets that match key moments of the real-world data: the estimated correlation between matched worker and firm types $\hat{\rho}$, the estimated variance of worker and firm types $\widehat{\sigma_\lambda^2}$ and $\widehat{\sigma_\mu^2}$, as well as the variance of log wages, the number of workers and firms, the distribution of the number of matches per worker M_i and per firm N_j, and the distribution of match durations. We take these moments from our estimates, e.g. column (3) in Table 2, and we take the distributions of the number of matches per firm and per worker and the distribution of match duration from directly from the data.

In each data set, we calculate ρ_b and $\hat{\rho}_b$ as in Appendix B.3. Again let $e_b = \hat{\rho}_b - \rho_b$ be the estimation error in sample b. We find values $\underline{e}$ and $\bar{e}$ such that

$$P(e_b \leq \underline{e}) = 0.025 \text{ and } P(e_b > \bar{e}) = 0.025.$$

The 95 percent confidence interval for ρ is $[\hat{\rho} + \underline{e}, \hat{\rho} + \bar{e}]$, not necessarily centered around $\hat{\rho}$.

Our procedure for constructing artificial data sets assumes that wages are homoscedastic conditional on worker and firm types, but it is straightforward to relax this assumption. This does not affect our confidence intervals. We have also constructed artificial data sets where types are correlated with the number of observations. In particular, we assumed that the worker types λ_i are distributed normally with a mean and variance that depends on M_i, and that the firm types μ_j are distributed normally with a mean and variance that depends on N_j. We constructed conditional distributions directly from the data by looking at the relationship between $\hat{\lambda}_i$ and M_i, and $\hat{\mu}_j$ and N_j. Our estimated confidence interval for $\hat{\rho}$ is robust to this change.

G Time-Varying Types

Consider the following model of time-varying types. Time is discrete and denoted by $t \in \{1, 2, \ldots\}$. There are a continuum of workers $i \in [0, 1]$ and a continuum of firms $j \in [0, 1]$. For simplicity, each firm hires one worker in each period. A worker i has a permanent characteristic $\bar{\lambda}_i$, distributed in the population with mean zero and variance $\bar{\sigma}_\lambda^2 > 0$. A firm j has a permanent characteristic $\bar{\mu}_j$, again distributed with mean zero and variance $\bar{\sigma}_\mu^2 > 0$.

Types are changing over time but follow an ergodic distribution with mean $\bar{\lambda}_i$ for each worker i. At time t, worker i has a type (expected wage) $\lambda_{i,t}$ with cross-sectional mean zero and cross-sectional variance $\sigma_\lambda^2 > \bar{\sigma}_\lambda^2$, the cross-sectional variance of $\bar{\lambda}_i$. Assume that the covariance of $\lambda_{i,t}$ and $\bar{\lambda}_i$ is $\bar{\sigma}_\lambda^2$ and the conditional mean of $\lambda_{i,t}$ is linear in $\bar{\lambda}_i$ and vice versa. This implies that the expected value of $\lambda_{i,t}$ given $\bar{\lambda}_i$ is $\bar{\lambda}_i$, so $\bar{\lambda}_i$ is a reasonable measure of the time-averaged type. Conversely, the expected value of $\bar{\lambda}_i$ given $\lambda_{i,t}$ is $(\bar{\sigma}_\lambda^2/\sigma_\lambda^2)\lambda_{i,t}$.

Symmetrically, at time t firm j has a type (expected wage) $\mu_{j,t}$ which has an ergodic distribution with mean $\bar{\mu}_j$. In the cross section, the mean of $\mu_{j,t}$ is zero and the variance is $\sigma_\mu^2 > \bar{\sigma}_\mu^2$. The covariance of $\mu_{j,t}$ and $\bar{\mu}_j$ is $\bar{\sigma}_\mu^2$ and the expected value of $\mu_{j,t}$ given $\bar{\mu}_j$ is $\bar{\mu}_j$ (so $\bar{\mu}_j$ is the time-averaged type), while the expected value of $\bar{\mu}_j$ given $\mu_{j,t}$ is $(\bar{\sigma}_\mu^2/\sigma_\mu^2)\mu_{j,t}$.

Next, we assume that at any time t, the correlation between matched types $\lambda_{i,t}$ and $\mu_{j,t}$ is ρ. Moreover, the expected value of $\lambda_{i,t}$ given $\mu_{j,t}$ is $\frac{\rho\sigma_\lambda}{\sigma_\mu}\mu_{j,t}$ and the expected value of $\mu_{j,t}$ given $\lambda_{i,t}$ is $\frac{\rho\sigma_\mu}{\sigma_\lambda}\lambda_{i,t}$, i.e. conditional expectations are again linear.

Finally, we assume that only the time-varying type determines who matches with whom. This implies two moment conditions. First, the distribution of $\mu_{j,t}$ conditional on $\lambda_{i,t}$ and $\bar{\lambda}_i$ does not depend on $\bar{\lambda}_i$. In particular, the difference between $\mu_{j,t}$ and its expected value conditional on $\lambda_{i,t}$ is orthogonal to $\bar{\lambda}_i$:

$$\mathbb{E}\left(\left(\mu_{j,t} - \frac{\rho\sigma_\mu}{\sigma_\lambda}\lambda_{i,t}\right)\bar{\lambda}_i\right) = 0 \Rightarrow \mathbb{E}(\mu_{j,t}\bar{\lambda}_i) = \frac{\rho\sigma_\mu}{\sigma_\lambda}\mathbb{E}(\lambda_{i,t}\bar{\lambda}_i) = \frac{\rho\sigma_\mu}{\sigma_\lambda}\bar{\sigma}_\lambda^2. \tag{49}$$

The last equation uses the fact that the covariance between $\lambda_{i,t}$ and $\bar{\lambda}_i$ is $\bar{\sigma}_\lambda^2$.

Second, think of a set of workers with the same value of $\lambda_{i,t}$ matched to a set of workers with the same value of $\mu_{j,t}$. These workers may differ in their permanent type, but we do not allow sorting on that basis. Formally, we impose that the distribution of $\bar{\mu}_j$ conditional on the time-varying type $\mu_{j,t}$ and the partner's permanent type $\bar{\lambda}_i$ does not depend on $\bar{\lambda}_i$. In particular, the difference between $\bar{\mu}_j$ and its expected value conditional on $\mu_{j,t}$ is orthogonal to $\bar{\lambda}_i$:

$$\mathbb{E}\left(\left(\bar{\mu}_j - \frac{\bar{\sigma}_\mu^2}{\sigma_\mu^2}\mu_{j,t}\right)\bar{\lambda}_i\right) \Rightarrow \mathbb{E}(\bar{\mu}_j\bar{\lambda}_i) = \frac{\bar{\sigma}_\mu^2}{\sigma_\mu^2}\mathbb{E}(\mu_{j,t}\bar{\lambda}_i) = \frac{\rho\bar{\sigma}_\lambda^2\bar{\sigma}_\mu^2}{\sigma_\lambda\sigma_\mu}.$$

The last equation uses equation (49) to eliminate $\mathbb{E}(\lambda_{i,t}\bar{\lambda}_i)$.

The correlation between the time-averaged types is the covariance $\mathbb{E}(\bar{\mu}_j\bar{\lambda}_i)$ divided by the product of the standard deviations, $\frac{\rho\bar{\sigma}_\lambda\bar{\sigma}_\mu}{\sigma_\lambda\sigma_\mu}$. Since the standard deviations of the time-averaged types are smaller than the standard deviations of the time-varying types, $\bar{\sigma}_\lambda\bar{\sigma}_\mu < \sigma_\lambda\sigma_\mu$, this is smaller than ρ, proving the result.

H Methods for Estimation of AKM Correlation

H.1 OLS Estimation of the AKM Correlation

The OLS estimates of the AKM fixed effects solve the moment conditions

$$\hat{\alpha}_i = \frac{\sum_{m=1}^{M_i}(w_{i,m}^w - \hat{\psi}_{\mathbf{j}_{i,m}})}{M_i} \text{ and } \hat{\psi}_j = \frac{\sum_{n=1}^{N_j}(w_{j,n}^f - \hat{\alpha}_{\mathbf{i}_{j,n}})}{N_j}.$$

We approximate the solution to these using a zig-zag algorithm (Guimaraes and Portugal, 2010) with a single normalization, say $\hat{\alpha}_1 = 0$. Next we compute duration-weighted average fixed effects,

$$\hat{\bar{\alpha}} \equiv \frac{\sum_i T_i^w \hat{\alpha}_i}{\sum_i T_i^w} \text{ and } \hat{\bar{\psi}} \equiv \frac{\sum_j T_j^f \hat{\psi}_j}{\sum_j T_j^f}.$$

We then use these to find duration-weighted variances and covariances:

$$\hat{\sigma}_\alpha^2 \equiv \frac{\sum_i T_i^w (\hat{\alpha}_i - \hat{\bar{\alpha}})^2}{\sum_i T_i^w}, \qquad \hat{\sigma}_\psi^2 \equiv \frac{\sum_j T_j^f (\hat{\psi}_j - \hat{\bar{\psi}})^2}{\sum_j T_j^f},$$
$$\text{and } \hat{c}_{AKM} \equiv \frac{\sum_i \sum_{m=1}^{M_i} t_{i,m}^w (\hat{\alpha}_i - \hat{\bar{\alpha}})(\hat{\psi}_j - \hat{\bar{\psi}})}{\sum_i T_i^w}.$$

Our OLS estimate of the duration-weighted AKM correlation is then $\hat{\rho}_{AKM} = \hat{c}_{AKM} / \sqrt{\hat{\sigma}_\alpha^2 \hat{\sigma}_\psi^2}$.

H.2 Bias Correction Methods for the AKM Correlation

$\hat{\sigma}_\alpha^2$ and $\hat{\sigma}_\psi^2$ are known to be biased up and $\hat{c}_{AKM}$ is known to be biased down, creating a bias in the correlation. We apply bias correction methods proposed by Andrews, Gill, Schank, and Upward (2008) and Kline, Saggio, and Sølvsten (2019) and compute the AKM correlation following Bonhomme, Lamadon, and Manresa (2019). We refer to these estimators as AGSU, KSS and BLM, respectively.

Andrews, Gill, Schank, and Upward (2008) and Kline, Saggio, and Sølvsten (2019) both derive a bias correction for the variance of the worker and firm fixed effects and their covariance but under different assumptions. Andrews, Gill, Schank, and Upward (2008) assume that errors in the wage equation are homoskedastic, while Kline, Saggio, and Sølvsten (2019) allow for heteroscedasticity. Both methods then use the bias-corrected moments to compute the AKM correlation. We note that the correlation computed this way is not unbiased, but Andrews, Gill, Schank, and Upward (2008) use simulations to show that in practice this bias is negligible.

Both Andrews, Gill, Schank, and Upward (2008) and Kline, Saggio, and Sølvsten (2019) offer exact formulae for the bias correction. However, these formulae require calculating the inverse of an $(I+J)\times(I+J)$ matrix where I is the number of workers and J is the number of firms. When $I+J$ is in the order of millions, as in our data sets, the exact formulae are not tractable. Kline, Saggio, and Sølvsten (2019) show how to use the Johnson-Lindenstrauss approximation to compute bias corrections in large data sets. The same approximation can be applied to the AGSU methodology.

Bonhomme, Lamadon, and Manresa (2019) propose a two-step procedure to estimating the correlation between the AKM fixed effects. In the first step, they use k-means clustering to classify firms into a given number of classes (ten in our case) based on their wage distributions. We follow them in using ten firm classes. In the second step, they estimate their model using maximum likelihood. Given the firm classification and a specified number of worker types, five in our specification, they assume that the log wage paid by a given firm type to a given worker type is log normal with unknown mean and standard deviation. They then project this non-linear wage equation onto a linear function of worker and firm types to obtain the AKM correlation.

In this section we estimate our correlation and the AKM correlation using different methods. Different estimators put different restrictions on the sample and so we create a sample which satisfies all these restrictions at the same time. Methods proposed by Kline, Saggio, and Sølvsten (2019) and Bonhomme, Lamadon, and Manresa (2019) are designed for short panels. In their main applications, they have exactly two wage observations per worker. AKM is identified on a connected set of firms and workers. Kline, Saggio, and Sølvsten (2019) further require that this set remains connected after dropping any worker from the set, and call it the leave-one-out connected set. Finally, recall that our estimator requires that every worker and every firm in the sample has at least two observations.

Our starting point is the sample we used for the AKM estimation in Section 7. In this sample, we keep only the first two observations for each worker to create a short panel, as in Kline, Saggio, and Sølvsten (2019) and Bonhomme, Lamadon, and Manresa (2019). We find the largest leave-one-out connected set and keep only workers and firms belonging to this set, as in Kline, Saggio, and Sølvsten (2019). There are two observations per worker and at least two observations per firm in the resulting sample, and hence it satisfies our restriction.

Using this sample, we first calculate $\hat{\rho}$ using our approach and $\hat{\rho}_{AKM}$ using OLS. Since AGSU, KSS, and BLM do not weigh spells by duration, we also do not weigh spells when we compute these two moments. That is, we set the duration of each match to 1, $t^{w}_{i,m} = t^{f}_{j,n} = 1$ for each $i = 1, \dots I$, $j = 1, \dots, J$, $m = 1, 2$, and $n = 1, \dots, N_j$. We compute the AGSU and KSS bias corrected correlations using the code created by Kline, Saggio, and Sølvsten (2019) which is available on GitHub, version 2.15.[33] We choose the Johnson-Lindenstrauss approximation method and set the precision parameter to $\varepsilon = 0.005$. We also use their code to find the largest leave-one-out connected set. We estimate BLM using the code provided by the authors on GitHub, version from March 15, 2019.[34]

Table 5 in the body of the paper shows the results. Imposing additional restrictions on the sample reduces the number of workers by less than 7% and firms by less 43%, with the exception of column (1). This sample size reduction does not have an economically

[33] `https://github.com/rsaggio87/LeaveOutTwoWay`
[34] `https://github.com/tlamadon/rblm`

significant impact on our estimates $\hat{\rho}$. The AKM correlation $\hat{\rho}_{AKM}$ decreases compared to Table 4 and is now negative in each sample. The AGSU and KSS corrections are similar. The bias-corrected correlations are both bigger than the OLS estimates of $\hat{\rho}_{AKM}$, but remain close to zero. The BLM estimates of the AKM correlation yields a higher estimate of the AKM correlation, averaging 0.296 for men and 0.289 for women for the mixture model. Still, these are much smaller than the estimates of our correlation.

Finally, we reconstruct the time series figure using these methods. For each year, we start from the sample we used in Section 6.5 but keep only the first two wage observations for each worker. We then find the largest leave-one-out connected set where each worker and firm has at least two observations. We then compute unweighed $\hat{\rho}$ and the OLS, AGSU, KSS, and BLM estimates of $\hat{\rho}_{AKM}$. Figure 7 depicts the results.

The level comparison is similar to that seen in Table 5: the OLS estimate of the AKM correlation is the lowest, followed by AGSU and KSS. The BLM estimate is closer to $\hat{\rho}$, which is the highest. For both men and women, the time trend in the AGSU and KSS estimates of $\hat{\rho}_{AKM}$ are similar to the time-trend in $\hat{\rho}$: increasing for men and decreasing for women. The BLM estimates of the AKM correlation shows little time trend, especially for women, although its level is close to $\hat{\rho}$.

References

Andrews, Martyn J., Leonard Gill, Thorsten Schank, and Richard Upward (2008). "High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias?" *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171.3, pp. 673–697.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa (2019). "A Distributional Framework for Matched Employer Employee Data". *Econometrica* 87.3, pp. 699–738.

Guimaraes, Paulo and Pedro Portugal (2010). "A Simple Feasible Procedure to Fit Models with High-Dimensional Fixed Effects". *The Stata Journal* 10.4, pp. 628–649.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten (2019). *Leave-out Estimation of Variance Components*. NBER Working Paper No. 26244.

Malmberg, Hannes and Ola Hössler (2014). "Probabilistic Choice with an Infinite Set of Options: An Approach Based on Random Sup Measures". *Modern Problems in Insurance Mathematics*. Ed. by Dmitrii Silvestrov and Anders Martin-Löf. Springer. Chap. 18, pp. 291–312.

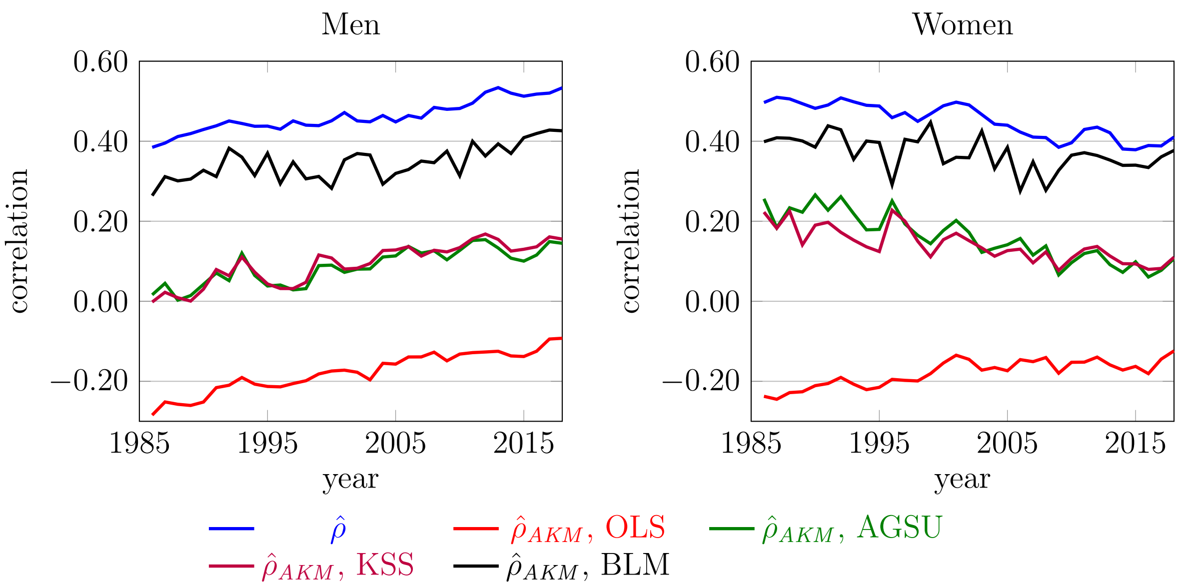

Figure 7: Year-by-year estimates of the correlation ρ and the AKM correlation using different methods, using AMDB 1986–2018 and the independence assumption III. The initial data construction is identical to that in Figure 3 but we further restrict the sample in each year to the largest leave-one-out connected set where every worker has exactly two distinct employers, and every firm employs at least two workers. The estimates are not weighted by duration.