High Wage Workers Work for High Wage Firms*

Katarína Borovičková
New York University

Robert Shimer
University of Chicago

November 26, 2019

Abstract

We develop a new approach to measuring the correlation between the types of matched workers and firms. Our approach is accurate in data sets with many workers and firms, but a small number of independent observations for each. Using administrative data from Austria, we find that the correlation lies between 0.4 and 0.6. We use artificial data sets with correlated worker and firm types to show that our estimator is accurate. In contrast, the Abowd, Kramarz and Margolis (1999) fixed effects estimator suggests no correlation between types in our data set. We show both theoretically and empirically that this reflects an incidental parameter problem.

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

*We are grateful for comments from John Abowd, Fernando Alvarez, Stephane Bonhomme, Jaroslav Borovička, Thibaut Lamadon, Rasmus Lentz, Ilse Lindenlaub, Elena Manresa, Derek Neal and Martin Rotemberg, 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.

fixed characteristics of workers and their employers is close to zero and sometimes 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 argues that this conclusion is unmerited. The finding that there is no sorting is a consequence of a well-known statistical problem with the fixed effects estimator proposed by AKM, a version of the incidental parameter problem which is often dubbed "limited mobility bias" (Abowd, Kramarz, Lengermann and Pérez-Duarte, 2004; Andrews, Gill, Schank and Upward, 2008). We propose a novel, simple, and accurate measure of the extent of sorting in the labor market and apply it to Austrian data. We find that the correlation between the unobserved types of workers and their employers is at least 0.4, probably above 0.5, and possibly as high as 0.6. In contrast, the AKM fixed effects estimator delivers a biased estimate of the correlation that is close to zero in our data set.

Measuring the correlation between types requires a cardinal measure of 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 very long period of time, her type would be the average log wage she receives. Similarly, a firm's type is defined 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 very long time series. This definition of type differs from the AKM fixed effects, but under natural distributional assumptions that we spell out in the body of the paper, the correlation between our notion of types is the same as the correlation between the AKM fixed effects, assuming both are measured without error.[2] That is, the difference between our results and those based on the AKM approach is not conceptual, but rather due to measurement issues.

The important difference between the two approaches is that real world data sets have few conditionally independent wage observations for most workers and firms. Our approach, in contrast to AKM, is well-suited to this type of environment. Wages are highly autocorrelated within worker-firm matches, so we think of the relevant unit of observation as being at the match level. In our data set we observe 4.1 million Austrian men working at 0.7 million

[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 (forthcoming) for Brazil, among others.

[2] Our definition of type is closer 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).

firms between 1972 and 2007. The median worker has two employers and the median firm has three employees over the entire time it is in the sample, although a few firms employ many more workers. It follows that the empirical average log wage is a noisy measure of a worker's or firm's type even with 36 years of data.

We therefore seek a measure of the correlation between types when we have a large number of workers and firms but the number of conditionally independent observations for each worker and firm is small. Our approach is to measure the correlation without measuring the type of any particular worker or firm, an important distinction from the AKM fixed effects approach. We assume that there is some underlying joint distribution of the types of matched workers and firms with finite first and second moments and we use a variance decomposition to recover those moments. This is similar to random effects, except we do not need to make any functional form assumptions on the joint distribution of matched types, beyond the finite second moment restriction.

Our approach allows the number of conditionally independent observations to be small but not too small. Our key identifying assumption is that for each worker, we have two or more observations of the actual wage received which are independently and identically distributed conditional on the worker's type; and for each firm, we have two or more observations of the actual wage paid which are independently and identically distributed conditional on the firm's type. Our measured correlation then pertains to the sample of workers and firms for whom this is true.

We first measure the correlation between types using annual wage data and find it is about 0.6 for both men and women. However, we recognize that annual wage observations might not be independent conditional on type, particularly for workers who do not switch employers. 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 wage in any two jobs which 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 0.49 for men and 0.43 for women.

A realistic model might also recognize that types change over time for reasons that we cannot observe. Because our approach is amenable to estimation using short time series, we can estimate the correlation between worker and firm types using only a single year's data, which should reduce the importance of time-varying types. Consistent with the hypothesis of time-varying types, our year-by-year estimates of the correlation are somewhat larger than our pooled estimates, averaging 0.53 for men and 0.47 for women.

We also estimate our model for each age and use a synthetic cohort approach to see how sorting evolves over the life cycle. We find a substantially rising correlation between worker and firm types for men, from 0.4 for men younger than 25 to above 0.6 for men in their thirties, finally approaching 0.8 for men older than 45. This is consistent with the view that learning about types takes time, but once types are known, the labor market sorts the high wage workers into high wage firms. The pattern for women is more complicated, possibly reflecting the exit and reentry of women from the labor force during years of peak fertility.

Finally, we allow workers' and firms' types to vary depending on the partners' observable characteristic. For example, we let firms have different types when matched with workers with different education levels. This raises the estimated correlation to 0.60 for men and 0.53 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 results differ from the existing literature based on AKM because our method for measuring the correlation differs. The key difference is that the AKM approach requires estimating a fixed effect for each worker and firm, a huge number of parameters. These estimates are consistent only in the limit when the number of workers, the number of firms, and the number of independent observations for each worker and firm all go to infinity. With a finite number of observations per worker and firm, the estimated fixed effects are noisy measures of the true types. Moreover, this noise is negatively correlated across matched workers and firms, biasing down or even negative the estimated correlation between matched worker and firm fixed effects. In contrast, our approach only requires two independent observations for each worker and firm.

We perform three exercises to show that this incidental parameter problem drives the estimated correlation between fixed effects. First, we show that the estimated correlation using our approach and using the fixed effects approach differs dramatically even when estimated on the same data set. Second, using Monte Carlo on artificial data sets that match the statistical properties of real-world data, we verify that our approach accurately measures the correlation between types while the fixed effects approach is biased. Third, we construct a simple matching model where we can measure the bias in the fixed effects estimator analytically. The model explains about half of the difference between our estimates and the fixed effects estimates given (i) our estimates of the first and second moments of the joint distribution of worker and firm types and (ii) the mean number of jobs held by each worker and the mean number of workers who work at each firm. Much of the remaining difference between the two estimators seems to reflect the fact that our model understates clustering in the matching graph, i.e. the fact that a worker's coworkers in one job are much more likely than other similar workers to be coworkers at another job. This leads our model to overstate the number of independent observations for each worker and firm and hence understate the bias in the AKM approach. We conjecture that violations of AKM's "exogenous mobility" assumption, that errors in the wage equation are orthogonal to worker and firm identities, may be important for explaining the remaining difference between the estimators.

Our main contribution lies in developing a simple and accurate measure of the correlation between worker and firm types. As previously noted, we are not the first to observe the bias of the AKM fixed effects estimator. Andrews, Gill, Schank and Upward (2008) propose estimating the AKM correlation and then applying a bias correction. Andrews, Gill, Schank and Upward (2012) instead suggest estimating the AKM correlation using a subsample of workers, which worsens the bias, and then extrapolating to estimate the true correlation. Jochmans and Weidner (2017) propose bounds on the variance of the fixed effects estimator and use those to analyze the bias in the AKM correlation. Our approach avoids the need for bias corrections, extrapolation, or bounds.

? offer a complementary approach to examining sorting patterns in the data. They propose a two-step estimator where firms are first classified into bins before estimating fixed effects. One advantage of our approach is its simplicity and transparency. We only need to estimate variances and covariances, while they need to first group firms into bins. A side effect of this is that our estimates appear to be more accurate. Using Monte Carlo, we show that we are able to recover the correlation and obtain tight confidence intervals using our approach in artificial data sets. In contrast, the estimator proposed by **?** appears to be biased and their confidence intervals are wider; see their Table 3. On the other hand, **?** are able to answer questions that we cannot address, in particular how a worker's wage depends on her employer's type.

A third approach is to think of the AKM correlation as a moment to match in a structural model. Two recent examples are Hagedorn, Law and Manovskii (2017) and Lopes de Melo (forthcoming).[3] Our assumption that the wages in jobs separated by an unemployment spell

[3]Lopes de Melo (forthcoming) shows that the correlation between a worker's AKM fixed effect and the

are independent conditional on a worker's type is satisfied in the models in both of those papers, and so our approach imposes fewer theoretical restrictions. The drawback to these structural approaches is that all the results, including the correlation between types, may be sensitive to the additional assumptions in the model. The payoff from the structural approach is that these papers can discuss issues that are beyond the scope of this paper. For example, Hagedorn, Law and Manovskii (2017) estimate the output of any worker in any firm, while we have nothing to say about the production function, only about measured sorting between high wage workers and high wage firms.

The remainder of the paper proceeds as follows. Section 2 defines our measure of the correlation between worker and firm types. In Section 3, we use several models as our laboratories to study how our measures the extent of sorting and compare it to the AKM measure of correlation. We propose an estimator in Section 4 and implement it on Austrian dataset, described 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 8 concludes.

2 Measuring Sorting in Theory

2.1 The Economy

We consider a cross-section of an economy with a fixed measure of employed workers and 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. We treat F and Φ_x as primitives in our environment and view them as coming from a snapshot of a structural dynamic model such as Burdett and Mortensen (1998), Shimer and Smith (2000), or Postel-Vinay and Robin (2002). That is, F is the cross-sectional distribution of employed workers' characteristics and Φ_x is the cross-sectional conditional distribution of their employers' characteristics. 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

AKM fixed effect of her coworkers is a useful moment in estimating his structural model. This moment is related to one we use, the correlation between a worker's log wage in her other jobs and the log wage of her coworkers in this job.

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. Using Bayes rule, we have $\Phi_x(y)F(x) \equiv \Psi_y(x)G(y)$ for all x and y.

We assume that a worker with characteristics x matched to a firm with characteristics y earns a wage that possibly depends on both vectors of 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 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. Of course, this is not true in real world data sets, and so Section 4 explains how we can estimate the correlation between expected log wages using the limited wage data that is available. Here we simply define expected log wages and the correlation between worker and firm types. Let

$$\begin{aligned} \lambda(x) &\equiv \int_Y \int_0^1 w(x, y, z)\, dz\, d\Phi_x(y) \\ \text{and } \mu(y) &\equiv \int_X \int_0^1 w(x, y, z)\, dz\, d\Psi_y(x) \end{aligned}$$

denote the expected log wage received by a worker with characteristics y and the expected

[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.

[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.

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.

We want to measure 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,

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

where

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

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

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

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

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

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

We highlight the special case where $\Phi_x(y) = \Phi(y)$ for all x and y. For example, each worker may be equally likely to work in every job, in which case $G(y) = \Phi(y)$. 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. This example emphasizes that there is nothing in our definition of types which pushes us towards a positive correlation. Later we offer examples where the correlation can be negative. 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) (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,

$$w(x_i, y_j, z) = \alpha_i + \psi_j + \eta \tag{1}$$

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.[6] 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 AKM model 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 the model is misspecified, we define the fixed effects in a structural model as the solution to the following moment conditions,

$$\alpha(x_i) = \int_Y \int_0^1 \left(w(x_i, y, z) - \psi(y)\right) \, dz \, d\Phi_{x_i}(y)$$
$$\psi(y_j) = \int_X \int_0^1 \left(w(x, y_j, z) - \alpha(x)\right) \, dz \, d\Psi_{y_j}(x),$$

which is equivalent to running OLS on data containing all matched firms. Again, α and ψ are uniquely defined up to an additive constant if (and only if) there is no way to partition the workers and firms into two sets A and B such that workers and firms in set A (B) only match with firms and workers in set A (B).

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

We then compute the correlation ρ_{AKM} in the matched pairs as

$$\bar{\alpha} \equiv \int_X \alpha(x)\, dF(x), \quad \bar{\psi} \equiv \int_Y \psi(y)\, dG(y),$$
$$\sigma_\alpha \equiv \sqrt{\int_X (\alpha(x) - \bar{\alpha})^2\, dF(x)}, \quad \sigma_\psi \equiv \sqrt{\int_Y (\psi(y) - \bar{\psi})^2\, dG(y)}$$
$$\rho_{AKM} \equiv \frac{\int_X \int_Y (\alpha(x) - \bar{\alpha})(\psi(y) - \bar{\psi})\, d\Phi_x(y)\, dF(x)}{\sigma_\alpha \sigma_\psi}.$$

We do not focus here on how to estimate ρ_{AKM}; there are well-known statistical problem 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 recover the idealized moments that one would get with infinitely much data and infinitely many switchers.

3 Models as Laboratories for Measuring Correlation

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

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 (1) is correctly specified and the joint density of matched worker and firm pairs, $\xi(\alpha, \psi)$, is elliptical[7] with the variance-covariance matrix

$$\begin{pmatrix} \sigma_\alpha^2 & \rho_{AKM}\sigma_\alpha\sigma_\psi \\ \rho_{AKM}\sigma_\alpha\sigma_\psi & \sigma_\psi^2 \end{pmatrix}.$$

[7] The joint distribution is *elliptical* if the associated density function ξ can be expressed as

$$\xi(\alpha, \psi) = \tilde{\xi}\left(\frac{(\alpha - \bar{\alpha})^2}{\sigma_\alpha^2} - \frac{2\rho_{AKM}(\alpha - \bar{\alpha})(\psi - \bar{\psi})}{\sigma_\alpha\sigma_\psi} + \frac{(\psi - \bar{\psi})^2}{\sigma_\psi^2}\right)$$

for some function $\tilde{\xi}$, i.e. if the level curves of the density functions are ellipses. The bivariate normal and the bivariate t-distributions satisfy this property.

We prove in the appendix that in this case, the conditional expected value of ψ_j in a match is linear in α_i, $\int_Y \psi(y)dG_{x_i}(y) = \kappa_0 + \kappa_1\alpha(x_i)$ for all i. The definition of λ and the wage equation (1) then imply

$$\lambda_i = \int_Y (\alpha(x_i) + \psi(y))d\Phi_{x_i}(y) = \kappa_0 + (1+\kappa_1)\alpha_i.$$

Symmetrically, the conditional expected value of α_i in a match is linear in ψ_j, $\int_X \alpha(x)d\Psi_{y_j}(x) = \theta_0 + \theta_1\psi(y_j)$ for all j, and

$$\mu_j = \int_X (\alpha(x) + \psi(y_j))d\Psi_{y_j}(x) = \theta_0 + (1+\theta_1)\psi_j.$$

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 $\kappa_1 < -1$ or $\theta_1 < -1$. However, κ_1 and θ_1 can be expressed in terms of variance-covariance matrix of α, ψ and hence we can detect the sign flip.

The following Proposition summarizes this result.

Proposition 1 *Assume that the joint distribution of α and ψ is elliptical and $\rho_{\mathrm{AKM}} \in (-1,1)$. Then λ and μ are linear transformations of α and ψ with correlation ρ and standard deviations $\sigma_\lambda = |\sigma_\alpha + \rho_{\mathrm{AKM}}\sigma_\psi|$ and $\sigma_\mu = |\sigma_\psi + \rho_{\mathrm{AKM}}\sigma_\alpha|$. Moreover,*

$$(\sigma_\alpha + \rho_{\mathrm{AKM}}\sigma_\psi)(\sigma_\psi + \rho_{\mathrm{AKM}}\sigma_\alpha) \gtreqless 0 \Rightarrow \begin{cases} \rho = \rho_{\mathrm{AKM}} \text{ and } (\sigma_\lambda - \rho\sigma_\mu)(\sigma_\mu - \rho\sigma_\lambda) > 0 \\ \rho \text{ is undefined} \\ \rho = -\rho_{\mathrm{AKM}} \text{ and } (\sigma_\lambda - \rho\sigma_\mu)(\sigma_\mu - \rho\sigma_\lambda) < 0. \end{cases}$$

The proof in Appendix A.1 establishes linearity of conditional expected values for elliptical distributions and finds conditions, both in terms of variance-covariance matrix of (α, ψ) and (λ, μ), under which both transformations are increasing.

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

3.2 Shimer and Smith (2000) with Match Productivity

We next examine search model with two-sided heterogeneity and match-specific heterogeneity, as in 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 types, but instead depends on the idiosyncratic shock as well.

The model is formulated in continuous time. There is measure 1 of workers and measure 1 of firms. Each worker is characterized by his productivity x, distributed in the population according $F(x)$. Similarly, each firm is described by its productivity type y, distributed according to $G(y)$.

Search is random and only unmatched firms and workers can search. Let $u(x)$ be the unemployment rate among workers of type x, and $v(y)$ vacancy rate among firms with type y. An unemployed worker meets a vacancy at the rate θ and the firm type is randomly drawn from the distribution G. If the firm has a filled job, it is as if the meeting never happened.[8] If it has a vacancy, with probability $v(y)$, the pair draws the match specific productivity $z \geq 0$ from distribution ζ and decides 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 γ. Assume $H(x,y)$ is strictly positive for almost all x and y. Matches randomly separate at the rate δ. Agents discount future at the rate r.

Let $U(x)$ and $V(y)$ be the value of being an unemployed worker and a vacant firm, 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 $z \geq \underline{z}(x,y)$ where $\underline{z}(x,y)$ is such that $S(x,y,\underline{z}(x,y)) = 0$. The system of equations which fully describe the model are in Appendix A.2, here we focus on analysis of wages. The wage setting 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

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

If the distribution of match productivity is exponential, we can prove that the expected log

[8] This is the quadratic technology is Shimer-Smith. A more standard assumption that unemployed workers only meet vacant firms is equivalent (for the purposes of this paper) to a rescaling of θ.

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.[9] This is in contrast to Shimer and Smith (2000), where a given x's wage is maximized at some value of y, typically an interior point, even if H is strictly increasing.

Another difference from Shimer and Smith (2000) is that all matches can be created as long as the match specific productivity is high enough. For example, with enough complementarity in the production function, a worker with the lowest type would never match with the highest type 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 low type worker employed by low type firms most of the time.

We solve the model for discrete number of types n, distributed uniformly on $X = Y = \{\frac{1-0.5}{n}, \frac{2-0.5}{n}, \dots \frac{n-0.5}{n}\}$ and so $dF(x) = dG(y) = \frac{1}{n}$. We use the CES production function

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

where $c \geq 0$ is the elasticity of substitution and $a \in [0,1]$ is worker's share in production. We assume that the distribution of match productivity shocks is Pareto, with some minimum value and variance σ_z^2.[10] Our benchmark uses the following parameter values: $r = 1, \delta = 10, \theta = 10^4, \gamma = 0.5, a = 0.5, c = 1, \sigma_z^2 = 0.1, min(z) = 1, n = 500$.

We are interested in exploring how two measures of sorting ρ and ρ_{AKM} vary as we change the meeting rate, the bargaining power, variance of the match specific shocks, and the elasticity of substitution in the production function. In each experiment, we compare ρ and ρ_{AKM} with the correlation between x and y, an intuitive measure of the extent of sorting that is generally not feasible in real-world data but that can easily be computed in the model.

Figure 1 shows results from these experiments. In the top left panel, we vary the meeting rate θ. When θ low, it is hard for workers to meet vacancies and thus they tend to accept any offer they receive, conditional on a favorable match specific shock. As match acceptance thresholds are similar for different types x and y, we see little sorting. As $\theta \to \infty$, workers receive offers very quickly, become selective at which offer to accept and sorting increases. The correlation between x and y (red) is increasing in θ, and our measure of correlation (blue) as well as the AKM correlation (green) capture this pattern properly.

We change the value of the bargaining power in our second experiment. As γ converges

[9] Numerically we find this to be true for Pareto distribution as well.

[10] 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.

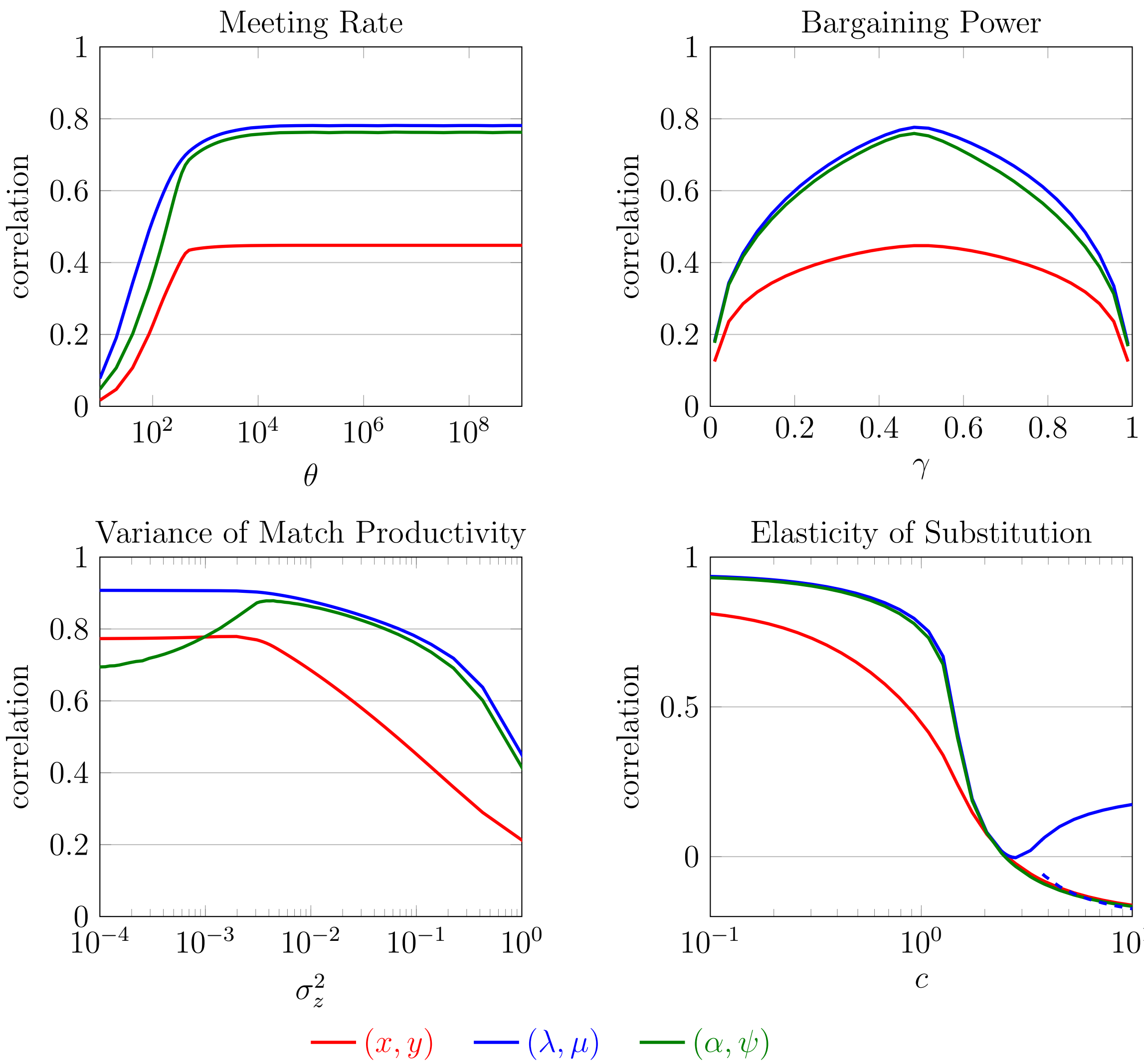

Figure 1: Comparative statics exercise in the Shimer and Smith (2000) model extended to include match specific productivity draws. The figures show correlation between three different measures of types: (x, y) (red lines), (λ, μ) (blue lines), (α, ψ) (green lines) in the matched pairs for different parameter values. In each experiment, we keep parameters at their benchmark values, $r = 1, \delta = 10, \theta = 10^4, \gamma = 0.5, a = 0.5, c = 1, \sigma_z^2 = 0.1, \min(z) = 1, n = 500$, and only vary one parameter at a time.

to 0 or 1, workers (or firms) are paid only their outside option regardless of who they match with, and sorting becomes weaker. The correlation between x, y is hump-shaped, see the top right in Figure 1, and again our correlation as well as the AKM correlation properly capture the extend of sorting.

In our third comparative static exercise, we vary 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 left bottom panel of Figure 1 shows that starting from a large enough variance, all three measures of sorting decrease as the variance increases further. However, for small values of the variance, the AKM correlation suggests that sorting is weaker for small leves of the variance even though the extent of sorting is barely affected. With small variance of the shocks, the model gets closer to the original Shimer and Smith (2000) model where the wage function is a non-linear function of the worker and firm type. As a result, the AKM wage equation is misspecified and the AKM correlation becomes a poor measure of sorting.

Finally, we turn to the elasticity of substitution in the production function. The production function is Leontief for $c = 0$ and linear for $c \to \infty$. The bottom right panel of Figure 1 shows correlations for $c \in [0.1, 10]$. Our measure properly captures the strength of sorting when the production function features enough complementarity between inputs. As we move towards the perfect substitution case, our measure recovers a positive correlation despite the fact that the correlation between x and y is negative. In this economy, high x workers tend to work for low y firms, but low y firms actually pay high wages. Hence, high wage workers work for high wage firms, a pattern which is picked up by our measure. From the 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. This point had been made before by Eeckhout and Kircher (2011). Using Proposition 1, we propose a test to detect the "sign flip". Proposition 1 states that when $(\sigma_\lambda - \rho\sigma_\mu)(\sigma_\mu - \rho\sigma_\lambda) < 0$, then ρ and ρ_{AKM} have the opposite sign. The blue dotted line shows the correlation adjusted for the flip sign, meaning that we plot $-\rho$ when this condition is satisfied. All three correlations depicted in the top right panel of Figure 1 have the similar pattern.

These experiments illustrate that our proposed measure of sorting reflects the extent of sorting in the model economy. The AKM correlation also reflects changes in sorting well. The reason is that expected log wages are close to log-linear in the worker and firm types, and hence AKM wage restrictions are close to being satisfied.

3.3 Discrete Choice Model

We next examine a static discrete choice model. There are 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 $G(y)$. There is no search in this model. Instead, each worker chooses a firm he wants to work for so as to maximize his utility. Workers' utility is the sum of log wage w and amenity value ε. The log wage depends on worker's and firm's characteristics. Worker i sees the amenity value he would get at each firm and chooses to work for firm j^* such that

$$j^* = \arg\max_j \left(w(x_i, y_j) + \varepsilon_{i,j} \right).$$

We assume that the wage function is bounded above and that amenities are drawn from an exponential distribution with mean (and hence standard deviation) s. This ensures that workers' choice of y_j has a non-trivial limit when the number of firms goes to infinity (Malmberg, 2013). In the limit, the probability that a worker with characteristics x chooses a firm with characterisitics y is

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

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 log wage is given by

$$w(x,y) = c_x x + c_y y - (\sqrt{c_{xx}}x - \sqrt{c_{yy}}y)^2$$

with c_x, c_y, c_{xx}, c_{yy} positive. Then, log wage of worker x is maximized at firm y^* such that

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

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

When the types x and y are distributed normally, then the joint distribution of match is normal and we obtain closed form expressions the types $\lambda(x), \mu(y)$, AKM types $\alpha(x), \psi(y)$ as well as their correlation ρ and ρ_{AKM}. We show some formulas in Appendix A.3.

For our benchmark, we choose the following parameter values: $x \sim N(m_x, \sigma_x)$ and

$G \sim N(m_y, \sigma_y)$ with $m_x = m_y = 1, \sigma_x = \sigma_y = 1$, $s = 1$ and $c_x = 1, c_y = 0, c_{xx} = c_{yy} = 1/a^2$ so effectively the wage function is $w(x, y) = x - (x - y)^2/a$. We again conduct several experiments by varying parameters of the model and measuring how sorting changes. We do comparative static exercise with respect to s, a, $m_x - m_y$ and σ_x. It turns out that varying parameters of the wage function is isomorphic to varying parameters of x and y distributions, and hence we do not focus on it.[11]

The top left panel of Figure 2 shows the comparative static exercise with respect to standard deviation of amenity. When it is zero, amenity does not play any role in worker's decision and each worker x chooses the firm $y^*(x)$. As a result, firm $y^*(x)$ employs only workers of type x, and hence the correlation between x and y is one. As the standard deviation increases, the role of wage, hence types x and y, decreases. As a result, sorting weakens and the correlation between (x, y) declines to zero. We observe that the correlation between $\lambda(x)$ and $\psi(y)$ exhibits the same pattern as the correlation between x and y – it is one in the case of perfect sorting, and then monotonically declines toward zero as sorting weakens. The correlation between $\alpha(x)$ and $\psi(y)$ follows a very different pattern. At $s = 0$, the AKM correlation is zero even though there is perfect sorting in the economy. The reason is that AKM attributes all wage variation to the worker fixed effects and no variation to the firm fixed effects, generating zero covariance between the types.[12] As the standard deviation increases, the AKM correlation becomes negative but remains close to zero. The correlation is not defined at $s = 2$ because $\sigma_\psi = 0$. In this experiment, the AKM correlation does not reflect changes in sorting in the underlying economy.

The results are qualitatively very similar in the experiment where we change 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 $y^*(x)$. The correlation between x and y, as well as between λ and μ, is one. As a increases, worker's wage depends less and less on firm type, sorting weakens and eventually converges to to no sorting as $a \to \infty$. The correlation between x and y declines from 1 to 0 in this experiment, and so does the correlation between λ and μ. We thus conclude that our measure of sorting properly captures changes in sorting. As in the previous experiment, the AKM correlation fails to capture the extent of sorting, especially in the region with perfect sorting.

In the next experiment, we vary the difference in means $m_x - m_y$. The extent of sorting as measured by the correlation between x and y does not change. Loosely speaking, increasing

[11]This is achieved through the following transformation: $\tilde{x}(x) = y^*(x)$, $\tilde{y}(y) = -\frac{1}{4}(2c_x c_y/\sqrt{c_{xx}c_{yy}} + c_y^2/c_{yy}) + (c_y + c_x c_{yy}/\sqrt{c_{xx}c_{yy}})y$, and $w(x, y) = \tilde{w}(\tilde{x}, \tilde{y}) = \tilde{x} - (\tilde{x} - \tilde{y})^2/\tilde{a}$, with $\tilde{a} = (c_x c_{yy} + c_y\sqrt{c_{xx}c_{yy}})^2/(c_{xx}c_{yy}^2)$.

[12]To be precise, in this case the covariance between α, ψ is zero and the standard deviation of ψ is zero so the correlation is undefined. However, the limit of the correlation as $s \to 0$ is well defined.

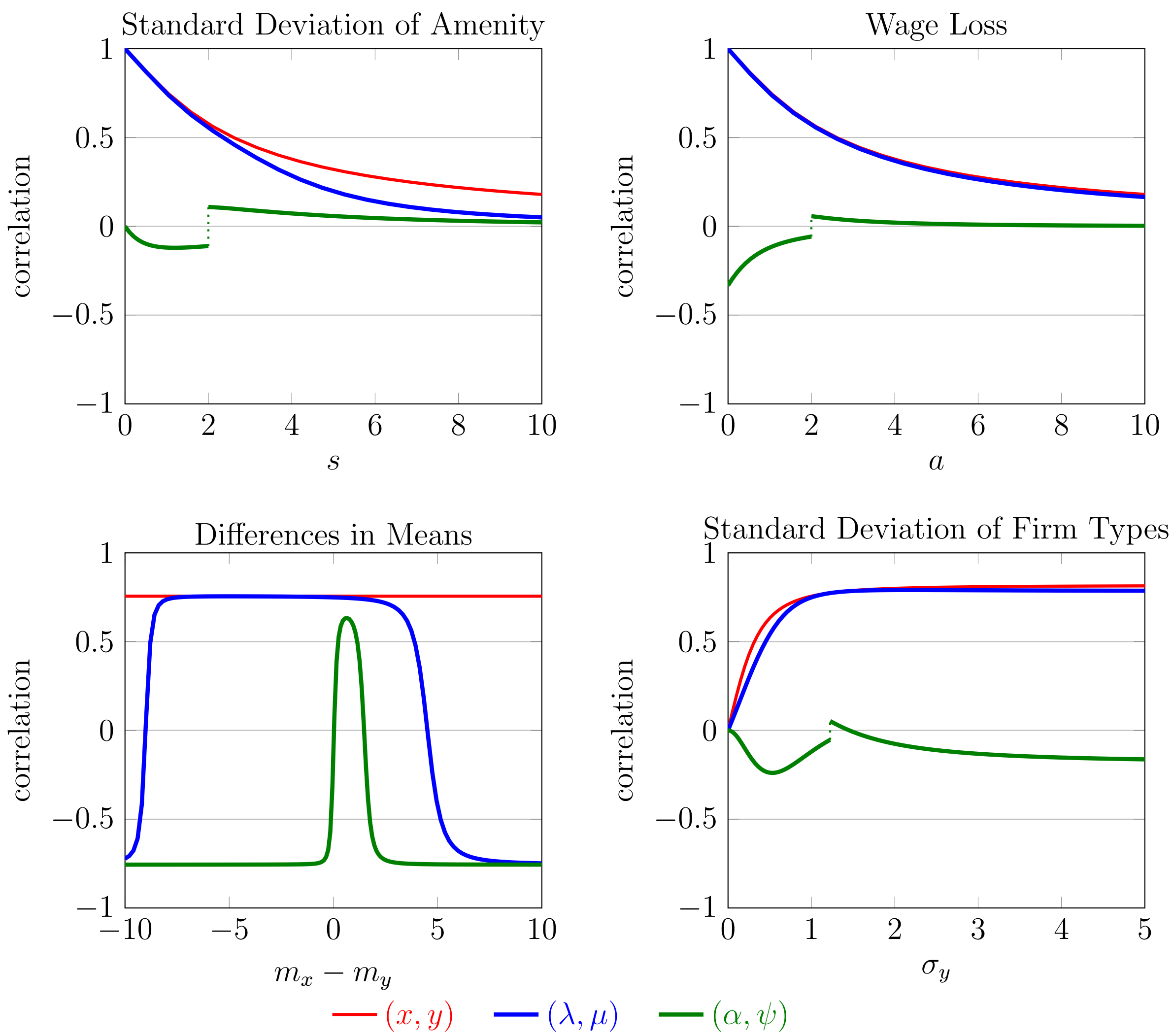

Figure 2: Comparative statics exercise in discrete choice model. The figures show correlation between three different measures of types: (x, y) (red lines), (λ, μ) (blue lines), (α, ψ) (green lines) in the matched pairs for different parameter values. In each experiment, we keep parameters at their benchmark values, $a = 1, s = 1, m_x = 0, m_y = 0, \sigma_x = 1, \sigma_y = 1$, and only vary one parameter at a time.

the mean of y relative to x only makes workers choose higher y but this increase is the same across x. Therefore, the correlation is not affected. The correlation between λ and μ (and also between α, ψ) depends on differences in means. As the mean of y increases relative to the mean of x, it becomes difficult to find the right firm since there are very few of them. This misalignment between which firms workers want to work at and which firms exist, is larger for low-type workers. As a result, high x workers work on average in high y firms, but high y firms pay lower average wages because of the large wage penalty incurred by the low x types employed by these firms. This is the reason for why our measure recovers negative correlation – high wage workers tend to work in low wage firms. Again, from the wage data alone, without the knowledge of the structure of the economy, it is not possible to make the inference that firms which pay low average wages are actually high type firms. This exercise also illustrates that the correlation between λ, μ can be negative.

In the last experiment, we increase the variance of worker types while keeping the variance of firm types unchanged. In the extreme case of $\sigma_y = 0$, all firms have the same type, hence there is no sorting. As the variance increases, more and more workers will have "the best firm $y^*(x)$" in their choice set and sorting increases. However, increasing the variance too much beyond the variance of worker types will not bring any additional improvement of sorting since the extreme type firms are not chosen by any worker. We indeed see that the correlation between x, y starts at zero when $\sigma_y = 0$, and then increases steeply until around $\sigma_y = 1$, after which it flattens. The correlation between λ, μ follows the same pattern, while the AKM correlation again fails to capture changes in sorting.

To summarize, in the discrete choice model, our measure of sorting properly captures sorting patterns but the AKM correlation does not. Maybe the most striking finding is that the AKM correlation is zero in two situations with perfect sorting, $s \to 0$ and $a \to \infty$. In this model, the true wage equation is non-monotone in types and hence the AKM wage equation is misspecified. Even thought if can potentially be a useful first order approximation, our calculations reveal that in this case the AKM correlation fails as a measure of sorting.

4 An Estimator of the Measure of Sorting

We return now to the cross-sectional correlation between λ and μ. A structural model determines sorting patterns, together with the wage and duration of each match. If we observed many conditionally independent matches, described by wage and duration, for each worker and firm, we could accurately measure λ and μ for everyone and hence directly measure their correlation. Unfortunately, in practice we have very few observations for most workers and most firms. This section proposes a strategy for measuring the correlation between λ and μ in realistic data sets. We start by defining a statistical model which encompasses and extends the structural models in Section 3. We then propose an estimator and show 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 imagine a finite set of possible characteristics 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 characteristics y and $J \equiv \sum_{y\in Y} J_y$ the total number of firms. We later consider replicating 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 characteristics 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 $M_i \in \{2, \ldots, \bar{M}\}$ matches indexed by $m = 1, \ldots, M_i$. 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 characteristics for that match. We assume that the worker's characteristics 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 characteristics 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 characteristics by $\bar{T}^w_x$. 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 $N_j \in \{2, \ldots, \bar{N}\}$ matches indexed by $n = 1, \ldots, N_j$. 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 $\{w^f_{j,n}, t^f_{j,n}, x_{j,n}\}_{n=1}^{N_j}$. When a firm matches with a worker with characteristic x, it is equally likely to match with any such worker and we let $\mathbf{i}_{j,n}$ denote the identity of the worker. 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 characteristics as $\bar{T}^f_y$. 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 $\mathbf{i}_{i,m}$ 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, the average log wage that worker i with characteristic x_i earns during his lifetime and the average log wage that firm j with characteristics y_j pays are

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

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 and variance of λ and μ:

$$\bar{\lambda} \equiv \frac{\sum_{x \in X} I_x \bar{T}^w_x \lambda(x)}{\sum_{x \in X} I_x \bar{T}^w_x} \text{ and } \bar{\mu} \equiv \frac{\sum_{y \in Y} J_y \bar{T}^f_y \mu(y)}{\sum_{y \in Y} J_y \bar{T}^f_y},$$

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

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 that $\bar{\lambda} = \bar{\mu}$, although the variances may be different.

Finally, we can compute the covariance between λ and μ in matched pairs:

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

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

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

This shows that the weight attached to each characteristic y firm is $\bar{T}^f_y$.

4.2 Auxiliary Assumptions and an Estimator

We now introduce some auxiliary assumptions and then define estimators of σ^2_λ, σ^2_μ, and c. Our main result in this section is that the estimators are consistent under the auxiliary assumptions.

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 first auxiliary assumption consists of two pieces. First, a worker's wages in different matches are independently identically distributed with a characteristic-specific mean and variance. Second, wage draws and durations are uncorrelated conditional on the worker's characteristic. The second auxiliary assumption imposes the same restrictions on firms. We recognize that these assumptions are restrictive, and so in Section **??** we develop approaches to handling real-world data that are designed to satisfy these assumptions.

The third auxiliary assumption 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 for the error in the firm's wage equation in one match and the employee type in other matches. We stress that this assumption allows the error terms to be correlated within a match, and indeed this will typically be the case.[13]

The first three auxiliary assumptions are useful for finding individual-level unbiased estimators of worker and firm types and the covariance between them. The fourth auxiliary assumption gives us a law of large numbers, ensuring that the average of these unbiased estimators is consistent as the economy grows large. This assumption rules out the possibility of correlated shocks. In the data, we handle aggregate shocks by deflating wages by the economy-wide average wage, but other correlation, 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 $\lambda(x) = \bar{w}_x^w$. A firm with characteristic y has type $\mu(y) = \bar{w}_y^f$.*

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_λ^2, σ_μ^2, and c. Start with the variance of worker types. Define

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

We show in the proof of Proposition 3 that these are unbiased estimators of $\lambda(x_i)$ and $\lambda(x_i)^2$ and use that to find an estimator of the variance

Proposition 3 *A consistent estimator of the variance of worker types is σ_λ^2 is*

$$\widehat{\sigma_\lambda^2} \equiv \frac{\sum_{i=1}^{\tau I} T_i^w \widehat{\lambda_i^2}}{\sum_{i=1}^{\tau I} T_i^w} - \left(\frac{\sum_{i=1}^{\tau I} T_i^w \hat{\lambda}_i}{\sum_{i=1}^{\tau I} T_i^w} \right)^2.$$

The consistency proof is a standard law of large numbers argument.

The logic for firms is identical. Define

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

[13] 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.

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

Proposition 4 *A consistent estimator of the variance of firm types σ_μ^2 is*

$$\widehat{\sigma_\mu^2} \equiv \frac{\sum_{j=1}^{\tau J} T_j^f \widehat{\mu_j^2}}{\sum_{j=1}^{\tau J} T_j^f} - \left(\frac{\sum_{j=1}^{\tau J} T_j^f \hat{\mu}_j}{\sum_{j=1}^{\tau J} T_j^f} \right)^2 .$$

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

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

Each of the $w_{i,m'}^w$ is an unbiased estimator of $\lambda(x_i)$ and each of the $w_{\mathbf{j}_{i,m},n'}^f$ 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 *A consistent estimator of the covariance c is*

$$\hat{c} \equiv \frac{\sum_{i=1}^{\tau I} \sum_{m=1}^{M_i} t_{i,m}^w \hat{c}_{i,m}}{\sum_{i=1}^{\tau I} T_i^w} - \left(\frac{\sum_{i=1}^{\tau I} T_i^w \hat{\lambda}_i}{\sum_{i=1}^{\tau I} T_i^w} \right)^2 .$$

Armed with consistent estimators of the covariance and two variances, it is straightforward to construct an estimator of the correlation as $\hat{c}/\sqrt{\widehat{\sigma_\lambda^2}\widehat{\sigma_\mu^2}}$. Assuming the variances of worker and firm types are both positive, this estimator is consistent.

4.3 Small Sample Properties of the Estimators

We next examine small sample properties of the estimator. We create artificial datasets from the structural models introduced in the previous section. We choose different values of I, J to see how the estimator performs in datasets of different sizes. Importantly, in each dataset we keep number of observations deliberately small, 3.8 on average. For each choice of I, we choose $J = I/5$, which guarantees that the number of observations per firm is also small, consistent with real world data.[14]

For each model, we create $B = 500$ artificial samples. We start with $I \in \{2500, 10^4, 10^5\}$ and $J = I/5$. The actual number of firms and workers in each sample can be smaller because

[14]We use distributions of observations per worker as is observed in the Austrian data, in the sample corresponding to Table 2, column (3).

	descriptive statistics					distribution of $\hat{\rho}_b - \rho_b$		
	I	J	M	N	ρ	5%	mean	95%
Shimer, Smith								
	2,412	500	3.9	18.7	0.754	-0.007	-0.002	0.006
	9,662	2,000	3.9	18.8	0.770	-0.003	0.000	0.003
	96,620	20,000	3.9	18.8	0.775	-0.001	0.000	0.001
Discrete choice model								
	2,496	500	3.9	19.7	0.743	-0.009	-0.001	0.006
	9,996	1,999	4.0	19.8	0.748	-0.004	0.000	0.004
	100,000	20,000	4.0	19.5	0.775	-0.001	0.000	0.001

Table 1: Monte Carlo simulations in Shimer and Smith (2000) 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. First five columns show several descriptive statistics computed as means across samples – number of workers I, number of firms J, number of job per worker M, number of workers per firm N and true sample correlation ρ. The last three columns show the mean, the 5^{th} and 95^{th} quartile of the error distribution, $\hat{\rho}_b - \rho_b$.

we drop firms and workers with less than two observations. For each sample b, we compute the true correlation between λ and μ using formulas in Section 2; we call this object ρ_b. We then use sample's wage and duration data to estimate $\hat{\rho}_b$ using formulas in 4.2. We report the mean value of ρ_b across samples and the distribution of error $\hat{\rho}_b - \rho_b$ for Shimer and Smith (2000) and the discrete choice models in Tables 1. Both these models satisfy our identifying assumptions.

We parametrize Shimer and Smith (2000) model such that the correlation between λ and μ in the infinite economy is 0.776. The realized correlation ρ_b varies across samples reflecting randomness in the matching process, and is on average lower than in the infinite economy due to finite number of agents, see the fifth column in Table 1. We observe that our estimator performs well even with $I = 1,000$ workers and $J = 500$ which is orders of magnitude smaller sample than a typical real world dataset. As the number of workers and firms increases, the error becomes smaller. In this model, the duration of the match is an exponentially distributed random variable which is uncorrelated with wage or workers' and firms' types. Therefore, each match is equally informative about the correlation and an efficient estimator would weigh all matches equally. We nevertheless weigh each worker by the total duration to use exactly the estimator we proposed.

Table 1 summarizes the results for the discrete choice model. With the chosen parameter values, the correlation in an infinite sample is 0.749. This is a repeated static model and as such it has no prediction for the duration of the match, and we therefore assume that each match lasts one period. As in the previous model, the correlation in each particular sample is different and typically lower than one corresponding to an infinite economy due to finite number of workers and firms. We again observe that the error in the estimator is very small, even in the sample with $I = 1,000$ workers and $J = 200$ firms. We point out that the average number of observations per worker is small, consistent with a typical real world dataset.

We conclude this estimator performs well in small samples. Even in samples orders of magnitude smaller than a typical real world data set, the error is on the third decimal place.

5 Data

5.1 Data Description

We measure the correlation between workers and jobs using two panel data sets from the Austrian social security registry (Zweimuller, Winter-Ebmer, Lalive, Kuhn, Wuellrich, Ruf and Buchi, 2009), ASSD and AMDB. The ASSD covers the universe of workers in the private sector from 1972 to 2007, the AMDB dataset from 1997 to 2017. Even though these two data sets cover the universe of Austrian labor force, we treat them as independent and report results separately for each data set. 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,[15] we observe earnings and days worked during the year.[16] We also have some limited demographic information on workers, including their birth year and sex. After 1986, we observe registered unemployment spells, which we use in much of our analysis. We also observe the education of most workers who experience a registered unemployment spell. Finally, we have some information about jobs, including region, industry, and whether the position is blue or white collar.

Following Card, Heining and Kline (2013), we focus on workers age 20–60. We look both 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 a week) and data that include an apprenticeship. We note that this dataset does 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.

[15] Formally, a firm is identified using its employer identification number (EIN). Some firms may have multiple EINs.

[16] Earnings are top-coded at the maximum social security contribution level, which rises over time. For example, in 2007, the cap is €3840 per month. The fraction of male worker-firm observations affected by top-coding fell from a peak of 25.3 percent in 1974 to 13.5 percent in 2007. Top-coding affects far fewer female worker-firm observations, varying from 3.6 to 6.5 percent during our sample period. We discuss the importance of top-coding for our results in Section 6.4.

Over the period 1994–2007, on average 4.7 percent of employed men and 34.0 of employed women worked part-time.[17] We take this into account when we interpret 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 earnings and log days worked. We then regress this on time-varying observable characteristics. These always include 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. In some specifications, we also include controls for realized experience. Our analysis focuses on these wage residuals.

5.2 Independence Assumptions

The identifying assumption A.W(a) and A.F(a) specify that we need each wage observation to be independent conditional on the worker identifier and conditional on the firm identifier and 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 36 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 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 as a single observation.[18] 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

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

[18] 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.

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 longest job during each employment spell.[19] This is independence assumption III: wages across employment spells are independent.

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 three employment spells (that is, workers with EUEUE transitions, where E represents an employment spell and U a registered unemployment spell). For these workers, we look alternatively at the first job, last job, and longest job during each employment spell. We call this independence assumption IV.

Our approach requires us to measure within and between wage inequality for both workers and firms, and so we need at least two observations for each. 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 her 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 only workers with multiple employers and employers with multiple workers. In our case the resulting data set is always non-empty.

6 Results

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 and of log wages. Different columns correspond to different independence assumptions.

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

[19] If a worker is ever recalled back to an old employer, we drop any intervening spells of unemployment from our analysis and so treat the entire episode as a single employment spell.

Estimated Correlation and Variances

	(1)	(2)	(3)	(4)	(5)	(6)
Men						
correlation of matched types $\hat{\rho}$	0.632	0.480	0.439	0.429	0.451	0.425
covariance of matched types $\hat{c}$	0.044	0.024	0.019	0.018	0.021	0.019
variance of worker types $\hat{\sigma}_\lambda^2$	0.073	0.047	0.039	0.038	0.042	0.039
variance of job types $\hat{\sigma}_\mu^2$	0.066	0.055	0.049	0.048	0.052	0.049
number of workers (thousands)	3,672	2,811	1,101	676	650	652
number of firms (thousands)	672	499	234	206	179	180
number of observations (thousands)	63,198	16,131	4,376	3,505	2,810	2,815
share of observations top-coded	0.186	0.134	0.078	0.060	0.033	0.041
Women						
correlation of matched types $\hat{\rho}$	0.608	0.390	0.418	0.424	0.457	0.435
covariance of matched types $\hat{c}$	0.087	0.036	0.028	0.027	0.032	0.028
variance of worker types $\hat{\sigma}_\lambda^2$	0.157	0.082	0.061	0.059	0.065	0.060
variance of job types $\hat{\sigma}_\mu^2$	0.130	0.104	0.075	0.071	0.074	0.071
number of workers (thousands)	3,128	2,359	951	540	503	504
number of firms (thousands)	760	522	238	196	160	162
number of observations (thousands)	46,635	11,103	3,190	2,336	1,771	1,773
share of observations top-coded	0.050	0.043	0.026	0.020	0.012	0.013
independence assumption	I	II	III	IV	IV	IV
observations included	all	all	longest	longest	first	last
first year of sample	1972	1972	1986	1986	1986	1986

Table 2: Estimates of correlations, covariances, and variances between matched workers' and firms' types for men. 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. For the naïve concept, we include all workers in the data. 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 three employment spells and treats either the longest (4), first (5), or last (6) job during each employment spell as independent observations. The last row in the table indicates the first year of the sample. The sample always ends in 2007.

our approach. This treats any two firm-year observations for a given worker as independent. We see that the correlation is high, above 0.6 for both men and women.

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 sharply, 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 (1) is biased. On the other hand, we lose a substantial number of 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 do the following experiment. We take firms and workers from column (2) and impose independence assumption I, that is, we again treat any two firm-year observations for a given worker as independent. We find that the correlation is 0.600 for men and 0.571 for women. This suggests that changes in the sample are not important and the difference between columns (1) and (2) is driven by the independence assumption and we therefore prefer estimates in column (2).

We next turn to independence assumption III, which treats wage observations as independent only if they are drawn from different employment spells, as in standard theories of on-the-job search. Column (3) shows a drop in the estimated correlation for men, and an increase for women but both changes are small which can reflect the combination of selection and bias. A first obvious difference is that the sample in (3) is shorter, and indeed this makes a difference. Estimating (2) on a sample covering years 1986–2007 gives the correlation of 0.539 for men and 0.439 for women. Furthermore, we take the sample of workers and firms from (3) and impose independence assumption II, that is, we use all matches of these workers. We find the correlation 0.461 for men and 0.432 for women. Thus, selection seems to play a more important role for men[20] (0.539 versus 0.461 versus 0.439) and bias for women (0.439 versus 0.432 versus 0.418), even though for women the effects are small.

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. From the perspective of Burdett and Mortensen (1998), the results in column (5) can be understood as measuring the correlation in the sampling distribution of wages, while those in column (6) should reflect the steady state distribution. These estimates are remarkably similar to the correlation in column (3), both for men and women.

[20]We want to explain the difference between 0.539 under II and 0.439 under III. Keeping the sample from column (3) and imposing the independence assumption II increases the correlation to 0.461 – this is the difference attributable to the bias from violating independence assumption. The rest, i.e. the difference between 0.539 and 0.461 can be attributed to sample selection.

In summary, the estimated correlation between types ranges from 0.425 to 0.632 for men, and from 0.390 to 0.608 for women. The exact number depends on the independence assumption. As we move from the independence assumption I to IV, the identifying assumption of conditionally independent and identically distributed wage observations is 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 recognize that sample selection probably biases the measured correlation down.

Column (3) shows that the standard deviation of worker types is 0.197 for men. The associated standard deviation of firm types is somewhat higher, 0.221. It follows that $\hat{\sigma}_\lambda > \hat{\rho}\hat{\sigma}_\mu$ and $\hat{\sigma}_\mu > \hat{\rho}\hat{\sigma}_\lambda$ and so Proposition 1 implies we are in the case where our correlation and the AKM correlation are equal. For women, both standard deviations are larger, 0.247 for workers and 0.274 for firms, but the conclusion is the same, $\rho_{AKM} = \rho$. This result holds in every specification in Table 2.

6.2 Confidence Intervals

We use a parametric bootstrap procedure to construct confidence intervals and examine the precision and accuracy of our estimator. Our main approach to the bootstrap involves constructing artificial data sets which differ from the actual data in terms of the exact number of workers and firms, the exact number of matches for each worker and firm, who matches with whom, and the wage paid in each match. The artificial data sets match the moments reported in Tables 2, including the variances of worker and firm types, the covariance of matched workers' and firms' types, 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. See Appendix C for details on the construction of the artificial data sets.

We construct $B = 500$ artificial data sets. For each data set $b = 1, \ldots, B$, we know each worker's and firm's type and so we can compare the actual correlation between types, ρ_b, with the correlation estimated using our approach, $\hat{\rho}_b$, which relies only on individual identifiers, wage data, and durations. We construct confidence intervals using the difference $\rho_b - \hat{\rho}_b$. 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 Table 2, column (3), the estimated correlation for men is $\hat{\rho} = 0.4912$, and the 95 percent confidence interval is $[0.4886, 0.4935]$. For women, the estimated correlation is $\hat{\rho} = 0.4290$ and the 95 percent confidence interval is $[0.4259, 0.4319]$. The results in the other columns are similar.

A drawback of this bootstrap procedure 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.

To capture this, we use an alternative bootstrap procedure which holds the set of matches fixed. Given the set of matches, we draw 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 $I+J$. This is computationally infeasible.[21] 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 would have generated estimates of $\hat{\rho}$ for men between -0.0098 and 0.0080. It is extremely unlikely that our data was generated from an economy without sorting.

6.3 Other Observable Characteristics

We now examine how controlling for fixed observable characteristics of workers and firms affects the estimated correlation. We start by reconsidering the assumption that the firm type is the same for all workers. Instead, imagine that a firm hires a collection of workers with different skills and the relevant firm type for a high skilled worker is potentially different than for a low type worker. Our approach effectively breaks a firm into different types for different skill levels and estimates the correlation on this adjusted data set. This differs from our approach in the time series and life cycle analysis, where we constructed a separate sample for each year or age. Although we could adopt that approach here, measuring the correlation within skill levels, this approach feels more natural to us when characteristics are fixed over time.

We start by treating a firm j as a cross between a firm identifier and an 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 Table **??**(3), i.e. using independence assumption III. We lose about ten percent of workers because they are missing education data, despite experiencing an unemployment spell.[22] We then drop some firms $\times$ education observations because they only appear once

[21] In the AKM fixed effects approach, types are known from the OLS estimates and only wages need to be generated for the bootstrap. This makes the bootstrap with a fixed network easy to perform. Confidence intervals are typically not reported in the literature, possibly because the AKM estimates are biased.

[22] 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 deviation higher than the average residual log wage of workers with recorded education. Furthermore, workers with missing education have

in the data set. This in turn forces us to drop some workers, etc. We then measure the correlation between the remaining worker and firm $\times$ education types.

Table 3 column (1) shows the results. Allowing firm types to differ by educational category raises the variance of firm types for both men and women. The bigger impact is on the covariance, and hence the correlation between matched types increases 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. 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 an individual at different positions as a different worker as well. This leads to an estimate of the correlation of 0.525 for men and 0.523 for women (Table 3 column (2)). 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, which are fixed at the firm level. We treat an individual with jobs in different industries as different workers. 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, ultimately about 38 percent of the observations for men and 37 percent for women. The correlation between the remaining matched workers and jobs is again higher, 0.580 for men and 0.527 for women (Table 3 column (3)).

6.4 Robustness

We first examine the sensitivity of our results to including work experience as an additional control when constructing the residual wages. We focus on results using independence assumption III. We construct work experience using the total number of days worked in the previous 14 years, taking advantage of data from before 1986 to get an accurate work history.[23] We then include a quartic polynomial in experience in addition to age and year dummies when we calculate the residual log wages. Column (1) of Table 4 shows the results for men and women. These are little changed from the corresponding results in column (4) of Table 2.

We next study the role of top-coding. In our baseline results, top-coding affects 7.8

fewer employment spells on average, 2.4 compared to 4.1 for men, and 2.3 compared to 3.4 for women.

[23]For example, in 1986, we measure experience as the number of days worked between 1972 and 1985.

Impact of Observables

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

Table 3: Results controlling for education, job classification, and industry. 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. In column (1), we treat each firm × education category as a separate firm. In column (2), we treat each worker × job position and firm × job position as different workers and firms. In column (3), we treat each worker × industry as different workers. The sample always runs from 1986–2007.

Robustness Results

	(1)	(2)	(3)
men			
correlation of matched types $\hat{\rho}$	0.405	0.439	0.417
covariance of mtched types $\hat{c}$	0.015	0.017	0.021
variance of worker types $\hat{\sigma}^2_\lambda$	0.031	0.034	0.043
variance of job types $\hat{\sigma}^2_\mu$	0.042	0.046	0.057
number of workers (thousands)	1,101	1,101	1,101
number of firms (thousands)	234	234	234
number of observations (thousands)	4,376	4,376	4,376
share of observations top-coded	0.078	0.117	0.078
women			
correlation of matched types $\hat{\rho}$	0.413	0.416	0.411
covariance of mtched types $\hat{c}$	0.026	0.027	0.027
variance of worker types $\hat{\sigma}^2_\lambda$	0.056	0.059	0.057
variance of job types $\hat{\sigma}^2_\mu$	0.069	0.073	0.075
number of workers (thousands)	951	951	951
number of firms (thousands)	238	238	238
number of observations (thousands)	3,190	3,190	3,190
share of observations top-coded	0.026	0.041	0.026
independence assumption	III	III	III
quartic in experience	yes	no	no
more severe top-code	no	yes	no
observations weighted equally	no	no	yes

Table 4: Robustness results for men and women. 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. In column (1) , we regress log wages on year, age, and a polynomial for work experience. Column (2) only regress log wages on year and age, but first reduce the top code by ten percent in each year. Column (3) again regresses log wages on year and age, but weighs all worker-firm observations equally by setting $t^w_{i,m} = t^f_{j,n} = 1$ for all i, j, m, n. All columns use independence assumption III, treating the longest jobs during each employment spell as independent observations. The sample always runs from 1986–2007.

percent of men's observations and 2.6 percent of women.[24] We ask here what would have happened if the top-coding threshold had been ten percent lower in every year, increasing the share of top-coded observations to 11.7 percent for men and 4.1 percent for women.[25]

Columns (2) of Table 4 show that more severe top-coding reduces the total variance of log wages as well as the estimated variance of both worker and firm types. It scarcely affects the estimated correlation $\hat{\rho}$ for women and mildly increases it for men. Appendix D examines what happens at other top-coding thresholds. We find 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, which suggests that in the absence of top-coding, the estimated correlation would be slightly higher.

Finally, instead of weighting each worker by his total duration, we weight all worker-firm matches equally. This corresponds to setting $t^w_{i,m} = 1$ for all $i = 1, \ldots, I$ and $m = 1, \ldots, M_i$ and $t^f_{j,n} = 1$ for all $j = 1, \ldots, J$ and $n = 1, \ldots, N_j$. Column (3) of Table 4 show that equally weighting all observations modestly reduces the estimated correlation. This is consistent with a higher correlation between worker and firm types in matches that last longer.

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 last job before the unemployment spell and the first job after the unemployment spell.[26]

Using only those workers who switch employers after an unemployment spell within a year reduces our sample size from 1.1 million workers to an average of 56 thousand workers per year for men, and from 1.0 million to 29 thousand 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

[24] We consider the log wage for a worker-firm pair to be top-coded if at least one annual wage 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 doubles to 17.5 percent for men and 5.4 percent for women if we weigh each observation by its duration.

[25] The usual approach to dealing with top-coded data 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.

[26] Appendix **??** shows the estimated time series correlation on data constructed using independence assumption II. This allows us to study the full time period from 1972–2007. The pattern for years 1986–2007 is similar, but it also reveals a large increase in the correlation in years 1972–1986 for men.

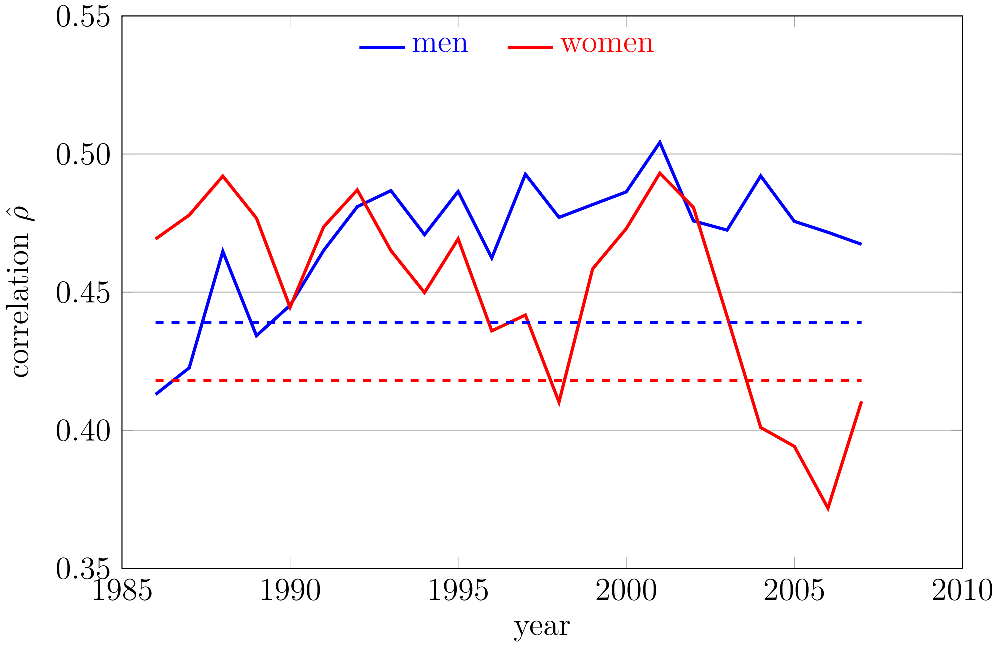

Figure 3: Correlation between worker and firm types using residual log wages under independence assumption III. Solid lines are computed year-by-year and shaded areas are bootstrapped 95 percent confidence intervals. For each year, the sample considers all workers who switched employers after an unemployment 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. Dashed lines are computed using the full sample, reported in Table 2, column (3).

0.46 in 1986 to around 0.55 in 1997, where it stayed until the last two years of the sample. The figure also shows that the correlation for women fluctuated over time, peaking at 0.52 in 2001 and then falling thereafter. 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.

Interestingly, the annual correlations average 0.53 for men and 0.47 for women, significantly more than the correlations of 0.49 and 0.43 reported in column (4) of Table 2using the full sample. We see two possible reasons for this. 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 address this, we pool the samples from the time series analysis and estimate a single correlation, 0.425 for men and 0.420 for women.[27] Sample differences are unimportant for women and

[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 624,917 men and 408,614 women.

actually enlarge the gap between the average annual correlation and the pooled correlation for men.

The second possibility is that types gradually change over time, so a worker's expected log wage when young is not the same as when old, even after accounting for the usual effect of aging on wages. This effectively makes λ and μ into noisy measures of the worker's and firm's types at a point in time, reducing the measured correlation; see Appendix E for details. This logic suggests that the annual observations more accurately reflect the correlation between worker and firm types at a point in time.

One possible concern with the results in this section is that, although the wage in the first and last job within an employment spell are independent, they are not drawn from the same distribution. Indeed, 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.

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,

$$\omega_{i,m}^{w} = x_{i,m}'\beta + \alpha_i + \psi_{k_{i,m}} + v_{i,m}, \tag{2}$$

where $x_{i,m}$ is a vector of match-varying observable characteristics for worker i and $k_{i,m}$ is the identifier of the firm that employs i in her m^{th} match. 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 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 5 verify that this finding holds in our data as well. 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. We estimate our correlation on this set. Comparing results in Tables 2 and 5 we see that this has little impact on estimates of the correlation using our approach.

Columns in Tables 5 correspond to the data sets used in Table 2, with the additional

Comparison with AKM

	(1)	(2)	(3)	(4)	(5)	(6)
Men						
$\hat{\rho}$	0.622	0.480	0.439	0.429	0.451	0.425
$\hat{\rho}_{AKM}$	0.035	0.057	0.033	0.033	0.015	-0.002
number of workers (thousands)	3,651	2,810	1,100	676	650	652
number of firms (thousands)	650	498	234	206	179	180
number of observations (thousands)	63,043	16,129	4,375	3,505	2,810	2,815
share of observations top-coded	0.186	0.134	0.078	0.060	0.033	0.041
Women						
$\hat{\rho}$	0.597	0.389	0.418	0.424	0.457	0.435
$\hat{\rho}_{AKM}$	0.007	0.068	0.037	0.067	0.055	0.038
number of workers (thousands)	3,088	2,358	951	540	503	504
number of firms (thousands)	716	522	238	196	160	162
number of observations (thousands)	46,275	11,101	3,190	2,336	1,771	1,773
share of observations top-coded	0.050	0.043	0.026	0.020	0.012	0.013
independence assumption	I	II	III	IV	IV	IV
observations included	all	all	longest	longest	first	last
first year of the sample	1972	1972	1986	1986	1986	1986

Table 5: Comparison of our estimates of correlation and AKM fixed effects estimates for men. The AKM correlation as well as correlation estimated using our method are estimated on the largest connected set. 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. For the naïve concept, we include all workers in the data. 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 three employment spells and treats either the longest (4), first (5), or last (6) job during each employment spell as independent observations. The last row in the table indicates the first year of the sample. The sample always ends in 2007. The asterix indicates that these are preliminary results.

restriction to the largest connected set. Using the fixed effects approach, the estimated correlation lies between -0.002 and 0.057 for men and 0.007 and 0.068 for women. Across the seven columns, the fixed effects correlation is about 0.50 below our estimate of the correlation for men and 0.45 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? We can think of three possible reasons. First, the two measures are conceptually different and hence could give different answers. Proposition 1 establishes that if the joint distribution of AKM fixed effects is elliptical, then our correlation should be equal to the true AKM correlation. Moreover, Section 3.2 showed that even in models where the joint distribution is not necessarily elliptical but the identifying assumptions of AKM are (almost) satisfied, our correlation and AKM correlation are close. Nevertheless, Section 3.3 gives an example of a model where the measures of correlation are very different. This may explain some of what is happening.

Second, identifying assumptions in the AKM approach are violated. This could be either because the wage equation is misspecified or because there is an "endogenous mobility" problem. We believe that the endogenous mobility assumption might not be very important, though. In the version of Shimer and Smith (2000) with match specific productivity shock this assumption is violated due to a selection problem: some matches are only formed if the match specific shock is high while other matches are formed with a bigger set of shocks. Nevertheless, Figure 1 shows that in practice this might not have a big impact.

On the other hand, misspecification of the wage equation might have tremendous impact on results. The discrete choice model examined in Section 3.3 shows how AKM correlation might not be a good measure of sorting when the wage equation is misspecified. If the Austrian data are not generated from an economy with (almost) log-linear wage equation, this is likely to be an important factor explaining the difference.

Finally, even if the identifying assumptions in the AKM approach are valid, the estimator of the AKM correlation is consistent only in the limit as the number of observations per worker and firm goes to infinity holding fixed the number of workers and firms (Postel-Vinay and Robin, 2006; Andrews, Gill, Schank and Upward, 2008). This is not a natural feature of real-world data sets. For example, even using 36 years of Austrian data, we find that the median worker has two employers and the median firm has three employees. This creates an incidental parameter problem which causes bias and inconsistency in the measurement of the correlation between the AKM fixed effects. Andrews, Gill, Schank and Upward (2008) derive a bias correction for the AKM correlation under some auxiliary assumptions, e.g. homoskedasticity of the error term in the wage equation.[28] We evaluate

[28]We use formulas (18), (20), (24), and (25) in Andrews, Gill, Schank and Upward (2008). These formulas

this formula using the samples from columns (3)–(6) of Table 5 and find that the results are barely affected for men and increase to about correlation of 0.1 for women. We cannot calculate the correction for columns (1)–(2) due to the sample size. Some authors suggest that the reason for a small bias correction is the assumption of homoscedastic errors Andrews, Gill, Schank and Upward (2008). Kline, Saggio and Sølvsten (2019) derive a bias correction when errors are heteroscedastic and find larger corrections, but unfortunately this correction is not implementable in datasets of our size.

8 Conclusion

This paper proposes and implements a simple, precise, and accurate approach to measuring whether high wage workers work for high wage firms. Using Austrian data, we find that they do. The correlation between a worker's type and her employer's type lies between 0.4 and 0.6 and is reasonably stable over time. We contrast our results with the existing literature based on the AKM fixed effects estimator. We show that the AKM estimator is significantly biased even in data sets with many worker and firm observations, due to the incidental parameter problem. This has led the previous literature to the incorrect conclusion that there is little sorting of high wage workers into high wage jobs.

Is a correlation of 0.4 to 0.6 large? This is a quantitative question that goes beyond the scope of this paper. Still, there are reasons to think that the true correlation is even larger. We have previously noted three reasons why our approach likely understates the true correlation: 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). Our estimated correlations therefore suggest that the labor market is very effective at getting the highest wage workers working together at the highest wage firms.

depend on the variance of the error term in the AKM wage equation, σ^2_η, and one needs to plug in a consistent estimate for the bias correction to work. We do not use the usual estimator of the variance based on the residuals from the AKM equation, because it is not consistent again due to incidental parameter problem. Instead, we use that equation (1) implies $Var[w_{i,j}] = Var[\alpha_i] + 2Cov[\alpha_i, \psi_j] + Var[\psi_j] + \sigma^2_{\eta_{i,j}}$. A consistent (bias-corrected) estimators of the first three terms on the right-hand side are linear in unknown σ^2_η. We can thus easily solve this equation for σ^2_η and use it as our estimate of the error variance in the formulas for the bias correction. Using Monte-Carlo simulations we verified that bias-corrected correlation calculated this way is unbiased.

References

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

Abowd, John M., Francis Kramarz, and David N. Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 1999, *67* (2), 251–333.

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

Andrews, Martyn J., Leonard Gill, Thorsten Schank, and Richard Upward, "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)*, 2008, *171* (3), 673–697.

_ , _ , _ , and _ , "High Wage Workers Match with High Wage Firms: Clear Evidence of the Effects of Limited Mobility Bias," *Economics Letters*, 2012, *117* (3), 824–827.

Arellano, Manuel and Stéphane Bonhomme, "Identifying Distributional Characteristics in Random Coefficients Panel Data Models," *Review of Economic Studies*, 2011, *79* (3), 987–1020.

Bagger, Jesper, François Fontaine, Fabien Postel-Vinay, and Jean-Marc Robin, "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics," *The American Economic Review*, 2014, *104* (6), 1551–1596.

_ , Kenneth L Sørensen, and Rune Vejlin, "Wage Sorting Trends," *Economics Letters*, 2013, *118* (1), 63–67.

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

Burdett, Kenneth and Dale Mortensen, "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 1998, *39* (2), 257–73.

Card, David, Jörg Heining, and Patrick Kline, "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics*, 2013, *128* (3), 967–1015.

Christensen, Bent Jesper, Rasmus Lentz, Dale T. Mortensen, George R. Neumann, and Axel Werwatz, "On-the-Job Search and the Wage Distribution," *Journal of Labor Economics*, 2005, *23* (1), 31–58.

Eeckhout, Jan and Philipp Kircher, "Identifying Sorting—In Theory," *Review of Economic Studies*, 2011, *78* (3), 872–906.

Garicano, Luis, "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 2000, *108* (5), 874–904.

Goussé, Marion, Nicolas Jacquemet, and Jean-Marc Robin, "Marriage, labor supply, and home production," *Econometrica*, 2017, *85* (6), 1873–1919.

Gruetter, Max and Rafael Lalive, "The Importance of Firms in Wage Determination," *Labour Economics*, 2009, *16* (2), 149–160.

Hagedorn, Marcus, Tzuo Hann Law, and Iourii Manovskii, "Identifying Equilibrium Models of Labor Market Sorting," *Econometrica*, 2017, *85* (1), 29–65.

Iranzo, Susana, Fabiano Schivardi, and Elisa Tosetti, "Skill Dispersion and Firm Productivity: An Analysis with Employer-Employee Matched Data," *Journal of Labor Economics*, 2008, *26* (2), 247–285.

Jochmans, Koen and Martin Weidner, "Fixed-Effect Regressions on Network Data," 2017. mimeo.

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

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

Ljungqvist, Lars and Thomas J. Sargent, "The European Unemployment Dilemma," *Journal of Political Economy*, 1998, *106* (3), 514–550.

Lopes de Melo, Rafael, "Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence," *Journal of Political Economy*, forthcoming.

Malmberg, Hannes, "Random Choice over a Continuous Set of Options," 2013. mimeo.

Pichelmann, Karl and Monika Riedel, "New Jobs or Recalls?," *Empirica*, 1992, *19* (2), 259–274.

Postel-Vinay, Fabien and Jean-Marc Robin, "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 2002, *70* (6), 2295–2350.

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

Shimer, Robert and Lones Smith, "Assortative Matching and Search," *Econometrica*, 2000, *68* (2), 343–369.

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

Tiebout, Charles M., "A Pure Theory of Local Expenditures," *Journal of Political Economy*, 1956, *64* (5), 416–424.

Weiss, Gideon, "Time-Reversibility of Linear Stochastic Processes," *Journal of Applied Probability*, 1975, *12* (4), 831–836.

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

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. Assume that the joint distribution of α and ψ is elliptical, that is, the associated density function ξ can be expressed as

$$\xi(\alpha,\psi) = \tilde{\xi}\left(\frac{(\alpha-\bar{\alpha})^2}{\sigma_\alpha^2} - \frac{2\rho_{AKM}(\alpha-\bar{\alpha})(\psi-\bar{\psi})}{\sigma_\alpha\sigma_\psi} + \frac{(\psi-\bar{\psi})^2}{\sigma_\psi^2}\right)$$

for some function $\tilde{\xi}$.

We first prove that the expected value of α conditional on ψ is $\theta_0+\theta_1\psi$, where $\theta_0 = \bar{\alpha}-\zeta\bar{\psi}$, $\theta_1 = \zeta$, and $\zeta \equiv \rho\sigma_\alpha/\sigma_\psi$. Towards this end, take any point (α_1,ψ) and let $\alpha_2 \equiv 2\big(\bar{\alpha}+\zeta(\psi-$

$\bar{\psi})\big) - \alpha_1$, so the mean of α_1 and α_2 is $\bar{\alpha} + \zeta(\psi - \bar{\psi})$. The definition of an elliptical distribution implies $\xi(\alpha_1, \psi) = \xi(\alpha_2, \psi)$. Using this, the conditional expected value satisfies

$$
\begin{aligned}
\frac{\int_{-\infty}^{\infty} \alpha\xi(\alpha,\psi)d\alpha}{\int_{-\infty}^{\infty} \xi(\alpha,\psi)d\alpha} &= \frac{\int_{-\infty}^{\bar{\alpha}+\zeta(\psi-\bar{\psi})} \alpha\xi(\alpha,\psi)d\alpha + \int_{\bar{\alpha}+\zeta(\psi-\bar{\psi})}^{\infty} \alpha\xi(\alpha,\psi)d\alpha}{\int_{-\infty}^{\bar{\alpha}+\zeta(\psi-\bar{\psi})} \xi(\alpha,\psi)d\alpha + \int_{\bar{\alpha}+\zeta(\psi-\bar{\psi})}^{\infty} \xi(\alpha,\psi)d\alpha} \\
&= \frac{\int_{-\infty}^{\bar{\alpha}+\zeta(\psi-\bar{\psi})} \alpha\xi(\alpha,\psi)d\alpha + \int_{-\infty}^{\bar{\alpha}+\zeta(\psi-\bar{\psi})} \big(2\big(\bar{\alpha}+\zeta(\psi-\bar{\psi})\big)-\alpha\big)\xi(\alpha,\psi)d\alpha}{2\int_{-\infty}^{\bar{\alpha}+\zeta(\psi-\bar{\psi})} \xi(\alpha,\psi)d\alpha} \\
&= \frac{\int_{-\infty}^{\bar{\alpha}+\zeta(\psi-\bar{\psi})} 2\big(\bar{\alpha}+\zeta(\psi-\bar{\psi})\big)\xi(\alpha,\psi)d\alpha}{2\int_{-\infty}^{\bar{\alpha}+\zeta(\psi-\bar{\psi})} \xi(\alpha,\psi)d\alpha} = \bar{\alpha}+\zeta(\psi-\bar{\psi})
\end{aligned}
$$

The first expression defines the conditional expectation. The first equality breaks the integrals into two terms. The second equality uses the key property of the elliptical distribution, $\xi(\alpha, \psi) = \xi(2(\bar{\alpha} - \zeta(\psi - \bar{\psi})) - \alpha, \psi)$, which allows us to change the variable of integration in the second integral in both the numerator and denominator. The third equation adds to the two integrands in the numerator. The fourth equation uses the fact that the integrand is constant.

A symmetric proof implies that the expected value of ψ conditional on α is $\bar{\psi} + \frac{\rho\sigma_\psi}{\sigma_\alpha}(\alpha - \bar{\alpha}) = \kappa_0 + \kappa_1\alpha$. The logic in the body of the paper then implies $\lambda = \kappa_0 + (1+\kappa_1)\alpha$ and $\mu = \theta_0 + (1+\theta_1)\psi$, with the coefficients given in equations (3) and (4),

$$\lambda_i = \bar{\psi} - \frac{\rho_{AKM}\sigma_\psi}{\sigma_\alpha}\bar{\alpha} + \left(1 + \frac{\rho_{AKM}\sigma_\psi}{\sigma_\alpha}\right)\alpha_i, \tag{3}$$

$$\mu_j = \bar{\alpha} - \frac{\rho_{AKM}\sigma_\alpha}{\sigma_\psi}\bar{\psi} + \left(1 + \frac{\rho_{AKM}\sigma_\alpha}{\sigma_\psi}\right)\psi_j. \tag{4}$$

If $\sigma_\alpha + \rho_{AKM}\sigma_\psi$ and $\sigma_\psi + \rho_{AKM}\sigma_\alpha$ are both positive, then λ_i is a linearly increasing function of α_i and μ_j is a linearly increasing function of ψ_j. Therefore the correlation between λ and μ is the same as the correlation between α and ψ, $\rho = \rho_{AKM}$. Moreover, equations (3) and (4) imply that the standard deviations of λ and ψ are $\sigma_\lambda = \sigma_\alpha + \rho_{AKM}\sigma_\psi$ and $\sigma_\mu = \sigma_\psi + \rho_{AKM}\sigma_\alpha$, both positive by the assumption at the start of this paragraph. Using this and $\rho = \rho_{AKM}$ gives us $\sigma_\lambda - \rho\sigma_\mu = \sigma_\alpha(1 - \rho_{AKM}^2) > 0$, and $\sigma_\mu - \rho\sigma_\lambda = \sigma_\psi(1 - \rho_{AKM}^2) > 0$. Hence indeed $(\sigma_\lambda - \rho\sigma_\mu)(\sigma_\mu - \rho\sigma_\lambda) > 0$.

Now suppose that $\sigma_\alpha + \rho_{AKM}\sigma_\psi > 0 > \sigma_\psi + \rho_{AKM}\sigma_\alpha$. Then λ_i is a linearly increasing function of α_i and μ_j is a linearly decreasing function of ψ_j. Therefore $\rho = -\rho_{AKM}$. Equations (3) and (4) imply that the standard deviations of λ and ψ are $\sigma_\lambda = \sigma_\alpha + \rho_{AKM}\sigma_\psi$ and $\sigma_\mu = -(\sigma_\psi + \rho_{AKM}\sigma_\alpha)$. Using this and $\rho = -\rho_{AKM}$ gives us $\sigma_\lambda - \rho\sigma_\mu = \sigma_\alpha(1 - \rho_{AKM}^2) > 0$ and $\sigma_\mu - \rho\sigma_\lambda = -\sigma_\psi(1 - \rho_{AKM}^2) < 0$. This proves $(\sigma_\lambda - \rho\sigma_\mu)(\sigma_\mu - \rho\sigma_\lambda) < 0$. The case with

$\sigma_\psi + \rho_{AKM}\sigma_\alpha > 0 > \sigma_\alpha + \rho_{AKM}\sigma_\psi$ is analogous.

Finally, if $\sigma_\alpha + \rho_{AKM}\sigma_\psi = 0$, equations (3) and (4) imply $\sigma_\lambda = 0$. If $\sigma_\psi + \rho_{AKM}\sigma_\alpha = 0$, then $\sigma_\mu = 0$. In either case, the correlation between λ and μ is undefined. ■

A.2 Shimer and Smith (2000) with Match Specific Shocks

We formulate equations for value functions $U(x), V(y)$ and steady state conditions for $u(x), v(y)$. Since it will become useful, define notation for 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 then is

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

Similarly, the value of a vacant firm is

$$rV(y) = \frac{\theta(1-\gamma)}{r+\delta} \int_X p(\underline{z}(x,y)) \big(\omega(\underline{z}(x,y)) H(x,y) - rU(x) - rV(y) \big) u(x) dF(x).$$

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

$$\delta(1-u(x)) = \theta u(x) \int_Y p(\underline{z}(x,y)) v(y) dG(y),$$
$$\delta(1-v(y)) = \theta v(y) \int_X p(\underline{z}(x,y)) u(x) dF(x).$$

The fraction of firm y's matches that are with worker x is proportional to $p(\underline{z}(x,y))u(x)dF(x)$. Thus the likelihood ratio of y and y' matching with x is proportional to $\frac{p(\underline{z}(x,y))}{p(\underline{z}(x,y'))}$. With an exponential distribution, $\zeta(z) = 1 - \exp(-z/s)$, the log-likelihood ratio is $(\underline{z}(x,y') - \underline{z}(x,y))/s$, increasing in $\underline{z}(x,y') - \underline{z}(x,y) = \frac{rU(x)+rV(y')}{H(x,y')} - \frac{rU(x)+rV(y)}{H(x,y)}$. There is no general monotonicity of this expression, i.e. it may be the case that some firms hire disproportionately many low productivity workers and others hire disproportionately many high productivity workers.

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

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

If ζ has an exponential distribution, this is

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

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] The model therefore breaks the link between the probability of matching and the expected log wage.

A.3 Discrete Choice Model

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

The result in Malmberg (2013) implies that the distribution of firm types y conditional on worker's type x is

$$\Phi_x(y) = \frac{1}{k_1} \exp\left(\frac{w(x,y)}{s}\right) dG(y),$$

where k_1 is the normalization which assures that $\int_0^1 d\Phi_x(y) = 1$. Under the assumption that y is standard normal distribution, we get that $\Phi_x(y)$ is also normal

$$y|x \sim N\left(\frac{2ax}{s+2a}, \frac{s}{s+2a}\right).$$

We use Bayes formula to find the distribution of x conditional on y, $d\Psi_y(x) \sim d\Phi_x(y)dF(x)$, which turns out to be again normal,

$$x|y \sim N\left(\frac{2a(s+2a)y}{s^2+2as+4a^2}, \frac{s^2+2as}{s^2+2as+4a^2}\right).$$

The knowledge of conditional distribution allows us to compute the types $\lambda(x), \mu(y)$ as well as AKM types $\alpha(x), \psi(y)$.

Finally, we need the joint distribution of (x, y) to compute the correlation between types. The joint distribution is the product of the conditional distribution $d\Phi_x(y)$ and the marginal

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

$dF(x)$. Since both of these are normal, it follows that the joint is also normal.

B Properties of Estimators

B.1 Consistency

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. 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 $\hat{\lambda}_i$. 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 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 $\frac{\sum_{i=1}^{\tau I} T_i^w \hat{\lambda}_i}{\sum_{i=1}^{\tau I} T_i^w}$ is a consistent estimator of $\bar{\lambda} \equiv \frac{\sum_{x \in X} I_x \bar{T}_x^w \lambda(x)}{\sum_{x \in X} I_x \bar{T}_x^w}$.

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 $\widehat{\lambda_i}^2$, 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. ■

Proof of Proposition 5. We start by expanding the definition of $\hat{c}_{i,m}$ using $w_{i,m'}^w = \bar{w}_{x_i}^w + \varepsilon_{i,m'}^w$ and $w_{j,n'}^f = \bar{w}_{y_j}^f + \varepsilon_{j,n'}^f$ with $j = \mathbf{j}_{i,m}$:

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

Now compute the average of this across all matches:

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

The first equation uses the definition 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

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

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 $\frac{\sum_{x\in X} I_x\mathbb{E}_x\sum_{m=1}^{M_i} t^w_{i,m}\lambda(x)\mu(y_{i,m})}{\sum_{x\in X} I_x\bar{T}^w_x}$. Finally, we have already shown in the proof of Proposition 3 that $\frac{\sum_{i=1}^{\tau I} T^w_i\hat{\lambda}_i}{\sum_{i=1}^{\tau I} T^w_i}$ is a consistent estimator of $\bar{\lambda}$. Since $\bar{\lambda} = \bar{\mu}$, it is a consistent estimator for this as well. ■

B.2 Finite Sample Properties

B.2.1 Construction of Artificial Datasets

We generate artificial datasets from two structural models, Shimer and Smith (2000) with match specific shocks and the discrete choice model.

In Shimer and Smith (2000), we proceed as follows. We solve the steady state of the model and use steady state decision rules, value functions and distribution of unemployed and vacancies to create an artificial dataset. We choose the number of worker and firms, I, J, and assign each worker and firm its type according to unconditional distribution of types F, G. We start the economy with some workers employed and some unemployed, respecting their type-specific unemployment rates. For each unemployed worker, we determine whether he gets an opportunity to meet a vacant firm. If so, firm's type y is drawn from the distribution of vacancies $v(\cdot)$. The worker-firm pair then draws a match specific productivity from $\zeta(\cdot)$ and determines whether to create a match or not using the steady state decision rule. If they decide to create a match, we assign firm's name to this match according to how many firms of that type exist in the economy. If there are K firm with that type, then each firm gets this worker with probability $1/K$. Every match breaks at the rate δ, in which case worker become unemployed. We repeat this sequence of step for each period. We repeat these steps for several periods; we choose number of periods so that the median worker holds 4.9 jobs.

We proceed a little differently in the discrete choice model. We choose number of workers and firms, I, J and draw the type for each of them from the unconditional distributions F, G, respectively. For each worker, we further draw number of jobs he will hold, using the actual distribution of jobs per worker in our dataset. We draw the firm type from the worker's conditional distribution of jobs. We assume that each match has the same duration.

In both cases, if a worker ends up having multiple jobs with the same firm, we keep only one. We drop all workers and firms with only one observations. In the final dataset, we compute the "true correlation" using the types λ and μ computed in the infinite (steady state) economy and then also estimate the correlation using only the wage data. We report the distribution of the error.

B.2.2 Parametric Bootstrap

We construct artificial datasets from the structural models as described above.

For each set of parameters, we construct $B = 100$ samples. In each sample, we compute the realized correlation ρ_b as well as its estimate $\hat{\rho}_b$. Let $e_b = \hat{\rho}_b - \rho_b$ be the 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 $[\rho + \underline{e}, \rho + \bar{e}]$. Note that the interval does not have to be centered. We use $I = 10,000$ workers and $J = 2,000$ firms for data creation, so rather conservative values. Nevertheless, we observe that the confidence intervals are centered and very tight.

We should be precise in saying that we start the economy with I workers and J firms, but the number of workers and firms in the resulting sample might be lower. To apply our estimator, we only keep workers and firms with at least observations, which means that some of the initial workers and/or firms might have to be dropped.

Figure 4 shows the confidence intervals in the Shimer, Smith model with match productivity shocks. We see that intervals are very tight for across the entire range of parameter values we consider.

Figure 5 shows the confidence intervals in discrete choice model. The confidence intervals are again very tight, with the exception of when the difference in means $m_x - m_y$ exceeds 1.5. The reason is the size of the sample. 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 most firms end up with one or zero workers, and many workers have multiple jobs at the same firm. Our estimator requires that each worker in the dataset has at least two distinct employers and each firm employs at least two distinct workers. In this environment, only very few workers and firms satisfy this requirement, and hence we end up with samples of approximately $I = 2,000$ workers and $J = 10$ firms, and those seem to be also highly selected.

C Standard Errors

We use bootstrap to construct standard errors.

C.1 Constructing Artificial Data

We construct artificial data sets that match a few key moments: 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, and the distribution of the number of matches per worker M_i, the number of matches per firm N_j and the joint distribution of durations $t^w_{i,\cdot}$. We draw these from our estimates, e.g. in Tables **??** and **??**, and we take distributions of $M, N, t^w_{i,\cdot}$ directly from the data.

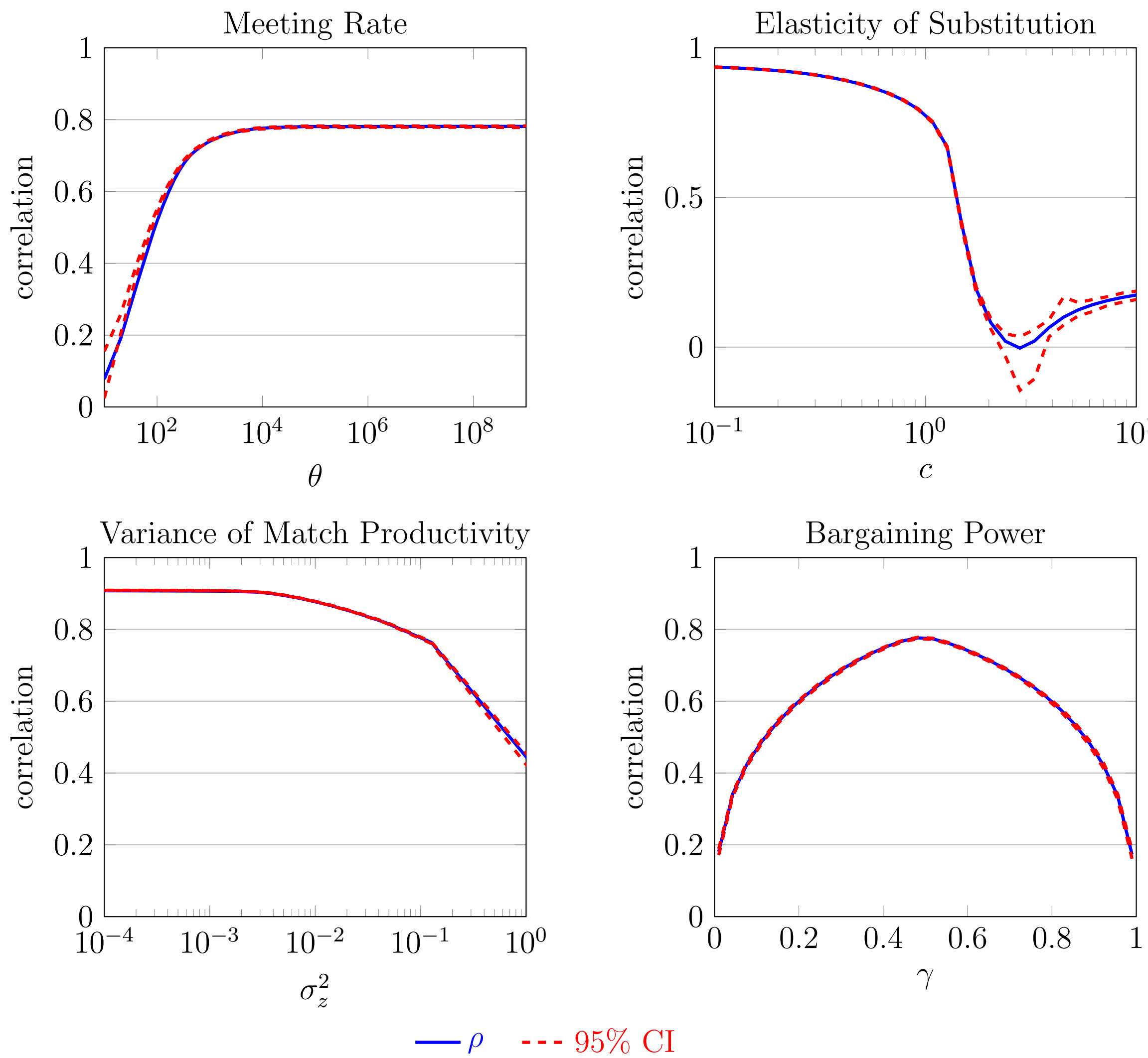

Figure 4: Confidence intervals in Shimer and Smith (2000) with match specific shocks. We plot the correlation between λ and μ in an infinite sample (blue line) and the bootstrapped confidence intervals. For the given set of parameter values, we create $B = 100$ artificial samples from the discrete choice model with $I = 10,000$ workers and $J = 2,000$ firms. In each sample we compute the error, the difference between the estimated and realized correlation, and use the 2.5% and 97.5% quantile of the error distribution to construct the confidence interval.

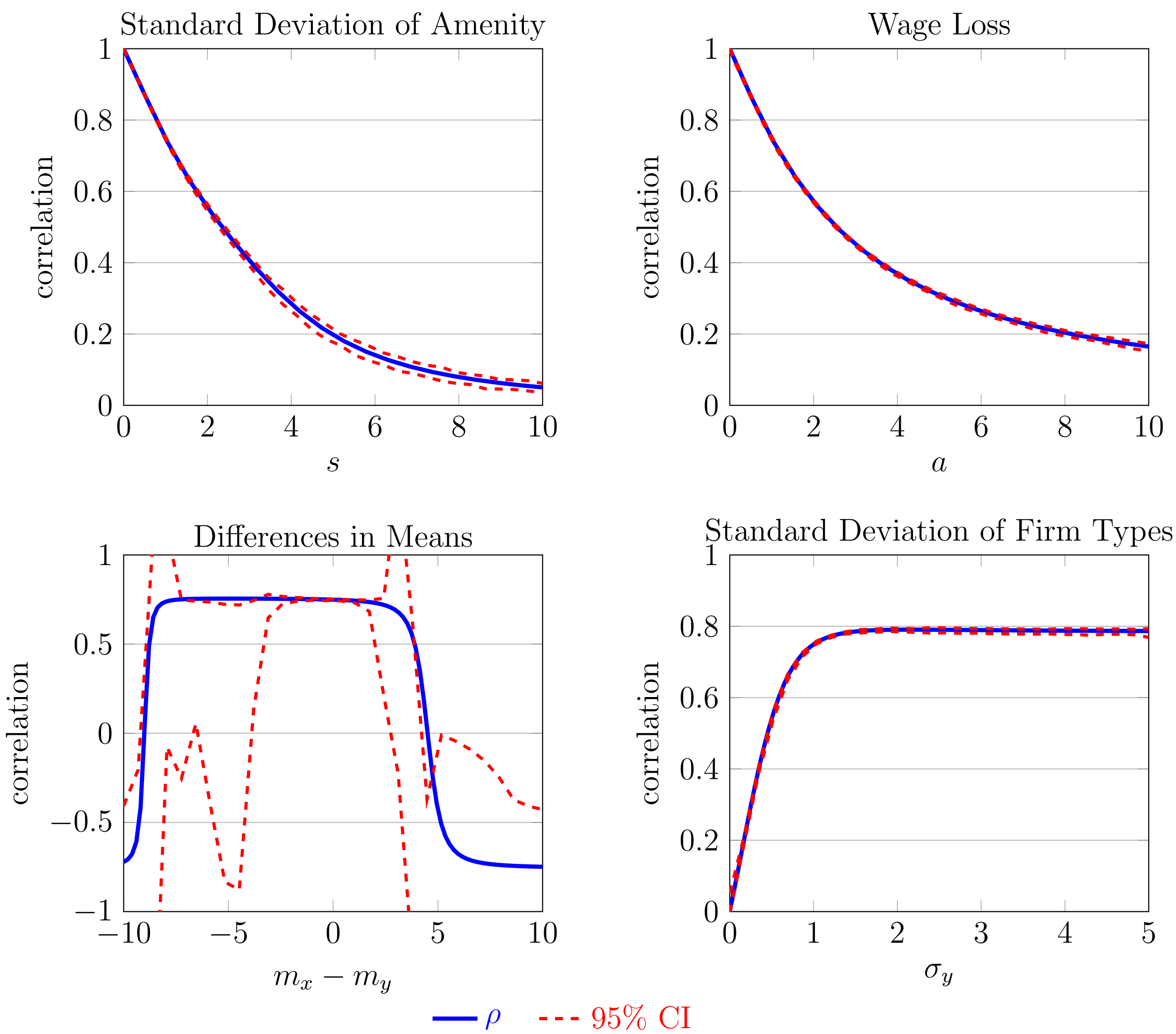

Figure 5: Confidence intervals in the discrete choice model. We plot the correlation between λ and μ in an infinite sample (blue line) and the bootstrapped confidence intervals. For the given set of parameter values, we create $B = 100$ artificial samples from the discrete choice model with $I = 10,000$ workers and $J = 2,000$ firms. In each sample we compute the error, the difference between the estimated and realized correlation, and use the 2.5% and 97.5% quantile of the error distribution to construct the confidence interval.

In each iteration of the bootstrap $b \in \{1, \ldots, B\}$, we construct an artificial data set that replicates these moments, use it to measure the correlation between λ and μ in matches, ρ_b, and then use it to estimate the correlation using our procedure, giving us $\hat{\rho}_b$. In practice, ρ, ρ_b, and $\hat{\rho}_b$ will not be the same. The difference between the first two reflects the fact that the artificial data set is finite. The difference between the latter two reflects limitations in our estimator. We focus on this difference.

We proceed as follows:

1. We choose the number of workers $\tilde{I}$ and firms $\tilde{J}$ as in the data.

2. For each worker $i \in \{1, \ldots, \tilde{I}\}$ 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 job directly from the data. For each $j \in \{1, \ldots, \tilde{J}\}$, we draw the number of employees N_j. We use the distribution of N from the data. The model imposes the restriction that $\sum_i M_i = \sum_j N_j$. We start with large $\tilde{I}$ and $\tilde{J}$ and 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 $I \geq \tilde{I}$ workers and $J \geq \tilde{J}$ firms.

3. For each worker i (firm j), we choose a random λ_i (μ_j) from a normal distribution with mean 0 and variance σ^2_λ (σ^2_μ).

4. 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 $\frac{\lambda_i \rho \sigma_\mu}{\sigma_\lambda}$ and variance $\sigma^2_\mu(1-\rho^2)$. We rank these values. 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.

6. We drop any duplicate matches between i and j. If this leaves us with any workers or firms with a single match, we drop those as well.

7. We measure correlation ρ_b using types λ and μ, and the job durations t^w.

8. We compute the log wage. For worker i's m^{th} job, the log wage is $\omega^w_{i,m} = a\lambda_i + b\mu_{k_{i,m}} + v_{i,m}$, 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

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

and the variance of the log wage shock satisfies

$$\sigma^2_v = \sigma^2 - \frac{\sigma^2_\lambda + \sigma^2_\mu - 2\rho\sigma_\lambda\sigma_\mu}{1-\rho^2}.$$

9. We estimate $\hat{\rho}_b$ using our approach (as described in the text).

10. We find the largest connected set and keep only workers and firms in this set. We estimate $\hat{\rho}_{AKM,b}$ following AKM methodology.

11. We are primarily interested in $\delta_b = \hat{\rho}_b - \rho_b$ and $\delta_{AKM,b} = \hat{\rho}_{AKM,b} - \rho_b$, the difference between the estimated and true correlation in the b^{th} sample.

We construct $B = 500$ samples and find values $\underline{\delta}$ and $\bar{\delta}$ such that

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

The 95 percent confidence interval for ρ is $[\rho + \underline{\delta}, \rho + \bar{\delta}]$. Note that this will not be centered around ρ if the estimator is biased. In our case, it is centered and the difference $\bar{\delta} - \underline{\delta}$ is small.

We similarly construct confidence intervals using $\delta_{AKM,b}$. These turn out not to be centered around ρ, reflecting the bias in the AKM estimate of the correlation between fixed effects.

Finally, we can use the same procedure to bootstrap confidence intervals around other parameters, e.g. σ_λ and σ_μ.

Our procedure assumes that worker and firm types are homoscedastic but it is straightforward to relax this assumption. We have constructed artificial data sets where types are correlated with the number of observations. In particular, we assume 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 measure the conditional distributions directly from the data, following the approach in Section **??**. Our estimated confidence interval for ρ is robust to this assumption.

C.2 Properties of the Artificial Data

This section shows that ρ_b, constructed as described above, is equal to ρ in an infinitely large data set. We do this by finding all the first and second moments:

1. The unconditional mean of $\chi_{i,m}$ is 0 by the law of iterated expectations.

2. 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

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

Thus the distribution of $\chi_{i,m}$ and μ_j are the same and hence $\mu_{i,m} = \mu_{k_{i,m}}$, the type of the firm that employs i in her m^{th} match.

3. The expected value of $\lambda\mu$ conditional on λ is $\lambda^2\rho\sigma_\mu/\sigma_\lambda$. Thus the unconditional expected value is $\rho\sigma_\mu\sigma_\lambda$. This is the covariance between λ and μ.

4. The correlation is the ratio of the covariance to the product of the two standard deviations, and hence is ρ.

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. Starting from the wage cap as in the data, we decrease it gradually by 2 percent, 4 percent,... , and up to 40 percent. We then censor wages at the wage cap, construct data using Concept III as described in the main text and estimate the correlation and variances.

Figure 6 shows the results. In the top row, we display the estimated correlation $\hat{\rho}$ for data sets with different top-coding as a function of the share of top-coded observations. For women, the correlation varies very mildly, staying around 0.43 even when almost 20 percent of observations are top-coded.

Top-coding matters for men. Setting the maximum wage to 40 percent of what it is in Austria increases the share of top-coded observations from 7.8 percent to 43.5 percent, and results in an increase of the correlation from 0.491 to 0.864.

Our intuition is that the impact of top-coding on 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. Viewed through this lens, the correlation among high-wage women is similar to the rest. For men, it is useful to think about the components of the correlation separately. The covariance (not plotted) decreases with top coding from initial 0.018 to 0.007 when top code is 40 percent of the top wage in Austria. This suggests that the covariance is stronger among high-wage workers. We see in Figure 6 that the correlation increases with severity of top-code, which is driven by the sharp decline in the variance of worker types.

The standard deviation of log wages declines with severity of top-coding. The drop over the depicted range of top-coding is significant for all three standard deviations. The decline is similar for men and women: increasing the share of top-coded observations by 10 percentage

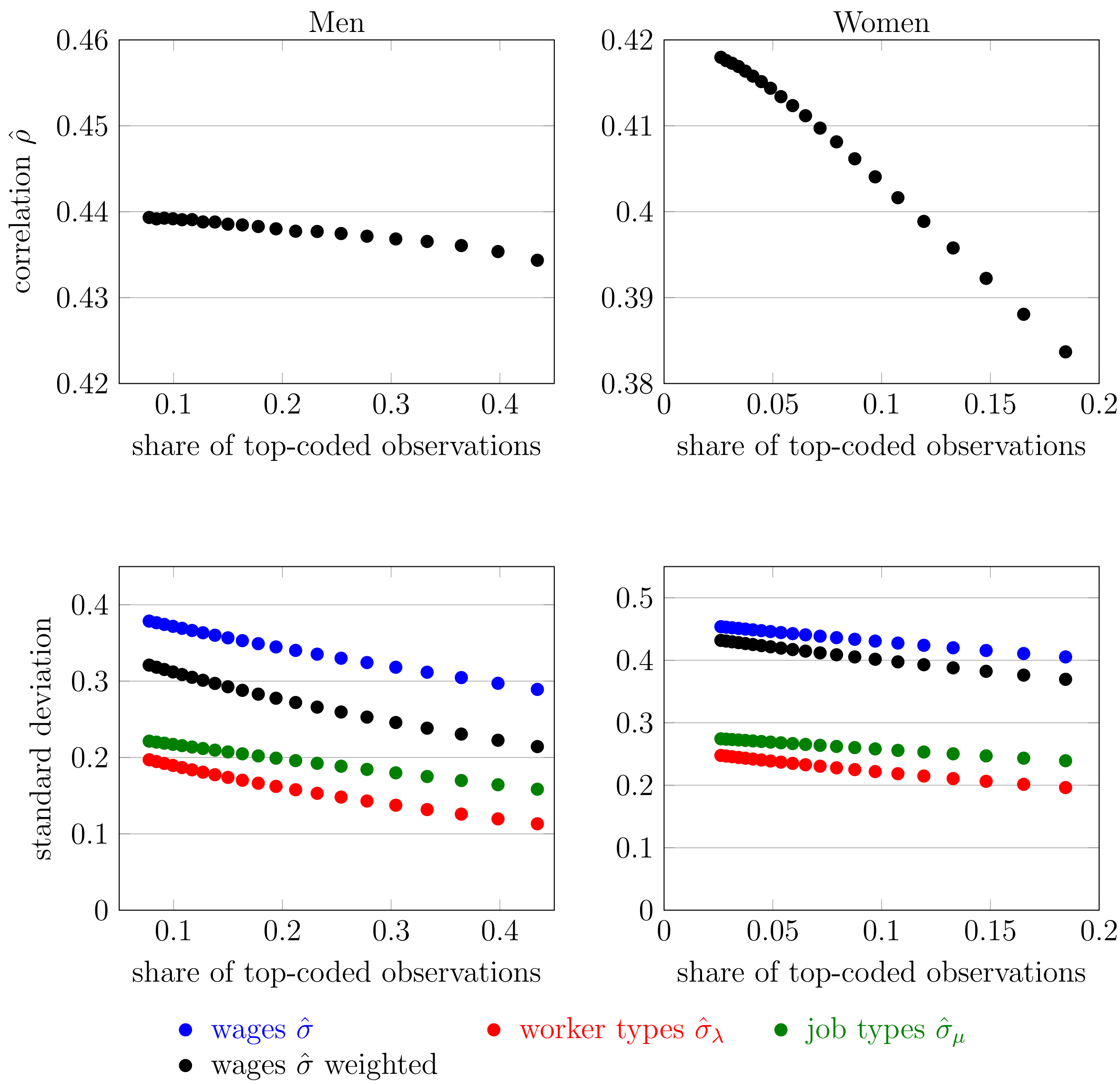

Figure 6: Impact of top-coding on estimated correlation and standard deviation of wages for men and women. Each dot corresponds to a sample where we decreased the top-code by $0, 2, 4, \ldots 40$ percent every year and truncated all wages at this new top-code. The sample of workers and firms is chosen according to Concept 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.

points decreases $\hat{\sigma}$, $\hat{\sigma}_\lambda$, and $\hat{\sigma}_\mu$ by 7.9 percent, 12.1 percent and 8.7 percent, respectively, for men and 6.9 percent, 11.9 percent, 8.0 percent, respectively, for women.

E Time-Varying Types

Consider a variant of the model where both workers' and firms' types change over time, and hence across matches. We are interested in understanding what our estimator would measure in this environment.

Assume that the m^{th} log wage observation for worker i is $\omega^w_{i,m} = \lambda_{i,m} + \varepsilon_{i,m}$. Conditional independence of wage draws implies that $\varepsilon_{i,m}$ is independently distributed with mean 0 and a distribution that may depend on the time-varying type $\lambda_{i,m}$. Similarly, the n^{th} log wage observation for firm j is $\omega^f_{j,n} = \mu_{j,n} + \eta_{j,n}$, where $\eta_{j,n}$ is independently distributed with mean 0 and a distribution that may depend on the time-varying type $\mu_{j,n}$.

Types themselves are autocorrelated. Assume $\lambda_{i,m+1} = r\lambda_{i,m} + \upsilon_{i,m+1}$ and $\mu_{j,n+1} = s\mu_{j,n} + \nu_{j,n+1}$, where $r \in [0,1)$, $s \in [0,1)$ and υ and ν are independent mean zero normal shocks with fixed variances σ^2_υ and σ^2_ν, respectively. The cross-sectional distribution of λ and μ is invariant across matches. Since υ and ν are normal, the stationary distributions of λ and μ are also normal, with zero means and variances $\sigma^2_\lambda = \sigma^2_\upsilon/(1-r^2)$ and $\sigma^2_\mu = \sigma^2_\nu/(1-s^2)$.

Since λ and μ are stationary normal processes, Theorem 1 in Weiss (1975) implies that they are time-reversible. That is, we can write $\lambda_{i,m} = r\lambda_{i,m+1} + \tilde{\upsilon}_{i,m}$ and similarly $\mu_{j,n} = s\mu_{j,n+1} + \tilde{\nu}_{j,n}$ where $\tilde{\upsilon}$ and $\tilde{\nu}$ are independent mean zero normal shocks with variances σ^2_υ and σ^2_ν, respectively. We will use this property to simplify the expression for the estimated covariance.

Finally, assume that there is measure I of workers, and to simplify the algebra, assume that all workers and firms have 2 matches, each of duration 1.

Our estimate of the variance of worker types in this environment is

$$\begin{aligned}
\hat{\sigma}^2_\lambda &= \frac{1}{2I}\int_0^I \left((\omega^w_{i,1} - \bar{w})^2 + (\omega^w_{i,2} - \bar{w})^2 - \left(\omega^w_{i,1} - \omega^w_{i,2}\right)^2\right) di \\
&= \frac{1}{I}\int_0^I \omega^w_{i,1}\omega^w_{i,2} di - \left(\frac{1}{I}\int_0^I \frac{\omega^w_{i,1} + \omega^w_{i,2}}{2} di\right)^2 \\
&= \frac{1}{I}\int_0^I (\lambda_{i,1} + \varepsilon_{i,1})(r\lambda_{i,1} + \upsilon_{i,2} + \varepsilon_{i,2}) di - \left(\frac{1}{I}\int_0^I \frac{(1+r)\lambda_{i,1} + \varepsilon_{i,1} + \upsilon_{i,2} + \varepsilon_{i,2}}{2} di\right)^2 \\
&= r\sigma^2_\lambda.
\end{aligned}$$

The first line uses the assumption that $t^w_{i,m} = 1$ and $M_i = 2$ to derive the Bessel correction factor $\beta_i^w = 2$. It also eliminates $\hat{\lambda}_i$ using its definition $\frac{1}{2}(\omega_{i,1}^w + \omega_{i,2}^w)$. The second line uses the definition of $\bar{w} = \frac{1}{2I}\int_0^I (\omega_{i,1}^w + \omega_{i,2}^w)di$ and expands all the squares. The third line uses the distributional assumptions to express $\omega_{i,m}^w$ in terms of $\lambda_{i,1}$ and shocks. The last line leverages the independence of the shocks to get that the measured variance is biased down by the autocorrelation.

Similarly, we can use the formula in Section 4 to show that $\hat{\sigma}_\mu^2 = s\sigma_\mu^2$.

Finally, our estimate of the covariance is

$$
\begin{aligned}
\hat{c} &= \frac{1}{2I}\int_0^I \Big((\omega_{i,2}^w - \bar{w})(\omega_{k_{i,1},2}^f - \bar{w}) + (\omega_{i,1}^w - \bar{w})(\omega_{k_{i,2},1}^f - \bar{w})\Big)\, di \\
&= \frac{1}{2I}\int_0^1 \big((\lambda_{i,2} + \varepsilon_{i,2})(\mu_{k_{i,1},2} + \eta_{k_{i,1},2}) + (\lambda_{i,1} + \varepsilon_{i,1})(\mu_{k_{i,2},1} + \eta_{k_{i,2},1})\big)di \\
&= \frac{1}{2I}\int_0^1 (r\lambda_{i,1} + \upsilon_{i,2} + \varepsilon_{i,2})(s\mu_{k_{i,1},1} + \nu_{k_{i,1},2} + \eta_{k_{i,1},2})di \\
&\quad + \frac{1}{2I}\int_0^1 (r\lambda_{i,2} + \tilde{\upsilon}_{i,1} + \varepsilon_{i,1})(s\mu_{k_{i,2},2} + \tilde{\nu}_{k_{i,2},1} + \eta_{k_{i,2},1})di \\
&= rs\rho\sigma_\lambda\sigma_\mu.
\end{aligned}
$$

The first line again uses the assumption that $t_{i,m}^w = 1$ and $M_i = N_j = 2$ to simplify the expression. We also order workers and firms so that if firm j is worker i's m^{th} employer, worker i is firm j's m^{th} employee. Since the average wage $\bar{w}$ is zero, we can drop that from subsequent lines. The second line rewrites wages as the sum of time-varying types and i.i.d. shocks. The third line writes the time-varying types in terms of the types in the period when the worker and firm are matched, taking advantage of time-reversibility in the case where the two are matched in the second period but we are looking at wages in the first period. The final line again uses independence of shocks to get that the measured covariance is also biased down.

Combining these results, the estimated correlation would be $\hat{c}/(\hat{\sigma}_\lambda\hat{\sigma}_\mu) = \rho\sqrt{rs} < \rho$. Thus to the extent that types vary over time, our approach underestimates the correlation between types at a point in time.

It may be possible to extend our approach to handle time-varying types. Identification results would build on the ideas in Arellano and Bonhomme (2011), using workers and firms with three or more observations, to distinguish between time-varying types and a low correlation between types in matched pairs.