High Wage Workers Work for High Wage Firms*

Katarína Borovičková
New York University

Robert Shimer
University of Chicago

November 22, 2017

Abstract

We develop a new approach to measuring the correlation between the types of matched workers and firms. Our approach accurately measures the correlation 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 between worker and firm types 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 simple, novel, 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. The AKM fixed effects estimator delivers a correlation 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 conditions which we spell out in the body of the paper, the correlation between our 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 and 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 these data might not contain independent observations of the wage conditional on type. To construct such data, we rely on economic theory. First, since wages are highly autocorrelated within matches, two observations of the same worker in the same job are not independent. We therefore average all our wage data to the worker-firm match level. Second, 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 data on all workers who have at least two jobs and all firms that have at least two employees in our data set. Third, in a more realistic search model with on-the-job search, as in Burdett and Mortensen (1998), 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 empirical 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. 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. This is consistent with the hypothesis of time-varying types.

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 across matches depending on the partners' observable characteristic. For example, we let firms have different types when matched with workers with different skill 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 the incidental parameter problem drives the estimated correlation in the fixed effects literature. 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. Finally, 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.

Bonhomme, Lamadon and Manresa (2016) 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 Bonhomme, Lamadon and Manresa (2016) appears to be biased and their confidence intervals are wider; see their Table 3. On the other hand, Bonhomme, Lamadon and Manresa (2016) 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) identify 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 describes our measure of the correlation between worker and firm types and compares it to the AKM measure of correlation. Section 3 discusses the data that we use in our analysis. Section 4 gives our main empirical results, showing that the correlation between worker and firm types lies between 0.4 and 0.6. Section 5 compares our results with those from the AKM estimator and develops a random graph approach for quantifying the importance of the incidental parameter problem. Section 6 briefly concludes.

2 Measuring Correlation

2.1 Measuring Correlation in Theory

We consider a cross-section of an economy with a measure I of employed workers indexed by i uniform on $[0, I]$ and a measure J of firms indexed by j uniform on $[0, J]$. Workers and firms are distinguished by their characteristics, $y_i \in Y$ and $z_j \in Z$, respectively. Let $F(y)$ denote the distribution of workers' characteristics. Let $G_y(z)$ denote the distribution of the employer's characteristics conditional on the worker's characteristics. We treat F and G as primitives in our environment and view these objects as coming from a snapshot of a dynamic matching model. That is, F is the cross-sectional distribution of employed workers' characteristics and G_y is the cross-sectional conditional distribution of their employers' characteristics. In such a model, differences in G across y 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

$$\Phi(z) \equiv \int_Y G_y(z) dF(y)$$

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.

to be the unconditional distribution of the characteristics of *jobs* in the economy. This is distinct from the distribution of the characteristics of firms to the extent that firms with different characteristics employ different numbers of workers. We also define $\Psi_z(y)$ to be the conditional distribution of the worker's characteristics given the firm's characteristics. Using Bayes rule, we have $G_y(z)F(y) \equiv \Psi_z(y)\Phi(z)$ for all y and z.

We assume that a worker with characteristics y matched to a firm with characteristics z earns a wage that possibly depends on both vectors of characteristics and on a shock. Let $w(y, z, u)$ denote the u^{th} quantile of the cross-sectional log wage distribution in an (y, z) match.[4] In a competitive environment where y captures all productivity-relevant characteristics of a worker, the wage should depend only on y. If there are search (or other) frictions or if we are only able to measure y with noise, the equilibrium wage may be correlated with z and other unobserved characteristics captured by u.

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 y and z 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, G, Φ and Ψ, as well as the wage function w. Of course, this is not true in real world data sets, and so Sections 2.3–2.7 explain 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

$$\lambda(y_i) \equiv \int_Z \int_0^1 w(y_i, z, u)\, du\, dG_{y_i}(z)$$
$$\text{and } \mu(z_j) \equiv \int_Y \int_0^1 w(y, z_j, u)\, du\, d\Psi_{z_j}(y)$$

denote the expected log wage received by worker i with characteristics y_i and the expected log wage paid by firm j with characteristics z_j, respectively. From now on, we identify a

[4]This is the distribution of log wages in the cross-section. If y and z reject some wage draws or turnover is higher following some wage draws, that is reflected in the matching distributions G 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.

worker by her expected log wage and call $\lambda(y_i)$ her type. Symmetrically, we identify a firm by the expected log wage it pays and call $\mu(z_j)$ 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

$$\sigma_\lambda \equiv \sqrt{\int_Y (\lambda(y) - \bar{w})^2 \, dF(y)} \quad \text{and} \quad \sigma_\mu \equiv \sqrt{\int_Z (\mu(z) - \bar{w})^2 \, d\Phi(z)}$$

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

$$c \equiv \int_Y \int_Z (\lambda(y) - \bar{w})(\mu(z) - \bar{w}) \, dG_y(z) \, dF(y)$$

is the covariance between worker and job types in an employment relationship, and

$$\bar{w} \equiv \int_Y \int_Z \int_0^1 w(y, z, u) \, du \, dG_y(z) \, dF(y) = \int_Y \lambda(y) dF(y) = \int_Z \mu(z) d\Phi(z)$$

is the mean log wage, also equal to both the mean worker type and the mean job type. We assume throughout that all of these first and second moments are finite.

We highlight the special case where $G_y(z) = G(z)$ for all y and z. For example, each worker may be equally likely to work in every job, in which case $G(z) = \Phi(z)$. In this case, we can rewrite the covariance as

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

The inner integral 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 type which pushes us towards a positive correlation. Instead, the correlation depends on whether high wage workers are particularly likely to work at high wage firms.

2.2 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' starting point is an assumption that workers and firms have one-dimensional characteristics, so $Y = Z = \mathbb{R}$, the real line, and

that expected log wages are linear in those characteristics,

$$\int_0^1 w(y_i, z_j, u)du = \alpha(y_i) + \psi(z_j) \tag{1}$$

where $\alpha(y_i) \equiv \alpha_i$ is the worker effect and $\psi(z_j) \equiv \psi_j$ is the firm effect.[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). In section 5, we consider the biases in regressing log wages on worker and firm fixed effects using ordinary least squares. Here we ask a different question: Suppose this equation is correctly specified and α_i and ψ_j are known for all workers i and firms j. How is their correlation related to the correlation of $\lambda_i \equiv \lambda(y_i)$ and $\mu_j \equiv \mu(z_j)$?

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, we show here that they are more closely related than appears at first blush. Indeed, in an important special case, the correlation between the two measures is the same.

To show this, assume the AKM wage equation (1) is correctly specified. Also assume that the conditional expected value of ψ_j in a match is linear in α_i, $\int_Z \psi(z)dG_{y_i}(z) = \kappa_0 + \kappa_1\alpha(y_i)$ for all i. Then the definition of λ and the wage equation (1) imply

$$\lambda_i = \int_Z (\alpha(y_i) + \psi(z))dG_{y_i}(z) = \kappa_0 + (1 + \kappa_1)\alpha_i.$$

Symmetrically, assume that the conditional expected value of α_i in a match is linear in ψ_j, $\int_Y \alpha(y)\Psi_{z_j}(y) = \theta_0 + \theta_1\psi(z_j)$ for all j. Then symmetrically

$$\mu_j = \int_Y (\alpha(y) + \psi(z_j))d\Psi_{z_j}(y) = \theta_0 + (1 + \theta_1)\psi_j.$$

The correlation coefficient between two random variables is unaffected by an increasing linear transformation. It follows that linearity of conditional expected values with $\kappa_1 > -1$ and $\theta_1 > -1$ implies that the correlation between α and ψ (the theoretical AKM correlation) is identical to the correlation between λ and μ (our theoretical correlation).

Linearity of conditional expected values is a property of an important family of bivariate distributions which includes the bivariate normal and the bivariate t-distribution as special cases. Let $\xi(\alpha, \psi)$ denote the density function of the joint distribution of matched workers

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

and firms. Let $\bar{\alpha}$ and $\bar{\psi}$ denote the means of α and ψ, respectively, and let σ_α and σ_ψ denote their standard deviations. Finally, let ρ_{AKM} denote their correlation. Then 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 conditional expected values are linear for elliptical distributions and that $\kappa_1 > -1$ and $\theta_1 > -1$ if and only if $\sigma_\alpha + \rho_{AKM}\sigma_\psi$ and $\sigma_\psi + \rho_{AKM}\sigma_\alpha$ are positive.

This leads to our main result comparing our measure of correlation to the correlation between the AKM worker and firm effects:

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

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

The proof in Appendix A establishes linearity of conditional expected values for elliptical distributions and, following the logic in the text, shows

$$\lambda_i = \bar{\psi} - \frac{\rho_{AKM}\sigma_\psi}{\sigma_\alpha}\bar{\alpha} + \left(1 + \frac{\rho_{AKM}\sigma_\psi}{\sigma_\alpha}\right)\alpha_i \quad \text{and} \quad \mu_j = \bar{\alpha} - \frac{\rho_{AKM}\sigma_\alpha}{\sigma_\psi}\bar{\psi} + \left(1 + \frac{\rho_{AKM}\sigma_\alpha}{\sigma_\psi}\right)\psi_j. \tag{2}$$

If $\sigma_\alpha + \rho_{AKM}\sigma_\psi$ and $\sigma_\psi + \rho_{AKM}\sigma_\alpha$ are both positive, λ_i and μ_j are increasing linear transformations of α_i and ψ_j, respectively.[7] If one is positive and one is negative, one of the transformations is decreasing and so our approach flips the sign of the AKM correlation. If one is equal to zero, either λ or μ has a degenerate distribution and so our correlation ρ is undefined.

The Proposition also provides a simple diagnostic tool for detecting this sign flip: if our approach delivers both $\sigma_\lambda - \rho\sigma_\mu$ and $\sigma_\mu - \rho\sigma_\lambda$ positive, $\rho_{AKM} = \rho$. According to our estimates in Section 4, this is the case in Austrian data. If one of $\sigma_\lambda - \rho\sigma_\mu$ and $\sigma_\mu - \rho\sigma_\lambda$ is negative,

[7] $\sigma_\alpha + \rho_{AKM}\sigma_\psi$ and $\sigma_\psi + \rho_{AKM}\sigma_\alpha$ cannot both be negative since $\rho_{AKM} \geq -1$. Similarly, $\sigma_\lambda - \rho\sigma_\mu$ and $\sigma_\mu - \rho\sigma_\lambda$ cannot both be negative since $\rho \leq 1$.

then $\rho_{AKM} = -\rho$. Finally, in the borderline case where one is zero, our approach cannot measure the correlation.

Proposition 1 implies that our framework is consistent with any correlation pattern, even one where high wage workers typically work in low wage firms. In particular, if $\sigma_\alpha = \sigma_\psi$ and $\rho_{AKM} > -1$, then our approach gives $\rho = \rho_{AKM}$. If $\sigma_\alpha \neq \sigma_\psi$ and $\rho_{AKM} < 0$, it is possible that our approach delivers the opposite correlation, $\rho = -\rho_{AKM}$, but our diagnostic tool identifies this situation.

We view the restriction that the joint distribution of α and ψ is elliptical as a reasonable starting assumption, but it might be violated in reality. Still, we believe the link between ρ and ρ_{AKM} is likely to be robust. For one thing, there are other distributions with linear conditional expectations. For example, let $\xi(\alpha, \psi) = \frac{1}{4} + \frac{3}{4}\rho_{AKM}\alpha\psi$ with support $[-1, 1]^2$. This is a proper density function if $|\rho_{AKM}| \leq \frac{1}{3}$, in which case one can verify that the conditional expectation of α given ψ is $\rho_{AKM}\psi$ and symmetrically the conditional expectation of ψ given α is $\rho_{AKM}\alpha$. It follows that the results in Proposition 1 go through even though the probability distribution is not elliptical. But more generally, conditional expectations may be nonlinear, in which case the correlation between α and ψ will generally differ from the correlation between λ and μ. Still, in this case we do not see an obvious reason to prefer the AKM measure of correlation to ours.

We do see one important advantage to our measure of types: it does not impose the structure of equation (1), a log-linear wage equation. Many models predict that a worker's wage is a nonmonotone function of the firm's type (Eeckhout and Kircher, 2011; Lopes de Melo, forthcoming; Bagger and Lentz, 2016). Although one can still estimate AKM fixed effects in data sets generated by models that do not have a log-linear wage equation, the fixed effects cannot be interpreted as structural parameters (see, for example Abowd and Kramarz, 1999, Section 4). Our approach allows for nonlinearities and non-monotonicities in the wage equation and so is equally well-suited to these more general environments.

2.3 Measuring Correlation in Practice

We return now to the cross-sectional correlation between λ and μ. If we observed many conditionally independent wage draws for each worker and firm, we could accurately measure $\lambda(y)$ and $\mu(z)$ for everyone and hence directly measure their correlation. Unfortunately, in practice we have very few observations for most workers and most firms. The remainder of this section proposes a strategy for measuring the correlation between λ and μ in realistic data sets. We start in this subsection by setting up the notation and clarifying our identifying assumptions.

We imagine a data set that includes worker identifiers, firm identifiers, wages, and the duration of employment relationships. This information is commonly available from administrative records in many countries. Let M_i denote the number of observations for worker i and N_j denote the number of observations for firm j. We label the log wage observations of worker i as $\omega^w_{i,1}, \ldots, \omega^w_{i,M_i}$ and the log wage observations of firm j as $\omega^f_{j,1}, \ldots, \omega^f_{j,N_j}$. Similarly let $t^w_{i,1}, \ldots, t^w_{i,M_i}$ denote the duration of worker i's jobs; and let $t^f_{j,1}, \ldots, t^f_{j,N_j}$ denote the duration of firm j's hires. Of course these observations are linked. Let $h_{j,n} \in [0, I]$ denote the worker employed by firm j in its n^{th} observation and let $k_{i,m} \in [0, J]$ denote the firm that employs worker i in her m^{th} job. Then for all i and m, $\omega^w_{i,m} = \omega^f_{j,n}$ and $t^w_{i,m} = t^f_{j,n}$ if $j = k_{i,m}$ and $i = h_{j,n}$.

We assume that worker i's expected wage at any moment when we observe her employed is some unknown constant λ_i, a weighted average of wage draws from some unspecified worker-specific distribution, with weights equal to the duration of the job. We also assume that each log wage observation and duration pair $(\omega^w_{i,m}, t^w_{i,m})$ is drawn independently from this worker-specific distribution.

Symmetrically, we assume that the expected wage that firm j pays each employee at any moment in time is some unknown μ_j, a weighted average of wage draws from some firm-specific distribution with weights equal to the duration of the job. We also assume that each log wage observation and duration pair $(\omega^f_{j,n}, t^f_{j,n})$ is drawn independently from this firm-specific distribution.

The only additional restriction we impose is that the total variance of wages weighted by duration is finite. Since the variance of wages is the sum of the variance of worker types λ_i plus the mean weighted variance of wage draws for each worker, we require that each of these objects is finite as well.

Our strongest assumption is independence of wage observations. It is satisfied for any two employment relationships in search models where workers may only search while unemployed, such as Shimer and Smith (2000). In this case, duration is an exponentially distributed random variable that is uncorrelated with the wage. In models with on-the-job search, such as Burdett and Mortensen (1998), the independence assumption is satisfied for workers as long as the two employment relationships are separated by an unemployment spell; and it is always satisfied for firms. In this case, high wages jobs typically last for longer than low wage jobs. The independence assumption is also consistent with certain specifications of measurement error in log wages. Our approach requires that the mean measurement error in log wages is the same for all workers but allows for arbitrary heteroskedasticity. We discuss later how we use these models to guide our measurement.

We only observe worker i for a finite amount of time, and so have just a small snapshot

of her potential wage draws. We assume that the jobs that we do observe during this time period are independent draws from the worker-specific wage-duration distribution, although recognize that the duration of the first and last observations may be censored. We include the worker in our analysis only if she has at least two wage observations, $M_i \geq 2$. Symmetrically, we only observe a finite number of workers at each firm. We make a symmetric independence assumption for firms and only include a firm if it has at least two wage observations, $N_j \geq 2$.

We stress that our results only apply to the sample of workers and firms, each of which has at least two wage observations. We cannot say much about how similar these workers are to other workers with only one observation, nor how similar these firms are to other firms with only one employee. Our intuition says that workers who keep a single job throughout their lifetime are probably better matched to their job, and hence the correlation between the worker's and firm's types is higher, than for the average worker.[8]

A naïve approach would be to measure the correlation between each matched worker's and firm's mean wage, but this gives a biased measure of both the covariance between matched pairs and the variance of types. The variance of mean wages is an upward biased measure of the variance of types because the mean wage in a finite sample is a noisy measure of type. The covariance between mean wages in matched pairs is an upward biased measure of the covariance between matched types because it includes the common wage observation for the pair. There is no reason to expect these two biases to cancel out.

Our approach deals with both of these biases. In contrast to the naïve approach, we do not attempt to measure any particular worker's or firm's type, but instead measure the variance of each and the covariance between them. We first perform unbiased within-between decompositions of the variance of log wages weighted by the duration of spells. We show that σ_λ^2 and σ_μ^2 correspond to the between-worker and between-firm variances. We then obtain the covariance c by noting that for a matched worker-firm pair, their other wages covary only because types covary in matches. For any particular matched pair, this yields a noisy measure of c, and hence we take their average to recover the desired moments. The next three subsections explains these measures in detail.

2.4 Measuring the Standard Deviation of Worker Types σ_λ

We start by obtaining an unbiased estimator of worker i's type, $\lambda_i = \lambda(y_i)$. The estimator is simply the weighted average of the M_i wages we actually observe,

$$\hat{\lambda}_i = \frac{\sum_{m=1}^{M_i} t_{i,m}^w \omega_{i,m}^w}{T_i^w}$$

[8] Appendix B offers some evidence consistent with this hypothesis.

where $T_i^w \equiv \sum_{m=1}^{M_i} t_{i,m}^w$. By assumption, the worker's expected wage is λ_i at each instant that we observe her. Weighting observations by duration captures the fact that we observe long duration jobs at more instants in time. Of course, $\hat{\lambda}_i$ is a noisy measure of λ_i unless i has a degenerate wage distribution. For this reason, we do not measure the variance of worker types directly from the variance of $\hat{\lambda}_i$.

We turn next to a measure of $\bar{w}$, the mean log wage in the economy at a point in time:

$$\bar{w} = \frac{\int_0^I \sum_{m=1}^{M_i} t_{i,m}^w \omega_{i,m}^w di}{\int_0^I T_i^w di} = \frac{\int_0^I T_i^w \hat{\lambda}_i di}{\int_0^I T_i^w di}. \tag{3}$$

This is a weighted average of the estimators of the workers' types, where the weights reflect the amount of time that the worker is in the data set, i.e. the likelihood of finding the worker in a particular cross-section. Because individual workers' observations are independent and we have a continuum of workers, we appeal to a law of large numbers and treat $\bar{w}$ as deterministic.

Next, we seek to measure the cross-sectional variance of log wages,

$$\sigma^2 \equiv \int_Y \int_Z \int_0^1 (w(y,z,u) - \bar{w})^2 \, du \, dG_y(z) \, dF(y).$$

Using our data set, this is simply

$$\sigma^2 = \frac{\int_0^I \sum_{m=1}^{M_i} t_{i,m}^w (\omega_{i,m}^w - \bar{w})^2 di}{\int_0^I T_i^w di}, \tag{4}$$

the empirical cross-section of log wages, weighting each observation by its duration. Again, this is deterministic in a large data set.

We next break the cross-sectional variance of log wages into the within and between components, or equivalently into the mean of individual variances and the variance of individual means. We start with the within-worker variance of log wages, defined in theory as

$$\sigma_{ww}^2 \equiv \int_Y (\sigma_i^w)^2 \, dF(y)$$

where

$$(\sigma_i^w)^2 \equiv \int_Z \int_0^1 (w(y_i,z,u) - \lambda(y_i))^2 \, du \, dG_{y_i}(z)$$

is the variance of worker i's log wage.

To measure the within-worker variance of log wages, we need an unbiased measure of $(\sigma_i^w)^2$. The following Lemma gives one such measure:

Lemma 1 *An unbiased measure of* $(\sigma_i^w)^2$ *is*

$$(\hat{\sigma}_i^w)^2 \equiv \frac{\beta_i^w \sum_{m=1}^{M_i} t_{i,m}^w \left(\omega_{i,m}^w - \hat{\lambda}_i\right)^2}{T_i^w} \tag{5}$$

where

$$\beta_i^w \equiv \frac{(T_i^w)^2}{(T_i^w)^2 - \sum_{m=1}^{M_i}(t_{i,m}^w)^2}$$

is the Bessel correction factor.

The proof is in Appendix A. The Bessel correction factor β_i^w accounts for the fact that $\hat{\lambda}_i$ is a noisy measure of λ_i in finite samples. Jensen's inequality implies $\beta_i^w \geq M_i/(M_i - 1)$ with equality if and only if $t_{i,m}^w = T_i^w/M_i$ for all m. That is, if all spells have the same duration, we get the standard Bessel correction, but otherwise the correction factor is larger, boosting the estimator of the worker's variance.

Aggregating these unbiased estimators is easy. The within-worker variance is a weighted average of $(\hat{\sigma}_i^w)$, where the weight again corresponds to the total duration of the spells:

$$\sigma_{ww}^2 = \frac{\int_0^I \beta_i^w \sum_{m=1}^{M_i} t_{i,m}^w \left(\omega_{i,m}^w - \hat{\lambda}_i\right)^2 di}{\int_0^I T_i^w di}. \tag{6}$$

Next, observe directly from their definitions that the variance of worker types satisfies $\sigma_\lambda^2 = \sigma^2 - \sigma_{ww}^2$. Since we have measures of both terms on the right hand side, we also have a measure of the standard deviation of worker types:

Lemma 2

$$\sigma_\lambda = \sqrt{\frac{\int_0^I \sum_{m=1}^{M_i} t_{i,m}^w (\omega_{i,m}^w - \bar{w})^2 di - \int_0^I \beta_i^w \sum_{m=1}^{M_i} t_{i,m}^w \left(\omega_{i,m}^w - \hat{\lambda}_i\right)^2 di}{\int_0^I T_i^w di}} \tag{7}$$

measures the standard deviation of worker types.

2.5 Measuring the Standard Deviation of Firm Types σ_μ

Our approach to measuring σ_μ, the standard deviation of firm types, is similar. The mean wage can also be computed by averaging across firms,

$$\bar{w} = \frac{\int_0^J \left(\sum_{n=1}^{N_j} t_{j,n}^f \omega_{j,n}^f\right) dj}{\int_0^J T_j^f dj} = \frac{\int_0^J T_j^f \hat{\mu}_j dj}{\int_0^J T_j^f dj}$$

where

$$\hat{\mu}_j = \frac{\sum_{n=1}^{N_i} t_{j,n}^f \omega_{j,n}^f}{T_j^f}$$

and $T_j^f \equiv \sum_{n=1}^{N_j} t_{j,n}^f$. That this is identical to the definition of the mean wage in equation (3) comes from the fact that the m^{th} observation for worker i can be mapped into an observation for its employer $j = k_{i,m}$ and vice versa. Similarly we can measure the variance of log wages across jobs as

$$\sigma^2 = \frac{\int_0^J \sum_{n=1}^{N_j} t_{j,n}^f (\omega_{j,n}^f - \bar{w})^2 dj}{\int_0^J T_j^f dj}.$$

Again, this is mathematically identical to the variance of log wages across workers in equation (4).

We turn next to measuring the mean of the variance of log wages across jobs. An unbiased measure of the variance of firm j's log wage is

$$(\hat{\sigma}_j^f)^2 \equiv \frac{\beta_j^f \sum_{n=1}^{N_j} t_{j,n}^f \left(\omega_{j,n}^f - \hat{\mu}_j\right)^2}{T_j^f} \tag{8}$$

where

$$\beta_j^f \equiv \frac{(T_j^f)^2}{(T_j^f)^2 - \sum_{n=1}^{N_j} (t_{j,n}^f)^2}$$

is the Bessel correction factor. The logic is identical to the variance of a worker's log wage and so we omit it.

Finally, a weighted average of these variances gives us the within-firm variance:

$$\sigma_{wf}^2 = \frac{\int_0^J \beta_j^f \sum_{n=1}^{N_j} t_{j,n}^f \left(\omega_{j,n}^f - \hat{\mu}_j\right)^2 dj}{\int_0^J T_j^f dj}. \tag{9}$$

Since $\sigma_\mu^2 = \sigma^2 - \sigma_{wf}^2$, we have a measure of the between variance of jobs.

Lemma 3

$$\sigma_\mu = \sqrt{\frac{\int_0^J \sum_{n=1}^{N_j} t_{j,n}^f (\omega_{j,n}^f - \bar{w})^2 dj - \int_0^J \beta_j^f \sum_{n=1}^{N_j} t_{j,n}^f \left(\omega_{j,n}^f - \hat{\mu}_j\right)^2 dj}{\int_0^J T_j^f dj}} \tag{10}$$

measures the standard deviation of firm types.

2.6 Measuring the Correlation of Matched Types ρ

The third step is to find the covariance c between λ and μ in matched worker-firm pairs. A naïve approach would be to directly measure the covariance between $\hat{\lambda}_i$ and $\hat{\mu}_{k_{i,m}}$ for every worker $i \in [0, I]$ and match $m \in \{1, \ldots, M_i\}$. This is biased by the common wage observation in the match between i and $k_{i,m}$. Instead, take any worker i and employer $k_{i,m}$. Suppose that worker i is firm $k_{i,m}$'s ${e_{i,m}}^{th}$ employee, i.e. $e_{i,m} \in \{1, \ldots, N_{k_{i,m}}\}$ and $h_{k_{i,m},e_{i,m}} = i$. The average log wage that i receives in her other jobs is $\lambda(y_i)$ plus noise. The average log wage that firm $k_{i,m}$ pays to its other employees is $\mu(z_{k_{i,m}})$ plus noise. Moreover, the two sources of noise are independent. Therefore the product of the average log wage that a worker receives in her other jobs and the average log wage that a firm pays to its other employees,

$$\left(\frac{\sum_{m' \neq m} t^w_{i,m'} \omega^w_{i,m'}}{\sum_{m' \neq m} t^w_{i,m'}}\right)\left(\frac{\sum_{n' \neq e_{i,m}} t^f_{k_{i,m},n'} \omega^f_{k_{i,m},n'}}{\sum_{n' \neq e_{i,m}} t^f_{k_{i,m},n'}}\right)$$

is a random variable with expected value

$$\int_Y \int_Z \lambda(y)\mu(z) dG_y(z) dF(y).$$

This is a standard "leave-one-out" estimator. Subtracting off unconditional means and averaging across workers and employers for each worker leads to our measure of the covariance $c = \int_Y \int_Z (\lambda(y) - \bar{w})(\mu(z) - \bar{w}) dG_y(z) dF(y)$:

Lemma 4

$$c = \frac{\int_0^I \sum_{m=1}^{M_i} t^w_{i,m} \left(\frac{\sum_{m' \neq m} t^w_{i,m'} \omega^w_{i,m'}}{\sum_{m' \neq m} t^w_{i,m'}} - \bar{w}\right)\left(\frac{\sum_{n' \neq e_{i,m}} t^f_{k_{i,m},n'} \omega^f_{k_{i,m},n'}}{\sum_{n' \neq e_{i,m}} t^f_{k_{i,m},n'}} - \bar{w}\right) di}{\int_0^I T^w_i di} \tag{11}$$

measures the covariance between a worker's type and the type of her employer.

Our main theoretical result follows immediately from these three Lemmas:

Proposition 2 *The correlation between a worker's type and the type of her employer can be measured using a data set with worker identifiers, firm identifiers, wages, and the duration of employment relationships, in which all workers and firms have at least two conditionally independent wage and duration observations.*

2.7 Estimators

Our actual estimators recognize that real world data sets are finite. With some abuse of notation, we let I denote the number of workers, now indexed by $i \in \{1, \ldots, I\}$, and J denote the number of firms, now indexed by $j \in \{1, \ldots, J\}$. Our estimator for the correlation between worker and firm types is the obvious finite analog of the previous measures:

$$\hat{\rho} \equiv \frac{\hat{c}}{\hat{\sigma}_\lambda \hat{\sigma}_\mu}$$

where

$$\hat{\sigma}_\lambda \equiv \sqrt{\frac{\sum_{i=1}^{I} \sum_{m=1}^{M_i} t_{i,m}^w (\omega_{i,m}^w - \hat{\bar{w}})^2 - \sum_{i=1}^{I} \beta_i^w \sum_{m=1}^{M_i} t_{i,m}^w \left(\omega_{i,m}^w - \hat{\lambda}_i\right)^2}{\sum_{i=1}^{I} T_i^w}},$$

$$\hat{\sigma}_\mu \equiv \sqrt{\frac{\sum_{j=1}^{J} \sum_{n=1}^{N_j} t_{j,n}^f (\omega_{j,n}^f - \hat{\bar{w}})^2 - \sum_{j=1}^{J} \beta_j^f \sum_{n=1}^{N_j} t_{j,n}^f \left(\omega_{j,n}^f - \hat{\mu}_j\right)^2}{\sum_{j=1}^{J} T_j^f}},$$

$$\hat{c} \equiv \frac{\sum_{i=1}^{I} \sum_{m=1}^{M_i} t_{i,m}^w \left(\frac{\sum_{m' \neq m} t_{i,m'}^w \omega_{i,m'}^w}{\sum_{m' \neq m} t_{i,m'}^w} - \hat{\bar{w}}\right) \left(\frac{\sum_{n' \neq e_{i,m}} t_{k_{i,m},n'}^f \omega_{k_{i,m},n'}^f}{\sum_{n' \neq e_{i,m}} t_{k_{i,m},n'}^f} - \hat{\bar{w}}\right)}{\sum_{i=1}^{I} T_i^w},$$

$$\hat{\bar{w}} \equiv \frac{\sum_{i=1}^{I} \sum_{m=1}^{M_i} t_{i,m}^w \omega_{i,m}^w}{\sum_{i=1}^{I} T_i^w},$$

with worker means $\hat{\lambda}_i$, firm means $\hat{\mu}_j$, and Bessel correction factors β_i^w and β_j^f defined in the text. Each of these objects is readily measured using a data set that contains worker and firm identifiers as well as wages and job durations.

Although we conjecture that $\hat{\rho}$ is a consistent estimator of ρ in a standard two-sided matching model, a proof goes beyond the scope of this paper. The basic difficulty is that individual observations are not independent in a finite agent matching model. For example, if a worker works for a particular firm, it is less likely that she works for any other firm. Azevedo and Leshno (2016) prove convergence in a simpler model of college-student matching, where the number of students goes to infinity and each student has only one match. Menzel (2015) examines similar issues in a marriage market where the number of men and women both go to infinity, but everyone can have only one match. We have large numbers and multiple matches on both sides of the market, further complicating a proof. Rather than trying to prove convergence analytically, we rely on simulations of model-generated data. In particular, in Section 4.2, we use a parametric bootstrap to compute confidence intervals. This approach informs us about the behavior of our estimator in samples with realistic properties: many

workers and firms but few conditionally independent observations for most of them. We find that the confidence intervals are small and centered around the true correlation.

3 Data

3.1 Data Description

We measure the correlation between workers and jobs using panel data from the Austrian social security registry (Zweimuller, Winter-Ebmer, Lalive, Kuhn, Wuellrich, Ruf and Buchi, 2009). The data set covers the universe of workers in the private sector from 1972 to 2007. For each worker, it contains information about every job they hold. More precisely, in every calendar year and for every worker-firm pair,[9] we observe earnings and days worked during the year.[10] 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 look only on full-time jobs and drop any data that includes an apprenticeship.

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.

3.2 Independence Assumptions

For our method to provide an accurate estimate of the correlation ρ, we need each wage observation to be independent conditional on the worker identifier and conditional on the

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

[10]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 4.3.

firm identifier. We approach this in several ways, always motivated by economic theories such as Burdett and Mortensen (1998) 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 residual 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.[11] We then select all workers who are employed by at least two employers and measure the correlation. We call this independence assumption II: wages are independent across matches.

We recognize that, due to job-to-job movements, residual wages might be correlated across employment relationships. To understand the problem, consider the job ladder model from Burdett and Mortensen (1998). There, an employed worker accepts a job offer from another firm if and only if it pays a higher wage. This means that the wage in jobs held before and after the job-to-job transition are correlated. According to this model, an unemployment spell breaks this correlation and so wages in two employment relationships separated by an unemployment spell are independent. Guided by these insights, we select all workers with at least two employment spells separated by a spell of registered unemployment and take the longest job during each employment spell.[12] 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,

[11] 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.

[12] 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.

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 nonempty.

4 Results

4.1 Main Results

Tables 1 and 2 show the main results for men and women, respectively. 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.

We start in column (1) by measuring the naïve correlation between $\hat{\lambda}_i$ and $\hat{\mu}_j$. For each worker and firm, including those with only a single observation, we compute the mean residual log wage that a worker earns and that a firm pays, weighting each observation by its duration. We estimate that the correlation is 0.598 for men and 0.578 for women. Although we have already argued that the naïve measure is biased, it is interesting to see that it is not wildly different than the other numbers we report in Tables 1 and 2.

Column (2) of Tables 1 and 2 uses independence assumption I to construct the correlation with our approach. This treats any two firm-year observations for a given worker as independent. For men, we see the (modest) bias in the naïve calculation: the covariance and variance of types falls slightly going from column (1) to column (2). In net, the correlation increases slightly for both men and women.

Column (3) uses the more plausible independence assumption II to construct the correlation, aggregating wage observations to the level of the worker-firm match. For men, each component of the correlation drops sharply, but the correlation barely changes. For women, we see a sharper drop in the correlation, driven by a larger decline in the covariance. Interpreting this drop is not trivial. On the one hand, we expect that independence assumption I is incorrect and so the resulting correlation in column (2) is biased. On the other hand,

Estimated Correlation and Variances: Men

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
correlation of matched types $\hat{\rho}$	0.598	0.642	0.617	0.491	0.450	0.435	0.418
covariance of matched types $\hat{c}$	0.051	0.046	0.026	0.018	0.017	0.020	0.017
variance of log wages $\hat{\sigma}^2$	0.155	0.130	0.113	0.103	0.102	0.123	0.115
variance of worker types $\hat{\sigma}^2_\lambda$	0.093	0.078	0.033	0.028	0.030	0.040	0.035
variance of job types $\hat{\sigma}^2_\mu$	0.077	0.066	0.054	0.049	0.049	0.053	0.049
number of workers (thousands)	4,171	3,672	2,811	1,101	676	650	652
number of firms (thousands)	782	672	499	234	206	179	180
number of observations (thousands)	63,798	63,198	16,131	4,376	3,505	2,810	2,815
share of observations top-coded	0.185	0.186	0.134	0.078	0.060	0.033	0.041
independence assumption	naïve	I	II	III	IV	IV	IV
observations included	all	all	all	longest	longest	first	last
first year of sample	1972	1972	1972	1986	1986	1986	1986

Table 1: 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 (3)–(7) aggregate residual wages to the worker-firm match level by taking a weighted average of wages within the match across years. We use a naïve measure of correlation in column (1), and our method in columns (2)–(7). 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.

Estimated Correlation and Variances: Women

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
correlation of matched types $\hat{\rho}$	0.578	0.617	0.435	0.429	0.428	0.448	0.436
covariance of matched types $\hat{c}$	0.088	0.093	0.040	0.028	0.027	0.031	0.028
variance of log wages $\hat{\sigma}^2$	0.272	0.270	0.237	0.187	0.174	0.189	0.177
variance of worker types $\hat{\sigma}^2_\lambda$	0.174	0.170	0.077	0.056	0.056	0.065	0.058
variance of job types $\hat{\sigma}^2_\mu$	0.132	0.133	0.110	0.077	0.072	0.074	0.070
number of workers (thousands)	3,439	3,128	2,359	951	540	503	504
number of firms (thousands)	878	760	522	238	196	160	162
number of observations (thousands)	47,054	46,635	11,103	3,190	2,336	1,771	1,773
share of observations top-coded	0.049	0.050	0.043	0.026	0.020	0.012	0.013
independence assumption	naïve	I	II	III	IV	IV	IV
observations included	all	all	all	longest	longest	first	last
first year of sample	1972	1972	1972	1986	1986	1986	1986

Table 2: Estimates of correlations, covariances, and variances between matched workers' and firms' types for women. See description of Table 1 for details.

we lose a substantial number of workers going from column (2) to column (3) and so worry that the drop reflects the changing sample. We argue in Appendix B that sample selection is probably not very important here and so prefer the estimates in column (3) over those 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 (4) shows a drop in the estimated correlation for men, with little additional change for women. We argue in Appendix B that the drop for men reflects a combination of selection and bias, both working to reduce the correlation in column (4).

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 (5), (6), and (7) look at the longest, first, and last job during multiple employment spells. From the perspective of Burdett and Mortensen (1998), the results in column (6) can be understood as measuring the correlation in the sampling distribution of wages, while those in column (7) should reflect the steady state distribution. For men, these estimates slightly reduce the measured correlation compared to column (4), while for women the results are scarcely changed.

In summary, the estimated correlation between types ranges from 0.42 to 0.64 for men, and from 0.43 to 0.62 for women. The exact number depends on the independence assumption. As we move from the naïve measure to independence assumption 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, leading to sample selection problems. We conclude in Appendix B that both bias and selection matter and choose to focus on the results in column (4) because we believe those are likely to satisfy the independence assumption while minimizing the sample selection issues in the last three columns. We view this choice as conservative, in the sense that sample selection probably still biases the measured correlation down.

Column (4) shows that the standard deviation of worker types is 0.17 for men. The associated standard deviation of firm types is somewhat higher, 0.22. 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, 0.24 for workers and 0.28 for firms, but the conclusion is the same. This result holds in every specification in Tables 1 and 2.

Finally, we compare our results to those of Bonhomme, Lamadon and Manresa (2016), who propose a new way to estimate the correlation between AKM fixed effects. Using Swedish administrative data, they recover a correlation of 0.42–0.49, depending on the model. This is remarkably similar to our estimate in column (4). Our results differ in the relative variance of worker and firm types. Bonhomme, Lamadon and Manresa (2016) find that the variance of the firm fixed effect is only four to seven percent of the variance of the worker fixed effects. In contrast, we find that the variance of firm types is 75 percent larger than the variance of worker types. There are three explanations for this difference: definitions of types, estimation method, and the data used. We believe that the definition of types in not the key reason. Using Proposition 1, we can translate estimated variances of worker and firm types into variances of the worker and firm fixed effects. This transformation shows that the variance of the firm fixed effect also exceeds the variance of the worker fixed effects. To see whether the difference reflects something about Sweden versus Austria, one needs to check the results from either applying our method to their Swedish data or their method to our Austrian data. Doing so goes beyond the scope of this paper.

4.2 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 1 and 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 1, column (4), the estimated correlation for men is $\hat{\rho} = 0.4912$, and the 95 percent confidence interval is $[0.4886, 0.4935]$. In Table 2, column (4), the estimated correlation for women 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.[13] 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.

4.3 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 as-

[13] 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.

Robustness Results for Men and Women

	men		women	
	(1)	(2)	(3)	(4)
correlation of matched types $\hat{\rho}$	0.493	0.514	0.451	0.430
covariance of matched types $\hat{c}$	0.015	0.016	0.026	0.027
variance of log wages $\hat{\sigma}^2$	0.092	0.071	0.170	0.181
variance of worker types $\hat{\sigma}^2_\lambda$	0.021	0.022	0.050	0.052
variance of job types $\hat{\sigma}^2_\mu$	0.041	0.045	0.069	0.106
number of workers (thousands)	1,101	1,101	951	951
number of firms (thousands)	234	234	238	238
number of observations (thousands)	4,376	4,376	3,190	3,190
share of observations top-coded	0.078	0.117	0.026	0.026
independence assumption	III	III	III	III
quartic in experience	yes	no	yes	no
more severe top-code	no	yes	no	yes

Table 3: 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 columns (1) and (3), we regress log wages on year, age, and a polynomial for work experience. Columns (2) and (4) only regress log wages on year and age, but first reduce the top code by ten percent in each year. All columns use independence assumption III, treating the longest jobs during each employment spell as independent observations. The sample always runs from 1986–2007.

sumption 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.[14] We then include a quartic polynomial in experience in addition to age and year dummies when we calculate the residual log wages. Table 3, column (1), which we hereafter refer to as Table 3(1), and Table 3(3) show the results for men and women, respectively. These are little changed from the corresponding results in Tables 1(4) and 2(4).

We next study the role of top-coding. In our baseline results in Tables 1(4) and 2(4), top-coding affects 7.8 percent of men's observations and 2.6 percent of women.[15] We ask here what would have happened if the top-coding threshold had been ten percent lower in every year.[16] This would have increased the share of top-coded observations to 11.7 percent

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

[15]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.

[16]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.

for men and 4.1 percent for women.

Table 3(2) and 3(4) shows 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 shows that for men, more severe top-code sharply decreases the variance of worker types which leads to a substantial increase in the correlation. For women, the estimated correlation is robust to the level of the top-code. We interpret this as suggesting that, if we had data without top-coding, the correlation would be slightly lower for men and little changed for women.

4.4 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.[17]

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 1 shows that the correlation between worker and firm types increased slightly for men, from an initial 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 Tables 1(4) and 2(4) using 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, 0.44 for men and 0.43 for women.[18] Sample differences are unimportant for women and actually

[17] Appendix F 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 patterns are broadly similar to those we report in this section.

[18] 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.

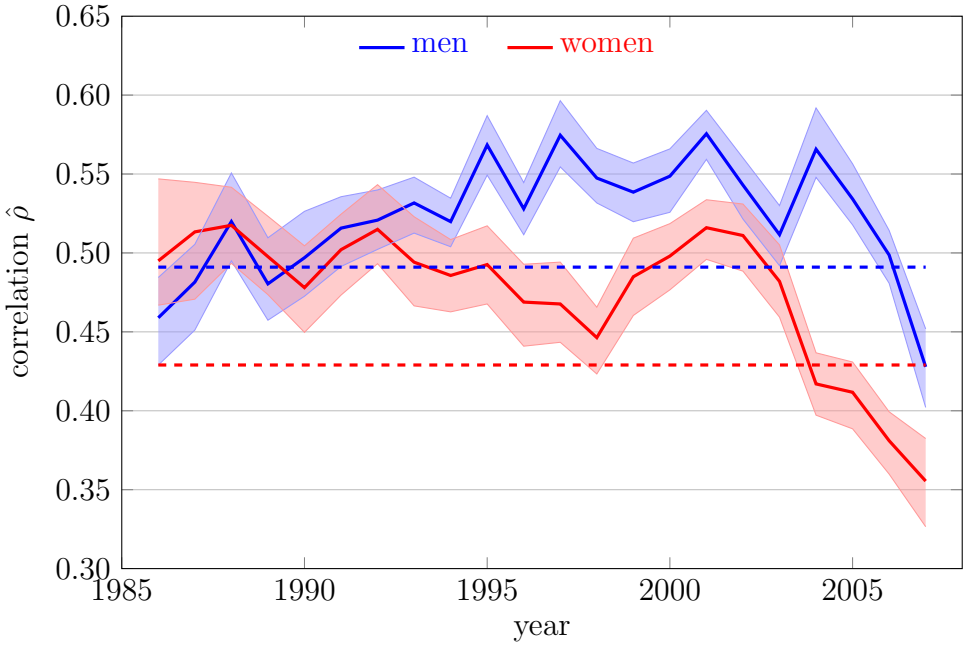


Figure 1: Correlation between worker and firm types using residual log wages under independence assumption III. Solid lines are computed year-by-year and shaded areas are bootstrapped 95 percent confidence intervals. For each year, the sample considers all workers who switched employers after an unemployment 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 Tables 1(4) and 2(4).

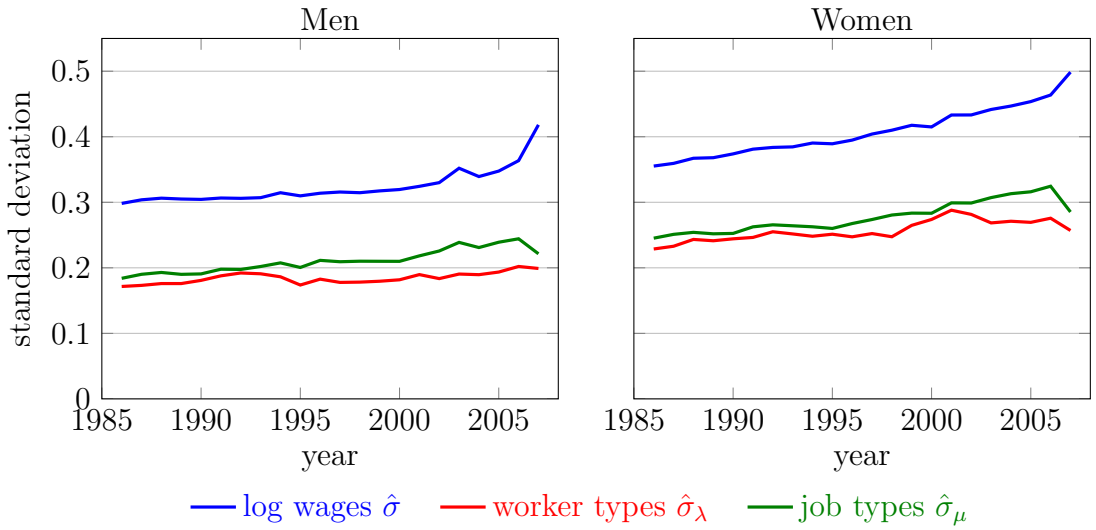


Figure 2: Standard deviations of log wages, worker types, and job types, using residual log wages under independence assumption III. Each line is computed year-by-year and uses one job per employment spell. See the description of Figure 1 for more details.

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.

Finally, Figure 2 shows the estimated standard deviation of residual log wages as well as the standard deviation of worker and job types for both men and women, using one job per employment spell. For both men and women, we find that the standard deviation of job types is slightly larger than the standard deviation of worker types in every year. This contrasts with the pooled data in Tables 1(4) and 2(4), which show a bigger gap between the two standard deviations. Again, the higher standard deviation of worker types here is consistent with time-varying types. Additionally, all four standard deviations show a modest increase over the sample period, until the last year of the sample.

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,

The sample contains 624,917 men and 408,614 women.

which that 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 Tables 1(5) and 1(6) for men and Tables 2(5) and 2(6) for women. Second, we have regressed log wages on the job's order within a spell, in addition to age and year dummies, before constructing wage residuals. This additional control has no quantitative impact on the correlations in Figure 1.

4.5 Life Cycle

We can also use our approach to create synthetic cohorts to examine how sorting evolves over the life cycle. We redo our analysis, now using only workers at a particular age a. We again use independence assumption III when constructing the data set.[19] We report results for workers aged 20–54. The sample size falls considerably at older ages, which makes the results noisy and, according to our bootstrap procedure, unreliable.[20]

Figure 3 shows the estimated life cycle pattern of the correlation between the types of matched workers and firms. For men, we find a remarkably steady increase, more than doubling from 0.40 at age 20 to 0.89 at age 54. The pattern for women is somewhat different: a steep rise from age 23 to 31, followed by a dip for the next decade, and then a gradual increase that accelerates after age 48, although the widening of the confidence interval makes this last increase statistically insignificant. The cumulative increase in the correlation for women is slightly smaller than the one for men.

Once again, the average correlation depicted in Figure 3 exceeds the correlation estimated using the full sample. Weighting the correlations in the figure by the number of workers observed in each age category gives an average correlation of 0.57 for men and 0.43 for women. We again recognize that these are estimated on a different sample, and so we estimate the correlation on a pooled sample of the observations used in Figure 3.[21] We find a correlation of 0.43 for men and 0.42 for women. Sample differences again enlarge the gap

[19] Appendix F shows the life cycle correlation estimated on data constructed using independence assumption II. The broad message is unchanged: the correlation increases over life cycle for men, and has a U-shaped pattern for women, with the exception of the first few years of the career.

[20] For men age 55, the standard deviation of worker types drops considerably compared to age 54, leading our estimate of the correlation to exceed 1.

[21] Note that the pooled sample in the time series and the life cycle analyses are different. The initial pooled selection of workers who switch an employer after an unemployment spell is the same for both. We then require that each employer has at least two employees in the considered category. A firm might have two workers in a calendar year, but not have two workers of the same age, in which case the firm only appears in the time series analysis. After dropping firms with one worker, we drop workers with a single employer in the data set and repeat until we have a sample with at least two observations for each worker and each firm. This gives us 464,828 men and 289,546 women in the life cycle analysis, significantly less than in the time series.

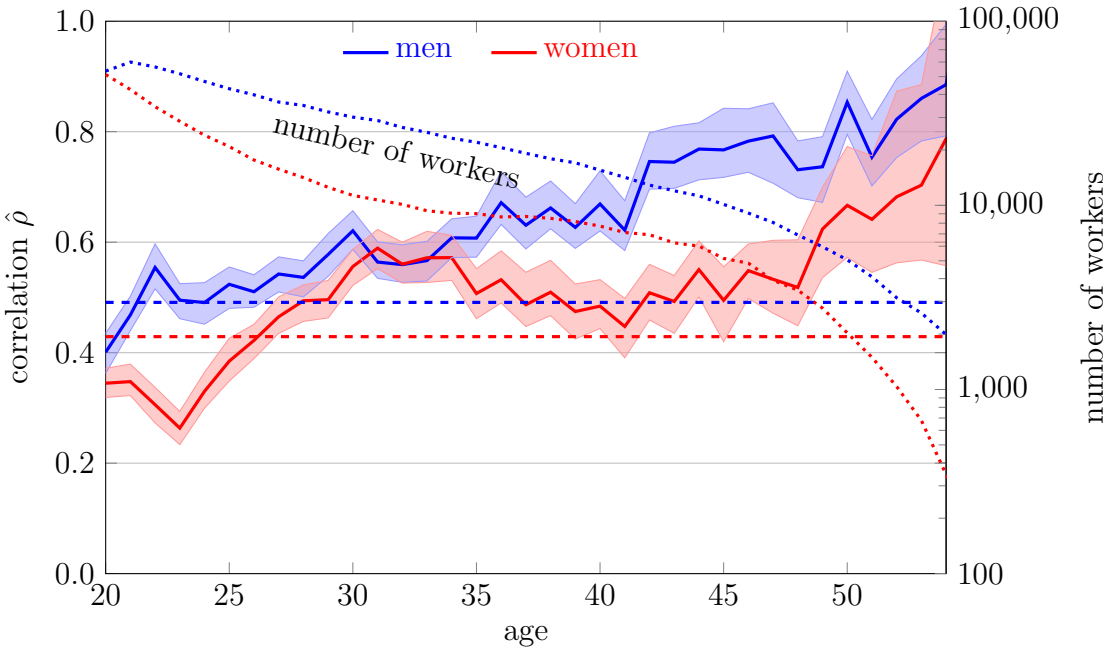


Figure 3: Correlation between worker and firm types by age using residual log wages under independence assumption III. Solid lines are computed age-by-age and shaded ares are bootstrapped 95 percent confidence intervals. For each age, the sample considers all workers who switched employers after an unemployment spell at that age, and includes one job for each employment spell of these workers. The sample only includes the wage observations for that age, even if the match continued at other ages. Dotted lines are the number of workers in the age bin who satisfy our selection criterion. For both men and women, we restrict attention to ages 20–54. Dashed lines are computed using the full sample, reported in Tables 1(4) and 2(4).



Figure 4: Standard deviations of log wages, worker types, and job types using residual log wags under independence assumption III. Each line is computed age-by-age and uses one observation per employment spell. See the description of Figure 3 for more details.

for men.

The remaining difference between the life cycle and pooled analysis reflects two factors. First, worker types vary over the life cycle, as we discussed in the previous subsection. Second, firms are collections of heterogenous jobs, and the job type for twenty-year-olds might be different than the job type for fifty-year-olds, even if they are working at the same firm. The life cycle analysis treats jobs for each age separately, but the pooled sample does not. Again, this suggests that the pooled analysis likely understates the true correlation between types.

Figure 4 shows the standard deviation of worker and job types over the life cycle. There is much less action in this figure than in Figure 3, particularly for men. The standard deviation of men's types increases from age 20 to 30 and then rises very slowly thereafter. For women, the standard deviation of worker and job types shows a life cycle pattern similar to the pattern in the correlation. We stress that the patterns depicted in Figure 3 are driven entirely by the behavior of the covariance between types, with the life cycle pattern of standard deviations working in the opposite direction.

Our preferred interpretation of these results is that workers gradually sort over the life cycle. At the start of their careers, there is a lot of uncertainty about a workers' type and so sorting is imperfect. As the worker grows older, the market learns the worker's type and the best workers sort into the best jobs. In particular, the sharp increase in the variance of worker types early in the life cycle is consistent with models of learning like Farber and

Gibbons (1996), augmented with a theory of sorting between workers and jobs.

We close by noting that sample selection issues may be important for these results. The dotted line in Figure 3 shows the number of individuals who fit our sample criterion at each age. For men, this declines from 53,565 at age 20 to 1,989 at age 54. For women, the decline is from 51,294 at age 20 to 345 at age 54. We have already discussed the concern that individuals who lose their job, become unemployed, and return to work are unusual. Those workers who do this near the end of their career may be even more unusual. This leads to the concern that the increasing correlation reflects at least in part a change in sample selection over the life cycle.

To address this concern, we extend our approach to allow for individual fixed effects; see Appendix G for details. This approach identifies the life cycle component of the increase in the correlation from the change in the correlation for those individuals who appear multiple times in our analysis, i.e. those workers who have two or more years when they work both before and after a registered unemployment spell. We find that controlling for individual fixed effects significantly moderates the increase in the correlation for men and leads to a u-shaped pattern in the correlation for women, with the trough occurring around age 35–40.

4.6 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, we imagine that a firm hires a collection of workers with different skills and the relevant firm type for a high skilled worker may be very different than for a low type worker. Our approach is to treat a firm j as a cross between a firm identifier and worker's skill, and estimate 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.

To examine this, we first treat 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 Tables 1(4) and 2(4), 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

[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

Impact of Observables for Men and Women

	men			women		
	(1)	(2)	(3)	(4)	(5)	(6)
correlation of matched types $\hat{\rho}$	0.596	0.590	0.616	0.526	0.558	0.530
covariance of matched types $\hat{c}$	0.022	0.024	0.027	0.036	0.042	0.037
variance of log wages $\hat{\sigma}^2$	0.099	0.098	0.099	0.178	0.183	0.174
variance of worker types $\hat{\sigma}^2_\lambda$	0.027	0.030	0.040	0.055	0.060	0.066
variance of job types $\hat{\sigma}^2_\mu$	0.052	0.053	0.047	0.086	0.093	0.073
number of workers (thousands)	949	1,045*	917*	786	895*	646*
number of firms (thousands)	337*	247*	181	315*	241*	163
number of observations (thousands)	3,895	3,975	2,706	2,660	2,757	1,787
share of observations top-coded	0.071	0.074	0.070	0.024	0.028	0.022
independence assumption	III	III	III	III	III	III
education	yes	no	no	yes	no	no
white/blue collar	no	yes	no	no	yes	no
industry	no	no	yes	no	no	yes

Table 4: 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. Columns (1)–(3) present results for men, (4)–(6) for women. In (1) and (4), we treat each firm × education category as a separate firm. In (2) and (5), we treat each worker × job position and firm × job position as different workers and firms. In (3) and (6), we treat each worker × industry as different workers. The sample always runs from 1986–2007.

only appear once 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 × education types.

Table 4(1) and 4(4) show the results for men and women, respectively. Allowing firm types to differ by educational category slightly raises the variance of both worker and firm types for both men and women. The bigger impact is on the covariance, and hence the correlation between matched types increases from 0.49 to 0.60 for men and from 0.43 to 0.53 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. This leads to an estimate of the

residual log wage of workers with recorded education. Furthermore, workers with missing education have fewer employment spells on average, 2.4 compared to 4.1 for men, and 2.3 compared to 3.4 for women.

correlation of 0.59 for men and 0.56 for women (Table 4(2) and 4(5)). 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, 0.62 for men and 0.53 for women (Table 4(3) and 4(6)).

5 Comparison With Abowd-Kramarz-Margolis (1999)

5.1 Methodology and Results

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,

$$\omega_{i,m}^{w} = x_{i,m}'\beta + \alpha_i + \psi_{k_{i,m}} + v_{i,m}, \tag{12}$$

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.

Tables 5 and 6 (again for men and women) verify that this finding holds in our data as well. We use the same approach as in Tables 1 and 2, with one difference: the AKM correlation is only identified on the largest connected set of workers and firms. We therefore redo our analysis on this set. In Tables 5(1) and 6(1), we use all worker-firm-year wage observations that belong to the largest connected set, including those with only one observation.[23] The AKM methodology delivers essentially zero correlation between the worker and firm fixed

[23]This is somewhat different that the standard AKM methodology which includes at most one worker-firm-year observation per year, the one with the highest earnings. Following this methodology, the estimated AKM correlation is 0.024 for men and -0.030 for women.

effects, 0.033 for men and 0.005 for women.[24]

The remaining columns in Tables 5 and 6 correspond to the data sets used in Tables 1 and 2, respectively, with the additional restriction to the largest connected set. Using the fixed effects approach, the estimated correlation lies between -0.002 and 0.057 for men and 0.005 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.

Figure 5 shows the estimated correlation between fixed effects only using workers who switch employers after an intervening unemployment spell within a given calendar year. The estimated correlation is smaller than -0.10 in every year for both men and women and significantly less than the correlation computed using the full sample. It is typically about 0.6 less than our estimates of the correlation.

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. 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. It is certainly possible that the joint distribution is not elliptical, but we believe that is unlikely to explain much of the difference.

Second, the identifying assumption in the AKM approach is that the error term in the wage equation $v_{i,m}$ has mean zero conditional on the identity of the worker i and firm $k_{i,m}$. In a version of Shimer and Smith (2000) with idiosyncratic match quality, for example, this assumption is likely to be violated due to a selection problem: some matches will only be formed if the idiosyncratic shock is high while other matches will be formed with a bigger set of idiosyncratic shocks. This "endogenous mobility" problem would lead to biased estimates of worker and firm fixed effects. We stress that our independence assumptions do not solve this problem, even when those assumptions are appropriate for our approach.

Finally, even if the identifying assumption in the AKM approach is 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. Otherwise, there

[24]Gruetter and Lalive (2009) find an AKM correlation of -0.21 for Austria. We attribute the difference to the fact that they only have a 25 percent sample of the Austrian private sector employment over an eight year period, while we have the full private sector over a longer period. An implication of Proposition 3 below is that increasing the number of matches per worker and per firm reduces the bias in the fixed effects estimates.

Comparison with AKM: Men

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\hat{\rho}$	0.589	0.633	0.616	0.491	0.450	0.435	0.418
$\hat{\rho}_{AKM}$	0.033	0.035	0.057	0.033	0.033	0.015	-0.002
$\hat{\rho} - \hat{\rho}_{AKM}$	0.555	0.598	0.559	0.458	0.418	0.420	0.419
number of workers (thousands)	4,138	3,651	2,810	1,100	676	650	652
number of firms (thousands)	750	650	498	234	206	179	180
number of observations (thousands)	63,630	63,043	16,129	4,375	3,505	2,810	2,815
share of observations top-coded	0.185	0.186	0.134	0.078	0.060	0.033	0.041
independence assumption	naïve	I	II	III	IV	IV	IV
observations included	all	all	all	longest	longest	first	last
first year of the sample	1972	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 (3)–(7) aggregate residual wages to the worker-firm match level by taking a weighted average of wages within the match across years. We use a naïve measure of correlation in column (1), and our method in columns (2)–(7). 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.

Comparison with AKM: Women

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\hat{\rho}$	0.569	0.606	0.435	0.429	0.428	0.448	0.436
$\hat{\rho}_{AKM}$	0.005	0.007	0.068	0.037	0.067	0.055	0.038
$\hat{\rho} - \hat{\rho}_{AKM}$	0.564	0.600	0.368	0.392	0.361	0.394	0.398
number of workers (thousands)	3,386	3,088	2,358	951	540	503	504
number of firms (thousands)	821	716	522	238	196	160	162
number of observations (thousands)	46,679	46,275	11,101	3,190	2,336	1,771	1,773
share of observations top-coded	0.050	0.050	0.043	0.026	0.020	0.012	0.013
independence assumption	naïve	I	II	III	IV	IV	IV
observations included	all	all	all	longest	longest	first	last
first year of the sample	1972	1972	1972	1986	1986	1986	1986

Table 6: Comparison of our estimates of correlation and AKM fixed effects estimates for women. See description of Table 5 for details.

is an incidental parameter problem which causes bias and inconsistency in the measurement of the correlation between the AKM fixed effects. We turn to this issue next.

5.2 Finite Sample Bias in Estimated Correlation

Estimates of the AKM fixed effects are unbiased but noisy when each worker has a finite number of jobs and each firm has a finite number of employees. This noise in turn affects the measured correlation. To see this, first suppose we measure the firm fixed effects accurately, because we have a lot of data for each firm. Due to idiosyncratic noise, we expect to still measure workers' fixed effects with noise, boosting its cross-sectional variance. Although this noise does not affect the covariance between worker and firm fixed effects, it biases the measured correlation towards zero.

Now suppose both fixed effects are measured with noise, as is the case when M_i and N_j are both finite. In some instance, a particular firm fixed effect is overestimated. Then the first order conditions from minimizing the sum of squared residuals in equation (12) implies that for a given set of wages, the fixed effects for that firm's workers will be underestimated. The opposite happens if the firm fixed effect is underestimated. This induces a negative bias in the covariance between the worker and firm fixed effects, potentially making the estimated covariance negative even if the true covariance is positive.

We develop an analytically tractable model economy to derive the potential magnitude of this bias. The model economy is simpler than the real world economy in a few ways. First, we assume that the AKM wage equation (12), is correctly specified. Second, we assume that

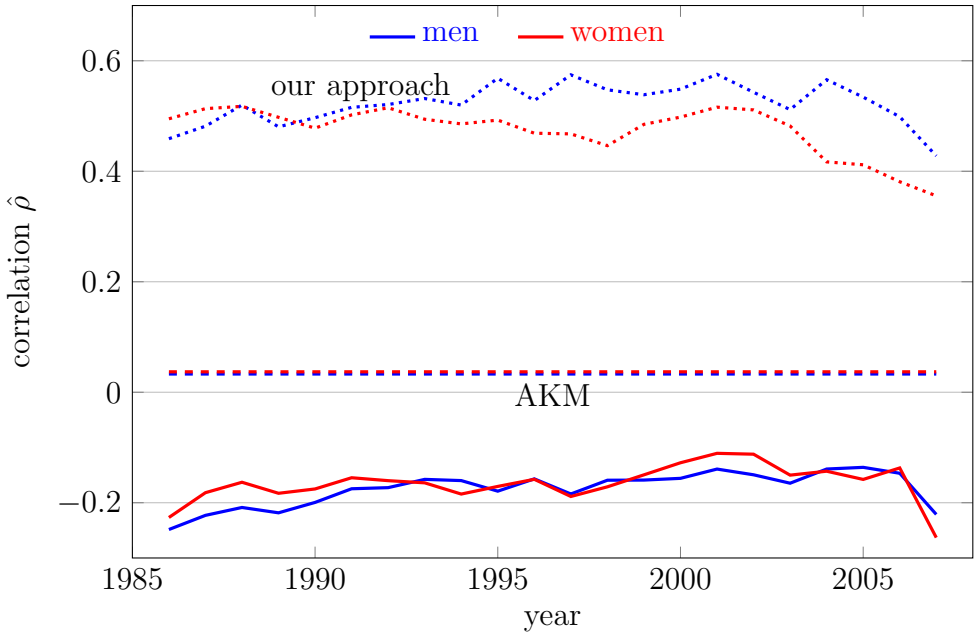


Figure 5: AKM correlation between worker and firm types using one job per spell under independence assumption III. Solid lines are computed year-by-year. For each year, the sample considers the largest connected set of workers who switched employers after an unemployment spell within that year, and includes 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 column Tables 1(4) and 2(4). Dotted lines show the estimates using our approach on the the largest connected set.

all workers have the same number of jobs and all firms have the same number of employees. Third, we assume that there are no loops in the matching graph, in a sense that we make precise below.

In our model economy, there are infinitely many workers indexed by i, each with an AKM wage effect $\alpha_i \in \mathbb{R}$. There are also infinitely many firms indexed by j, each with an AKM wage effect $\psi_j \in \mathbb{R}$. The workers' and firms' characteristics y and z and types λ and μ do not play a role here, and so we suppress them. Worker i is matched with M different employers and firm j is matched with N different workers. This notation again suppresses any explicit notion of time and dynamics since that is not essential to our analysis.

For simplicity we assume there are no match-varying covariates $x_{i,m}$. Wages are set according to equation (12), where $v_{i,m}$ is an independent shock with mean 0 and standard deviation σ_v. This means that the AKM model is correctly specified, although we measure wages at the match (rather than year) level.

We turn now to the matching graph. The matching graph is a set of nodes and links,

where the nodes represent workers and firms, and there is a link between a firm and a worker node if the firm ever employed the worker. A key assumption is that the graph has no loops. A loop would arise in the graph, for example, if there are workers i and i' and firms j and j' such that both i and i' work for j and j'. Loops can also be larger. For example i works for j and j', i' works for j' and j'', and i'' works for j and j''. As we have already observed, loops are present in real-world networks. We discuss later what happens if we relax this assumption.

Let σ_α denote the standard deviation of α, σ_ψ denote the standard deviation of ψ, and ρ_{AKM} denote the correlation between α and ψ in matched pairs. We do not impose any distributional assumptions, but remind the reader that if the joint distribution of matched α and ψ is elliptical and $(\sigma_\alpha + \rho_{AKM}\sigma_\psi)(\sigma_\psi + \rho_{AKM}\sigma_\alpha) > 0$, then the correlation between λ and μ, our measure of type, satisfies $\rho = \rho_{AKM}$; see Proposition 1.

We are interested in understanding what happens if we estimate equation (12) using a data set produced with this data generating process. The following Proposition gives the result:

Proposition 3 *Assume $M \geq 2$ and $N \geq 2$ with at least one inequality strict. Suppose we use ordinary least squares to estimate equation* (12) *on the largest connected set of workers and firms. Then the joint distribution of the estimated fixed effects $\hat{\alpha}$ and $\hat{\psi}$ in matched pairs has variance-covariance matrix*

$$\begin{pmatrix} \sigma_\alpha^2 + \frac{N(M-1)\sigma_v^2}{M(MN-M-N)} & \rho_{\text{AKM}}\sigma_\alpha\sigma_\psi - \frac{\sigma_v^2}{MN-M-N} \\ \rho_{\text{AKM}}\sigma_\alpha\sigma_\psi - \frac{\sigma_v^2}{MN-M-N} & \sigma_\psi^2 + \frac{M(N-1)\sigma_v^2}{N(MN-M-N)} \end{pmatrix}. \tag{13}$$

If $\rho_{\text{AKM}} \geq 0$, then the correlation between $\hat{\alpha}$ and $\hat{\psi}$ in matched pairs is smaller than ρ_{AKM}, and strictly so if the error in the wage equation has a positive variance.

The proof is in Appendix A. The proof proceeds by first noting that $\hat{\alpha}_i$ and $\hat{\psi}_j$ are unbiased estimators of α_i and ψ_j, as one would expect given that the error term in the wage equation is strictly exogenous. We call the difference $\hat{\alpha}_i - \alpha_i$ and $\hat{\psi}_j - \psi_j$ the AKM residuals. In the variance-covariance matrix (13), the first term in each expression corresponds to the variance-covariance matrix of the true effects α_i and ψ_j, while the second term corresponds to the variance-covariance matrix of the AKM residuals.

After constructing the AKM residuals, the proof then analyzes how shocks to the wage in one match spill through the matching graph, affecting the AKM residuals of workers and firms that are not necessarily directly affected by the wage shock. Finally, it uses the fact that the wage shocks are independent with known variance to compute the variance in the AKM residuals and the covariance between the AKM residuals of matched workers and firms.

If the number of observations per worker and firm, M and N, converge to infinity at the same rate, then all the second terms in the covariance matrix (13) converge to zero and the correlation between $\hat{\alpha}$ and $\hat{\psi}$ converges to ρ_{AKM}, the correlation between α and ψ. But in practice M and N are small and so that limit is not empirically relevant.

We use Proposition 3 to explore the quantitative biases in the fixed effects estimates. To do this, we need to feed in numbers. Our approach gives us precise estimates of the unconditional standard deviation of log wages σ, the standard deviation of worker types σ_λ, the standard deviation of job types σ_μ, and the correlation between worker and job types in matched pairs ρ. Under the assumption of an elliptical distribution of worker and firm types with $(\sigma_\lambda - \rho\sigma_\mu)(\sigma_\mu - \rho\sigma_\lambda) > 0$, the arguments in the proof of Proposition 1 imply

$$\sigma_\alpha = \frac{\sigma_\lambda - \rho\sigma_\mu}{1-\rho^2}, \quad \sigma_\psi = \frac{\sigma_\mu - \rho\sigma_\lambda}{1-\rho^2}, \quad \text{and } \rho_{AKM} = \rho.$$

Moreover, the variance of the residual in equation (12) satisfies

$$\sigma_v^2 = \sigma^2 - \left(\sigma_\alpha^2 + \sigma_\psi^2 + 2\rho_{AKM}\sigma_\alpha\sigma_\psi\right) = \sigma^2 - \frac{\sigma_\lambda^2 + \sigma_\mu^2 - 2\rho\sigma_\lambda\sigma_\mu}{1-\rho^2}.$$

The remaining numbers are M and N, the number of matches per worker and per firm. In the data, there is considerable dispersion and skewness in these numbers. For example, among the 1.1 million men in Table 1(4), the median value of M_i is 3 and the mean is 3.98. Among their 0.2 million employers, the median value of N_j is 5 and the mean is 18.68. The corresponding medians and means for women are 3 and 3.36 for M_i and 4 and 13.41 for N_j. The theory does not tell us which numbers to use.

We find that if we plug in the median values of M_i and N_j, the variance-covariance matrix (13) overpredicts the bias in the AKM correlations in Tables 5(4) and 6(4). For example, the equation predicts an AKM correlation of -0.008 for men and -0.066 for women, compared to 0.033 and 0.037 reported in the Tables.

On the other hand, using the mean numbers for M_i and N_j yields more modest biases, an AKM correlation of 0.218 for men and 0.186 for women. Interestingly, these are close to what we find on the artificial data sets that we use for our bootstrap procedure (see Appendix C). These data sets are designed to match the variance-covariance structure $(\sigma, \sigma_\lambda, \sigma_\mu, \rho)$ and the entire distribution of M_i and N_j, not just the means and medians. On average, we find that the AKM correlation is 0.217 for men and 0.184 for women in the artificial data sets. We interpret this result as suggesting that M and N in Proposition 3 should be interpreted as the mean values of these parameters.

The fact that an AKM correlation in our bootstrap procedure is higher than in the real-

world data implies that the artificial data and real-world data sets differ in an important dimension. One difference is the presence of loops in the matching graph. The proof of Proposition 3 relies on an assumption that there are no loops in the matching graph, since that ensures independence of shocks in the wage equation as we step away from a particular match.

What happens if there are loops? Our intuition is that loops reduce the number of workers and firms who are a given number of steps removed from a particular match, much like reducing M and N. An alternative way to think about this is that loops effectively create some perfectly correlated shocks, since we can reach the same node following different paths. Correlated shocks act like an increase in the variance of the shock in the wage equation. And an increase in the variance of the shock has the same effect on the correlation as a reduction in M and N. Thus loops should raise the variance of fixed effects and reduce the covariance of fixed effects measured using the AKM approach.

While we know loops exist in the data, it is unclear how to recreate the types and frequencies of loops that we see in artificial data sets. These loops presumably reflect the fact that there are clusters in the matching graph, with matches more likely within clusters than across them, even conditional on λ and μ. Modeling clusters is tricky even in simulated data (Schaeffer, 2007), and extending the results in Proposition 3 to handle realistic clusters goes beyond the scope of this paper. Nevertheless, we have found using Monte Carlo methods that introducing clusters further depresses the estimated correlation between fixed effects when using the AKM methodology, in line with our empirical results.

We can also examine how the AKM estimator behaves using our alternative bootstrap procedure, where we hold fixed the set of matches and draw uncorrelated types for each worker and firm. This is essentially the approach taken by Andrews, Gill, Schank and Upward (2008). We find that the AKM estimator is biased, but the bias is modest. When we feed in a correlation of 0, our 95 percent confidence interval for the correlation is $[-0.0457, -0.0431]$. The modest bias reflects the tension between a negative estimate of the covariance and an overestimate of the variance of worker and firm fixed effects.

The bottom line is that the bias in the AKM estimator is quantitatively significant even if the model is correctly specified and even if we have a long panel with many workers and firms. The bias in the correlation between matched worker and firm effects is worse if the true correlation is positive, since the overestimate of the variance of the worker and effects and the underestimate of the covariance between the worker and firm effect both push the measured correlation towards zero.

6 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 to 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.

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. and Francis Kramarz, "The Analysis of Labor Markets using Matched Employer-Employee Data," *Handbook of Labor Economics*, 1999, *3*, 2629–2710.

_ , _ , 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.

Azevedo, Eduardo M. and Jacob D. Leshno, "A Supply and Demand Framework for Two-Sided Matching Markets," *Journal of Political Economy*, 2016, *124* (5), 1235–1268.

Bagger, Jesper and Rasmus Lentz, "An Equilibrium Model of Wage Dispersion and Sorting.," 2016. Mimeo.

_ , 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.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa, "A Distributional Framework for Matched Employer Employee Data," 2016. University of Chicago Mimeo.

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.

Farber, Henry S. and Robert Gibbons, "Learning and Wage Dynamics*," *Quarterly Journal of Economics*, 1996, *111* (4), 1007–1047.

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

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.

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.

Menzel, Konrad, "Large Matching Markets as Two-Sided Demand Systems," *Econometrica*, 2015, *83* (3), 897–941.

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.

Schaeffer, Satu Elisa, "Graph Clustering," *Computer Science Review*, 2007, *1* (1), 27–64.

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 Omitted Proofs

Proof of Proposition 1. 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 equation (2).

If $\sigma_\alpha + \rho_{AKM}\sigma_\psi$ and $\sigma_\psi + \rho_{AKM}\sigma_\alpha$ are both positive, this equation implies 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, equation (2) implies 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 $\rho = -\rho_{AKM}$. Equation (2) implies 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 $\sigma_\alpha + \rho_{AKM}\sigma_\psi = 0$, equation (2) implies $\sigma_\lambda = 0$. If $\sigma_\psi + \rho_{AKM}\sigma_\alpha = 0$, then $\sigma_\mu = 0$. In either case, the correlation between λ and μ is undefined. ■

Proof of Lemma 1. For worker i, consider M_i independent observations from the worker-specific distribution of log wages and densities, say $(\omega_{i,m}^w, t_{i,m}^w)_{m=1}^{M_i}$. Now consider a random variable ω which is equal to $\omega_{i,m}^w$ with probability $t_{i,m}^w/T_i^w$, where as usual $T_i^w = \sum_{m=1}^{M_i} t_{i,m}^w$. The population variance of this random variable is

$$(s_i^w)^2 \equiv \frac{\sum_{m=1}^{M_i} t_{i,m}^w \left(\omega_{i,m}^w - \hat{\lambda}_i\right)^2}{T_i^w}, \tag{14}$$

where $\hat{\lambda}_i = \frac{\sum_{m=1}^{M_i} t_{i,m}^w \omega_{i,m}^w}{T_i^w}$.

Now let ω_1 and ω_2 denote two independent draws from this distribution. With probability $\sum_{m=1}^{M_i}(t_{i,m}^w/T_i^w)^2$, the two draws come from the same observation m and so $\omega_1 = \omega_2$. Otherwise they come from different observations m and m'. In this event, the two observations were themselves draws from a distribution with variance $(\sigma_i^w)^2$ and so $\mathbb{E}(\omega_1 - \omega_2)^2 = 2(\sigma_i^w)^2$. (Recall that if x_1 and x_2 are two independent draws of a random variable with variance

σ_x^2, then the expected value of their squared difference equals twice the variance of x, $\mathbb{E}(x_1 - x_2)^2 = 2\mathbb{E}(x^2) - 2(\mathbb{E}x)^2 = 2\sigma_x^2$.) Combining this gives us

$$\mathbb{E}(\omega_1 - \omega_2)^2 = 2\left(1 - \sum_{m=1}^{M_i}\left(\frac{t_{i,m}^w}{T_i^w}\right)^2\right)(\sigma_i^w)^2 = \frac{2(\sigma_i^w)^2}{\beta_i^w},$$

where the second equation uses the definition of β_i^w. Inverting this tell us that the variance of a single draw ω from this distribution is $(\sigma_i^w)^2/\beta_i^w$.

We have now found two different measures of the variance of a single draw from the M_i observations. Equating those measures gives us that the product of the Bessel correction factor and the population variance, $\beta_i^w(s_i^w)^2$, is a random variable with expected value $(\sigma_i^w)^2$. The unbiased estimator of the variance in equation (5) follows immediately from the expression for s_i^w in equation (14). ■

Proof of Proposition 3. The first order condition from minimizing the sum of squared residuals in equation (12) is equivalent to two moment conditions:

$$\hat{\alpha}_i = \frac{1}{M}\sum_{m=1}^{M}(\omega_{i,m}^w - \hat{\psi}_{k_{i,m}}) \text{ for all } i, \tag{15}$$

$$\hat{\psi}_j = \frac{1}{N}\sum_{n=1}^{N}(\omega_{j,n}^f - \hat{\alpha}_{h_{j,n}}) \text{ for all } j. \tag{16}$$

Standard results imply that the expected value of $\hat{\alpha}_i$ is α_i and the expected value of $\hat{\psi}_j$ is ψ_j.[25]

When the variance of the error in the wage equation, σ_v^2, is zero, the measured fixed effects are exactly proportional to the types, $\hat{\alpha}_i = \alpha_i$ and $\hat{\psi}_j = \psi_j$. In the more interesting case when the variance in the wage equation is positive, the differences $\hat{\alpha}_i - \alpha_i$ and $\hat{\psi}_j - \psi_j$, which we refer to hereafter as the "AKM residuals," are random variables with mean zero and some variance. Moreover, because of the structure of who matches with whom, the AKM residuals are correlated across matched pairs of workers and firms. The bulk of the proof consists of finding these variances and covariances.

As a preliminary step, we seek to understand how shocks in the wage equation affect estimated fixed effects. Consider the impact of raising the error in the wage equation (12) by 1 for the m^{th} job of some worker i. This directly affects the fixed effects estimates $\hat{\alpha}_i$

[25]The pair of equations is actually underidentified, so a correct statement is that the expected value of $\hat{\alpha}_i$ is $\alpha_i + \gamma$ and the expected value of $\hat{\psi}_j$ is $\psi_j - \gamma$ for some constant γ. The constant γ reflects an indeterminacy in the fixed effect measurement that does not affect the correlation between the fixed effects on any connected set of workers and firms; hereafter we normalize it to 0 for convenience.

and $\hat{\psi}_{k_{i,m}}$. Let $\Delta\alpha_0$ denote the change in $\hat{\alpha}_i$ and $\Delta\psi_0$ denote the change in $\hat{\psi}_{k_{i,m}}$. From the moment conditions (15) and (16), these satisfy

$$\Delta\alpha_0 = \frac{1}{M}\big(1 - \Delta\psi_0 - (M-1)\Delta\psi_1\big),$$
$$\Delta\psi_0 = \frac{1}{N}\big(1 - \Delta\alpha_0 - (N-1)\Delta\alpha_1\big)$$

where $\Delta\alpha_1$ and $\Delta\psi_1$ denote the change in the estimated fixed effect for all the other employees of $k_{i,m}$ and all the other employers of i. These changes in the fixed effects propagate through the network of workers and jobs. Let $\Delta\alpha_n$ and $\Delta\psi_n$ denote the change in the estimated fixed effect of workers and firms who are n steps removed from i or $k_{i,m}$, i.e. matched with a worker or firm that is $n-1$ steps removed. Again using the moment conditions, these satisfy

$$\Delta\alpha_n = \frac{1}{M}\big(-\Delta\psi_{n-1} - (M-1)\Delta\psi_{n+1}\big),$$
$$\Delta\psi_n = \frac{1}{N}\big(1 - \Delta\alpha_{n-1} - (N-1)\Delta\alpha_{n+1}\big).$$

The unique bounded solution to these equations is:

$$\Delta\alpha_n = \begin{cases} \frac{1}{M(M-1)^{n/2}(N-1)^{n/2}} & \text{if } n \text{ is even} \\ -\frac{1}{M(M-1)^{(n-1)/2}(N-1)^{(n+1)/2}} & \text{if } n \text{ is odd} \end{cases}$$
$$\Delta\psi_n = \begin{cases} \frac{1}{N(M-1)^{n/2}(N-1)^{n/2}} & \text{if } n \text{ is even} \\ -\frac{1}{N(M-1)^{(n+1)/2}(N-1)^{(n-1)/2}} & \text{if } n \text{ is odd} \end{cases}$$

This solution oscillates around 0, with even n corresponding to positive deviations and odd n corresponding to negative deviations. Moreover, if $M \geq 2$ and $N \geq 2$ with one inequality strict, the sequence converges to 0.

Now each worker i is matched with M firms. Each of those matches induces a variance of the fixed effect of $(\Delta\alpha_0)^2\sigma_v^2 = \frac{\sigma_v^2}{M^2}$. Since the shocks are independent across matches, all of those matches together create a variance of the fixed effect totalling $\frac{\sigma_v^2}{M}$.

Additionally, each of i's M employers has $N-1$ other employees. Each of those $M(N-1)$ matches induces a variance of the fixed effect of $(\Delta\alpha_1)^2\sigma_v^2 = \frac{\sigma_v^2}{(M(N-1))^2}$, independent across matches. Thus all of those matches together create a variance of the fixed effect totalling $\frac{\sigma_v^2}{M(N-1)}$.

Proceeding by induction, the matches that are n steps removed from i create a variance of the fixed effect that collectively accounts for $\frac{\sigma_v^2}{M(M-1)^{n/2}(N-1)^{n/2}}$ for n even and

$\frac{\sigma_v^2}{M(M-1)^{(n-1)/2}(N-1)^{(n+1)/2}}$ for n odd. Summing across n gives

$$\text{var}\left(\hat{\alpha}_i - \alpha_i\right) = \frac{N(M-1)\sigma_v^2}{M(MN-M-N)}.$$

A similar logic implies

$$\text{var}\left(\hat{\psi}_j - \psi_j\right) = \frac{M(N-1)\sigma_v^2}{N(MN-M-N)}.$$

These are the variances of the AKM residuals.

We can also compute the covariance between a matched worker i and firm $j = k_{i,m}$'s AKM residuals. First, the shock $v_{i,m}$ induces a covariance of $(\Delta\alpha_0)(\Delta\psi_0)\sigma_v^2 = \frac{\sigma_v^2}{MN}$.

Second, each of worker i's $M-1$ other employment relationships has a shock $v_{i,m'}$. This shock induces a covariance of $(\Delta\alpha_0)(\Delta\psi_1)\sigma_v^2 = -\frac{\sigma_v^2}{MN(M-1)}$ between the AKM residuals of i and j, since those workers are one step removed from firm j. The total of these shocks is $\frac{-\sigma_v^2}{MN}$. Third, each of the workers' $M-1$ other employers has $N-1$ other employees. The wage shock in each of those employment relationships induces a covariance of $(\Delta\alpha_1)(\Delta\psi_2)\sigma_v^2 = \frac{-\sigma_v^2}{MN(M-1)^2(N-1)}$ between the AKM residuals of i and j. This covariance totals $\frac{-\sigma_v^2}{MN(M-1)}$. Fourth, each of these employees has $M-1$ other employers, each inducing a covariance of $(\Delta\alpha_2)(\Delta\psi_3)\sigma_v^2 = \frac{-\sigma_v^2}{MN(M-1)^3(N-1)^2}$ between the AKM residuals of i and j. This covariance totals $\frac{-\sigma_v^2}{MN(M-1)(N-1)}$. Each successive step away from the worker divides the total covariance alternatively by $M-1$ or $N-1$. Summing across all the contributions from all the relationships emanating from i's other employers, the covariance totals $\frac{-\sigma_v^2(M-1)}{M(MN-M-N)}$.

A symmetric argument shows that summing across all the relationships emanating from j's other employees, the covariance totals $\frac{-\sigma_v^2(N-1)}{N(MN-M-N)}$. Finally, $\frac{1}{MN} - \frac{M-1}{M(MN-M-N)} - \frac{N-1}{N(MN-M-N)} = \frac{-1}{MN-M-N}$. Therefore the covariance between the AKM residuals of i and j is

$$\text{cov}\left(\hat{\alpha}_i - \alpha_i, \hat{\psi}_{k_{i,m}} - \psi_{k_{i,m}}\right) = -\frac{\sigma_v^2}{MN-M-N}.$$

The last step is to compute the unconditional variance-covariance matrix of α and ψ. We do this using the decomposition

$$\begin{aligned}\hat{\alpha}_i &= \alpha_i + \left(\hat{\alpha}_i - \alpha_i\right), \\ \hat{\psi}_j &= \psi_j + \left(\hat{\psi}_j - \psi_j\right).\end{aligned}$$

The variance-covariance matrix of the first term is exogenous and known. We have just found the variance and covariances of the second term. Finally, the two random variables are independent. The variance-covariance matrix (13) follows immediately.

Now if $\rho_{AKM} \geq 0$, the covariance between $\hat{\alpha}$ and $\hat{\psi}$ may be negative, in which case the correlation between $\hat{\alpha}$ and $\hat{\psi}$ is negative and hence smaller than ρ_{AKM}. Alternatively, the covariance is positive but smaller than $\rho_{AKM}\sigma_\alpha\sigma_\psi$. The standard deviation of $\hat{\alpha}$ exceeds σ_α and the standard deviation of $\hat{\psi}$ exceeds σ_ψ. Thus the correlation is less than ρ_{AKM}. ■

B Selection versus Bias

Standard theories of on-the-job search imply that the independence assumptions I-IV are increasingly likely to be satisfied. That is, it seems plausible that the correlation estimates in Tables 1(4)–1(7) and 2(4)–2(7) are unbiased, whereas the larger estimates in Tables 1(2)–1(3) and 2(2)–2(3) are biased. At the same time, the sample size drops dramatically as we impose more stringent requirements on the data, which means that the estimates in the later columns may not apply to the whole population. The goal of this section is to disentangle the extent to which changes in the estimated correlation reflect a reduction in bias versus a change in sample selection. We focus first on the change in the correlation between Tables 1(3) and 1(4) for men and present similar analysis for women at the end of the section.

The estimated correlation for men drops from 0.62 to 0.49 going from Table 1(3) to Table 1(4). An obvious difference between these estimates is the time period. Whereas in Table 1(3) we use data from 1972–2007, Table 1(4) drops the first 14 years of data because we only have registered unemployment data after 1986. To show that this shorter sample does not drive our results, we replicate Table 1(3) using only data from 1986–2007. Table 7(1) shows that this raises the estimated correlation to 0.65. Thus we seek to explain why changing the independence assumption from II to III causes a decline in the measured correlation from 0.65 to 0.49 for men. Does this reflect a bias coming from a reliance on independence assumption II or selection in the sample resulting from independence assumption III or both?

A suggestive piece of evidence that selection may be important comes from dividing the workers in Table 7(1) into two groups, those who are never unemployed in Table 7(2), versus those who have at least one registered unemployment spell in Table 7(3). We maintain independence assumption II and so include all jobs for both groups of workers. We find that about 44 percent of the workers are never unemployed and they have an estimated correlation of 0.81. About 55 percent of workers experience at least one unemployment spell and they have a substantially lower estimated correlation, 0.53.[26] The average estimated

[26]The remaining 1 percent of the workers are dropped from the sample because of the requirement that all firms have two workers in the relevant data set. Some firms only have one worker who is never unemployed and one worker who is unemployed once, resulting in us dropping the firm and then potentially both workers.

Selection versus Bias: Men

	(1)	(2)	(3)	(4)	(5)	(6)
correlation of matched types $\hat{\rho}$	0.647	0.807	0.530	0.559	0.666	0.502
covariance of matched types $\hat{c}$	0.034	0.036	0.025	0.029	0.037	0.019
variance of log wages $\hat{\sigma}^2$	0.123	0.115	0.115	0.119	0.119	0.101
variance of worker types $\hat{\sigma}^2_\lambda$	0.044	0.036	0.040	0.045	0.051	0.029
variance of job types $\hat{\sigma}^2_\mu$	0.061	0.055	0.056	0.060	0.060	0.050
number of workers (thousands)	2,066	904	1,133	473	1,263*	444
number of firms (thousands)	373	187	315	211	211	112
number of observations (thousands)	10,575	2,712	7,712	3,578	3,578	1,182
share of observations top-coded	0.123	0.242	0.081	0.083	0.083	0.097
independence assumption	II	II	II	II	II	III
observations included	all	all	all	all	all	longest
ever unemployed?	some	no	yes	some	no	yes

Table 7: Estimates of correlations, covariances, and variances between matched workers' and firms' types for women. All columns use residual log wages, obtained by regressing log wages on year and age dummies, aggregated to the worker-firm match level by taking a weighted average of wages within the match across years. All measures of correlation use our method after we iteratively drop firms and workers with a single wage observation. Each column uses a different sample to estimate the correlation. Column (1) includes all workers with at least two distinct employers and treats each employer as an independent observation. Columns (2) and (3) divide those workers up into those who did not and did experience at least one registered unemployment spell. Columns (4)–(6) includes a sample of workers with at least two employment spells, each of which has at least two employers. Column (4) treats the entire sample. Column (5) treats an observation as coming from a different worker if it comes from a different employment spell; thus it only measures the correlation using within-employment-spell data. Column (6) uses independence assumption III and treats the longest jobs during each employment spell as independent observations. The sample always runs from 1986–2007.

Selection versus Bias: Women

	(1)	(2)	(3)	(4)	(5)	(6)
correlation of matched types $\hat{\rho}$	0.502	0.555	0.462	0.494	0.565	0.435
covariance of matched types $\hat{c}$	0.046	0.054	0.035	0.042	0.052	0.030
variance of log wages $\hat{\sigma}^2$	0.228	0.243	0.199	0.206	0.206	0.188
variance of worker types $\hat{\sigma}^2_\lambda$	0.086	0.092	0.070	0.080	0.095	0.059
variance of job types $\hat{\sigma}^2_\mu$	0.097	0.102	0.083	0.089	0.090	0.082
number of workers (thousands)	1,768	730	996	298	705*	264
number of firms (thousands)	386	173	321	168	168	79
number of observations (thousands)	7,667	2,104	5,375	1,870	1,870	620
share of observations top-coded	0.042	0.078	0.028	0.034	0.034	0.039
independence assumption	II	II	II	II	II	III
observations included	all	all	all	all	all	longest
ever unemployed?	some	no	yes	some	no	yes

Table 8: Estimates of correlations, covariances, and variances between matched workers' and firms' types for women. See description of Table 7 for details.

correlation in Table 7(1) is essentially a weighted average of these two numbers.

It seems intuitively reasonable that workers who are well matched are less likely to become unemployed.[27] By insisting on a registered unemployment spell with independence assumption III, we select a sample of poorly matched (low correlation) workers. Indeed, the correlation in Table 7(3) and Table 1(4) are remarkably similar. The small remaining drop in the correlation from 0.53 to 0.49 appears to reflect the fact that the sample in Table 7(3) includes some workers who have multiple jobs during a single employment spell. The wage in those jobs is conditionally correlated, inflating the estimated correlation.

This reasoning might suggest that the drop in the correlation going from Table 7(1) to Table 1(4) largely reflects sample selection issues. If so, this would point towards relying on the less selected sample in Table 7(1). The problem is that the numbers in Table 7(1) may be more biased than the previous paragraph suggests. All the jobs for the workers in Table 7(2) are drawn from the same employment spell, whereas this is the case only for some of the jobs in Table 7(3). If the correlation is higher within spells than across spells, then workers with only one spell would have a higher correlation than workers with multiple spells even if selection is not an issue. That is, the difference between the estimated correlations in Tables 7(2) and 7(3) reflects a combination of bias and selection.

To illustrate and quantify this, we construct a sample of workers who have at least

[27] Workers who do not experience unemployment are different on a number of observable dimensions. They earn a residual log wage that is 0.17 standard deviations above the mean and have less than half as many jobs as men who go through unemployment.

two employment spells, with at least two employers per spell. For this sample of workers, Table 7(4) shows that the correlation constructed in the usual manner, under independence assumption II, is 0.56. Table 7(5) treats observations from different employment spells as if they come from different workers, and so effectively only measures the correlation within employment spells, 0.67. And Table 7(6) uses independence assumption III to measure the correlation using the longest job during each spell, 0.50. For this sample, we view this last number as the correct measure of the correlation, whereas the correlation in Table 7(4) is mixture of this and the upward-biased within-spell correlation measure. Since selection is not an issue in this sample, the difference between Tables 7(5) and 7(6) reflects the bias from independence assumption II.

Finally, we try quantify the magnitudes of bias and selection. Here we rely on the fact that Tables 1(4) and 7(6) are very similar not only in terms of the correlation but also in terms of the covariance and each of the variances. We treat these two estimates as unbiased for the selected samples. The bias due to the independence assumption not being satisfied for workers with one spell is then the difference between Tables 7(5) and 7(6), $0.666 - 0.502 = 0.164$. The contribution of selection is the difference between the correlation for the two samples of never unemployed workers, Tables 7(2) and 7(5), $0.807 - 0.666 = 0.141$. We conclude that bias and selection are of roughly similar importance in explaining the results in Table 1(3) and Tables 1(4).

We can perform a similar analysis for women. Table 8(1) shows that if we measure the correlation using independence assumption II on data after 1986, the measured correlation is 0.50, compared to 0.43 using independence assumption III (Table 2(4)). As in Table 7, this reflects a higher correlation for women who are never employed and a lower correlation for women who work on each side of an unemployment spell; see Tables 8(2) and 8(3). Again, we believe this suggests that the baseline estimates in Table 2 are a lower bound on the correlation in the full sample. Finally, Table 8(4) shows the correlation for women with at least two employment spells and at least two jobs in each spell, split into the within-spell correlation Table 8(5) and across spell correlation Table 8(6). That the decomposition in Table 8(1)–8(3) is similar to the decomposition in Table 8(4)–8(6) suggests that selection is not an important issue. Instead, the decline in the measured correlation going from Table 8(1) to Table 2(4) reflects the bias in independence assumption II for women.

For women, the drop in the correlation that comes from switching from independence assumption I to II is much larger than for men, 0.62 to 0.44; see Tables 2(2) and 2(3). Bias again appears to be behind this. Using the sample from Table 2(3) but treating each worker-firm-year observation according to independence assumption I, we get a correlation of 0.58 (not reported in the table), similar to the finding in Tables 2(2). This strongly suggests

the reasonable conclusion that wage observations in the same worker-firm match in different years are not conditionally independent.

C Bootstrap

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 1 and 2, and we take distributions of $M, N, t^w_{i,\cdot}$ directly from the data.

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

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

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 2. 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 $\chi^2_{i,m}$ conditional on λ_i is the conditional variance plus the square of the mean, $\sigma^2_\mu(1-\rho^2) + \frac{\lambda_i^2\rho^2\sigma^2_\mu}{\sigma^2_\lambda}$. Thus the unconditional expectation of $\chi^2_{i,m}$ is

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

   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.

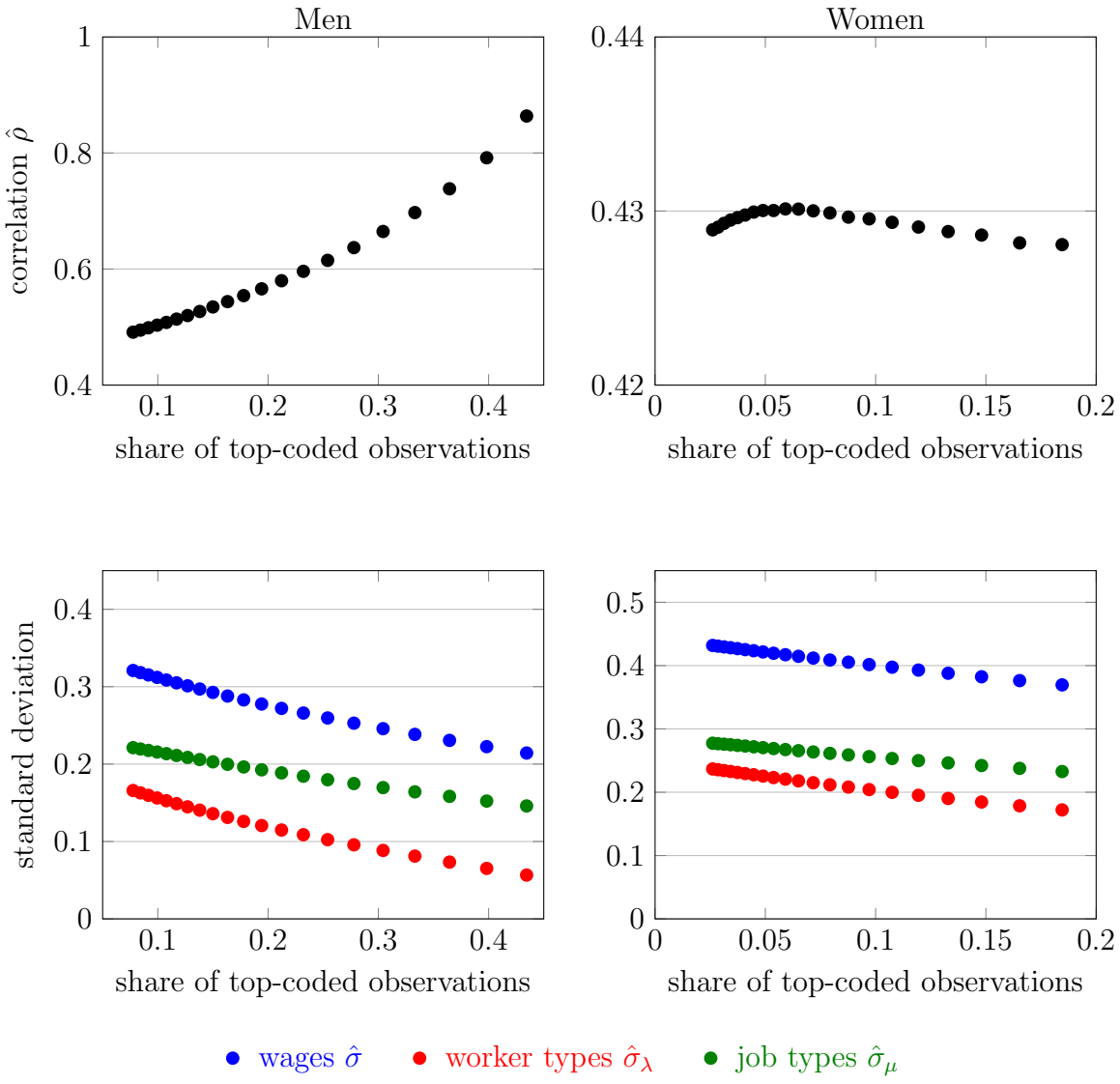


Figure 6: Impact of top-coding on estimated correlation and standard deviation of wages for men and women. Each dot corresponds to a sample where we decreased the top-code by $0, 2, 4, \ldots 40$ percent every year and truncated all wages at this new top-code. The sample of workers and firms is chosen according to Concept III, so the numbers are comparable to Column (4) of Table 1 for men and Table 2 for women. 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.

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 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 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.

Using Lemma 2, our estimate of the variance of worker types in this environment is

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

The first line uses the assumption that $t_{i,m}^w = 1$ 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 Lemma 3 to show that $\hat{\sigma}_\mu^2 = s\sigma_\mu^2$.

Finally, Lemma 4 implies that our estimate of the covariance is

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

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.

F Time Series and Life Cycle Results under Independence Assumption II

We replicate our time series and life cycle analysis on data sets constructed following using independence assumption II. We proceed in an analogous manner to the main text. For each year or age, we select workers who work for at least two distinct employers in the considered year/age. We estimate the correlation for each year/age separately. Figures 7 and 8 depict the results.

G Fixed Effects in Life Cycle Estimation

Our goal is to obtain an estimate of the correlation which controls for composition of workers. We start by noticing that the covariance $\hat{c}$ can be obtained as a coefficient from a linear regression. For each worker i in the sample, we can construct the summand,

$$c_i = \sum_{m=1}^{M_i} \frac{t^w_{i,m}}{T^w_i} \left(\frac{\sum_{m' \neq m} t^w_{i,m'} \omega^w_{i,m'}}{\sum_{m' \neq m} t^w_{i,m'}} - \bar{w} \right) \left(\frac{\sum_{n' \neq e_{i,m}} t^f_{k_{i,m},n'} \omega^f_{k_{i,m},n'}}{\sum_{n' \neq e_{i,m}} t^f_{k_{i,m},n'}} - \bar{w} \right)$$

where

$$\hat{\bar{w}} = \frac{\sum_{i=1}^{I} \sum_{m=1}^{M_i} t^w_{i,m} \omega^w_{i,m}}{\sum_{i=1}^{I} T^w_i},$$

and use weighted least squares with weights $\tau^w_i \equiv T^w_i / \sum_{i'=1}^{I} T^w_{i'}$ for worker i to regress this on the constant 1. The coefficient on the constant is the weighted average of c_i across workers, and is identical to $\hat{c}$ defined in Section 2.7.

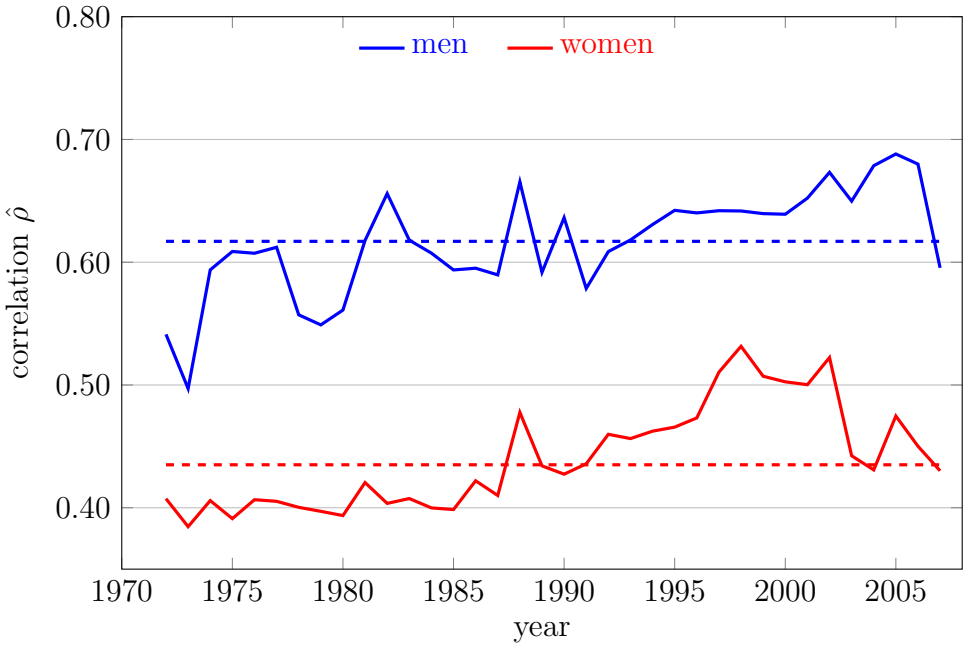


Figure 7: Correlation between worker and firm types using indpendence assumption II. Solid lines are computed year-by-year. For each year, the sample considers workers who switched employers within that year. The sample only includes the wage observations for that year, even if the match continued in other years. Dashed lines are computed using the full sample, reported in column Tables 1(3) and 2(3).



Figure 8: Correlation between worker and firm types by age using independence assumption II. Solid lines are computed age-by-age. For each age, the sample considers all workers who switched employers at that age. The sample only includes the wage observations for that age, even if the match continued at other ages. Dotted lines are the number of workers in the age bin who satisfy our selection criterion. Dashed lines are computed using the full sample, reported in Tables 1(3) and 2(3).

We can similarly obtain the variance of worker types by constructing

$$\sigma_{i,\lambda}^2 = \sum_{m=1}^{M_i} \frac{t_{i,m}^w}{T_i^w} \left(\omega_{i,m}^w - \hat{\bar{w}}\right)^2 - \sum_{m=1}^{M_i} \beta_i^w \frac{t_{i,m}^w}{T_i^w} \left(\omega_{i,m}^w - \hat{\lambda}_i\right)^2$$

and use weighted least squares with weights τ_i^w for worker i to regress this on the constant 1. The coefficient on the constant is the weighted average of $\sigma_{i,\lambda}^2$ across workers, and is identical to $\hat{\sigma}_\lambda^2$ defined in Section 2.7.

We next turn to our life cycle analysis. For the covariance, we construct $c_{i,a}$ for each worker i who has at least two jobs separated by an unemployment spell at each age a. In doing this, we use the age-specific mean wage in place of $\hat{\bar{w}}$ and age-specific weights $\tau_{i,a}^w$ for worker i at age a. We then regress this on a full set of age dummy variables. Again, the coefficients are the average covariance, $\hat{c}$, among workers with age a. We obtain the age-specific variance of worker types in the same manner.

Finally, to control for worker composition, we add worker fixed effects into the two regressions. We impose that the mean of the fixed effects is zero and look at the coefficients on the age dummies.

It is impossible to estimate the variance of job types controlling for worker fixed effects, because there is no obvious way to attach worker dummies to the summand in the formula for variance of job types. Since Figure 4 shows little life cycle pattern in the variance of job types, we feel comfortable assuming that selection is unimportant and simply rely on that measure.

We obtain the correlation by dividing the age-specific covariance, controlling for worker fixed effects, by the product of the age-specific standard deviation of worker types, again controlling for worker fixed effects, and the age-specific standard deviation of job types, not controlling for worker fixed effects.

Figure 9 presents results for men (left panel) and women (right panel). We start with the same data set as we used in Section 4.5, pooling together workers of all ages. This gives us an initial sample of 463,794 men and 289,724 women. The blue lines show the correlation without any fixed effects. This exactly replicates Figure 3.

The effect of age in the regressions including worker fixed effects is identified only off of workers whom we observe at two or more ages. Only 176,991 men and 78,434 women satisfy this restriction, a substantial reduction in the sample size. We are naturally concerned that these workers are different than their peers who only have one of these short unemployment episodes. To address this, we look at age-specific correlations for this subset of workers, depicted in the red lines in Figure 9. The difference between the red and blue lines is modest, mitigating our concern that this sample selection issue is important.

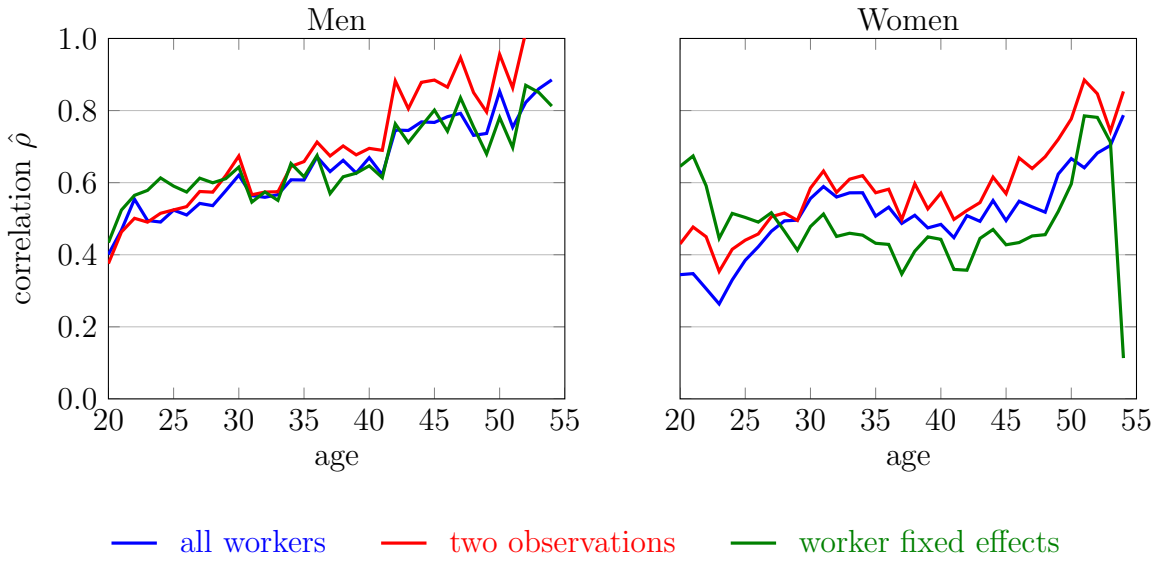


Figure 9: Correlation between worker and firm types by age using one job per employment spell (independence assuption III). The blue lines are computed age-by-age using the sample from Section 4.5; see notes under Figure 3 and the main text for more details. The red lines restrict the sample to workers who appear at least twice in the life cycle analysis, that is, they satisfy the relevant selection criteria at two or more ages. The green lines show the measure of correlation which controls for worker fixed effects.

Finally, the green lines measure the correlation after including worker fixed effects in the regression. For men, controlling for worker fixed effects dampens the rise in the correlation, reducing the slope by about a forty percent. Still, the figure shows a dramatic increase in the correlation, from around 0.4 for the youngest workers to above 0.8 for workers in their fifties.

For women the results are quite different. Selection is critical for women in their twenties. After controlling for worker fixed effects, the correlation is actually decreasing during this decade. This changes later in life. For women who are at least 30 years old, the estimates including worker fixed effects follow a similar pattern to the estimates that omit those fixed effects, albeit at a lower level. For the subsample of women whom we can observe at two or more ages, controlling for selection leads to a u-shaped pattern in the estimated correlation over the lifecycle. Finally, the estimated correlation drops dramatically at age 54 for women, but this result is based on a particularly small sample.