High Wage Workers Work for High Wage Firms*

Katarína Borovičková
New York University

Robert Shimer
University of Chicago

September 13, 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 0.5. Using Monte Carlo, we show that our estimator is accurate in realistic data sets. 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, Table 1). The one place where it has been hard to find evidence of sorting is in the labor market. A fair summary of an extensive literature following Abowd, Kramarz and Margolis (1999) (hereafter AKM) is that the correlation between the fixed characteristics of workers and their employers is close to zero and sometimes negative.[1] This is often interpreted as saying that there is no evidence that high wage workers

*We are grateful for comments from Fernando Alvarez, Stephane Bonhomme, Thibaut Lamadon, Elena Manresa, and Derek Neal. Any remaining errors are our own.

[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,

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 lies between 0.4 and 0.5, while the AKM fixed effects estimator delivers a correlation close to zero.

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 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.

Card, Heining and Kline (2013) for Germany, 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 wage firms being either high or low productivity firms, for the reasons discussed in Eeckhout and Kircher (2011).

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 restriction to finite second moments.

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.

To obtain independent observations of the wage conditional on type in the 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 only modestly on which data set we use, but 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 about 0.40 for men and 0.39 for women. The exact number depends on how we treat time-varying observable characteristics, including age and experience.

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 0.44 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 below 0.4 for men younger than 25 to above 0.6 for men in their forties, finally approaching 0.8 for men older than 55. 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 entry and exit of women from the labor force during years of peak fertility.

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, the incidental parameter problem. 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. The remaining difference between the two estimators seem to reflect the fact that our model ignores 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.

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 use cutting-edge techniques to 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 of this approach include Hagedorn, Law and Manovskii (2017) and Lopes de Melo (forthcoming).[3] Our assumption that the wages in jobs separated by an unemployment spell 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. Section 3 discusses the data that we use in our

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

analysis. Section 4 gives our main empirical results, showing that the correlation between worker and firm types lies between 0.4 and 0.5. Section 5 compares our results to the biased results from the fixed effects estimator and develops a simple framework for quantifying the importance of limited mobility bias. Section 6 briefly concludes.

2 Measuring Correlation

2.1 Matching Model

We consider an economy consisting of a measure I of workers indexed by i uniform on $[0, I]$ and a measure J of firms indexed by j uniform on $[0, J]$. Workers are distinguished by their constant characteristics, $y_i \in Y$, as well as the total number of jobs they hold, $M_i \in \{2, 3, \ldots\}$. For example, the underlying environment may be dynamic and some workers may switch jobs more often than others. Firms are distinguished by their constant characteristics, $z_j \in Z$, as well as the total number of workers they employ, $N_j \in \{2, 3, \ldots\}$. Again, if the environment is dynamic, these workers may be employed at different (overlapping or non-overlapping) points in time. In general y_i and z_j may be vector-valued and correlated with M_i and N_j, 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 in any match. Differences in G across y reflect the fact that different workers accept different jobs with different probabilities. We treat this as a primitive in our environment, although in a structural economic model, this would be an equilibrium outcome. We define

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

to be the unconditional distribution of the characteristics of *jobs* in the economy. This is distinct from the distribution of the characteristics of firms because the number of employees N_j differs across firms and may be correlated with the firm's characteristics z_j. 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 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

[4]This is the distribution of log wages in matches that actually take place. If y and z reject some wage draws, that is reflected in the matching distributions G and Ψ, not in the log wage distribution.

a worker's unobserved productivity-relevant characteristics vary over time, the equilibrium wage may be correlated with z and other unobserved characteristics captured by u.

Our key identifying assumption is that for any worker, the distribution of her employers' characteristics z and z' and the conditional log wage quantiles u and u' are independent across employment relationships conditional on the worker's characteristics y; and similarly that for any firm, the distribution of its worker characteristics y and y' and the conditional log wage quantiles u and u' are independent across employment relationships conditional on the firm's characteristics z. This assumption is satisfied, for example, in search models where workers may only search while unemployed, such as Shimer and Smith (2000). In models with on-the-job search, such as Burdett and Mortensen (1998), this 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.[5] We discuss later how we use these models to guide our measurement.

2.2 Measuring Correlation in Theory

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 may be vector-valued and in any case do not have even an ordinal interpretation.[6] 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.

This subsection defines our desired measure of expected log wages and the correlation between them precisely. We assume here 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 the Sections 2.4–2.8 explain how we can estimate the correlation between expected log wages using the limited wage data that is available.

[5]This 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.

[6]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 not possible to examine this notion of sorting using wage data alone.

We start by defining expected log wages. 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 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. Conditional independence of wage draws for a worker implies that the m^{th} log wage observation for worker i is $\omega^w_{i,m} = \lambda(y_i) + \varepsilon_{i,m}$, where $\varepsilon_{i,m}$ is independently distributed with mean 0 and a distribution that may depend on y_i. Conditional independence of wage draws for a firm implies that the n^{th} log wage observation for firm j is $\omega^f_{j,n} = \mu(z_j) + \eta_{j,n}$, where $\eta_{j,n}$ is independently distributed with mean 0 and a distribution that may depend on z_j.

We want to measure the correlation between the type of a worker and the type of her job in a typical employment relationship:

$$\rho \equiv \frac{\Sigma}{\sigma_{worker}\sigma_{\text{job}}},$$

where

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

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

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

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 take any job, in which case $G(z) = \Phi(z)$. For general wage functions and type distributions, the variance of worker and firm types will be positive in this case. We can rewrite the covariance as

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

The inner integral is zero by the definition of $\bar{w}$, hence the covariance is zero, and so is the correlation. 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 match with high wage firms.

Our measure of the correlation weighs all workers equally. We could also construct a correlation that weighs all matches equally or one that weighs all firms equally. To the extent that different workers and firms have different numbers of employment relationships and that this number is correlated with types, the measures would differ. Our measure seems like the most natural one to us, but following the approach in this paper, it is possible to construct the other measures.

2.3 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). In our notation and suppressing observable characteristics, the authors' starting point is an assumption that log wages satisfy

$$\omega^w_{i,m} = \alpha(y_i) + \psi(z_{k_{i,m}}) + v_{i,m}, \tag{1}$$

where $\alpha(y_i)$ is the worker effect, $\psi(z_{k_{i,m}})$ is the effect for the firm $k_{i,m}$ that employs i in her m^{th} match, and $v_{i,m}$ is independently and identically distributed with mean zero. An important goal in that research agenda is measuring the correlation between $\alpha(y_i)$ and ψ_{z_j} among matched worker-firm pairs (i, j). In section 5, we consider the biases in estimating this equation by treating α and ψ as fixed effects and estimating equation (1) using OLS. Here we ask a different question: Suppose $\alpha(y_i)$ and $\psi(z_j)$ are known for all workers i and j. How is their correlation are related to the correlation of $\lambda(y_i)$ and $\mu(z_j)$?

Our approach defines a worker's type $\lambda(y_i)$ to be equal to her expected log wage and a firm's type $\mu(z_j)$ to be equal to the expected log wage it pays. AKM define the units of types $\alpha(y_i)$ and $\psi(z_j)$ 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, so the wage is the sum of a worker effect, a firm effect, and an independently and identically distributed mean zero error term. Also assume that the conditional expected value of $\psi(z_{k_{i,m}})$ in a match is linear in $\alpha(y_i)$, $\int_Z \psi(z)dG_{y_i}(z) = \kappa_0 + \kappa_1\alpha(y_i)$ for all i. Then since $v_{i,m}$ has mean zero, the definition of λ and the wage equation (1) imply

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

Conversely, let $h_{j,n}$ denote the identifier of firm j's n^{th} match. Assume that the conditional expected value of $\alpha(y_{h_{j,n}})$ in a match is linear in $\psi(z_j)$, $\int_Y \alpha(y)d\Psi_{z_j}(y) = \theta_0 + \theta_1\psi(z_j)$ for all j. Then symmetrically

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

The correlation coefficient between two random variables is unaffected by a linear transformation. It follows that linearity of conditional expected values 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 natural benchmark. To show this, let $\tau(\alpha,\psi)$ denote the density function of the joint distribution of matched workers 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 denote their correlation. Then the density function τ is *elliptical* if it can be expressed as

$$\tau(\alpha,\psi) = T\left(\frac{(\alpha-\bar{\alpha})^2}{\sigma_\alpha^2} - \frac{2\rho(\alpha-\bar{\alpha})(\psi-\bar{\psi})}{\sigma_\alpha\sigma_\psi} + \frac{(\psi-\bar{\psi})^2}{\sigma_\psi^2}\right)$$

for some function T. The family of elliptical distributions includes many well-known bivariate distributions, including the bivariate normal and the bivariate t distribution.

A key property of elliptical distributions is that conditional expected values are linear. This leads to our main result comparing the correlation of our measure of types with the correlation between the AKM worker and firm effects:

Proposition 1 *Assume that the joint density of α and ψ is elliptical. Then λ and μ are linear transformations of α and ψ and the correlation between α and ψ is the same as the correlation between λ and μ.*

The proof in the appendix establishes linearity of conditional expected values for elliptical distributions. The linear transformation then follows immediately from the logic in the text.

We view equality of the AKM correlation and our correlation as a reasonable starting assumption, but of course, in reality the joint distribution of α and ψ might not be elliptical. In this case, although we would generally expect a different correlation between α and ψ than the one between λ and μ, we do not see an obvious reason to prefer one measure of the correlation over the other.

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. Our approach allows for nonlinearities and non-monotonicities in the wage equation and so is equally well-suited to these more general environments.

2.4 Measuring Correlation in Practice

We return now to the 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. Instead, we use the restrictions that $M_i \geq 2$ and $N_j \geq 2$ for all workers i and firms j and the assumption that wage observations are conditionally independent observations for each worker and firm to measure the correlation ρ. We stress that we never estimate any worker's or firm's type.

We imagine a data set that includes worker identifiers, firm identifiers, and wages. 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 $\omega^f_{j,1}, \ldots, \omega^f_{j,N_j}$. 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, $\omega^w_{i,m} = \omega^f_{j,n}$ if $j = k_{i,m}$ and $i = h_{j,n}$.

Using this data, we first perform within-between decompositions of the total variance of log wages. We show that σ^2_{worker} and σ^2_{job} correspond to the between-worker and between-job variances. We then obtain the covariance Σ 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 Σ, and hence we take their average to recover the desired moments. The next three subsections explains these measures in detail.

2.5 Measuring the Standard Deviation of Worker Types σ_{worker}

We start by measuring $\bar{w}$, the mean log wage. Consider a worker i with characteristics y. The expected value of that worker's mean log wage, $\frac{1}{M_i}\sum_{m=1}^{M_i}\omega_{i,m}^w$, is $\lambda(y)$. If we do not know the worker's characteristics, the expected value of her mean log wage is $\bar{w} = \int_Y \lambda(y)dF(y)$. Moreover, this is independent across workers and so averaging gives us $\bar{w}$:

$$\bar{w} = \frac{1}{I}\int_0^I \frac{\sum_{m=1}^{M_i}\omega_{i,m}^w}{M_i}di. \tag{2}$$

We note that this expression is not the raw average log wage in our sample, but instead weights all workers equally, consistent with the definition of $\bar{w}$.

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

$$\sigma^2 = \frac{1}{I}\int_0^I \frac{\sum_{m=1}^{M_i}(\omega_{i,m}^w - \bar{w})^2}{M_i}di, \tag{3}$$

the empirical cross-section of log wages, weighting each individual equally.

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

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

We turn next to measurement. For any individual worker i, recall that we can represent the log wage observations as $\omega_{i,m}^w = \lambda(y_i) + \varepsilon_{i,m}$ where $\varepsilon_{i,m}$ is independently and identically distributed across m with mean 0 and a variance that may depend on y_i.[7] Therefore

$$\frac{1}{M_i - 1}\sum_{m=1}^{M_i}\left(\omega_{i,m}^w - \frac{\sum_{m'=1}^{M_i}\omega_{i,m'}^w}{M_i}\right)^2$$

is an unbiased estimate of the mean within-worker variance in the population. Moreover,

[7] Because the distribution of $\varepsilon_{i,m}$ may differ across workers, we cannot apply Kotlarski's (1967) lemma to recover the distribution of λ and ε.

this is independent across workers and so averaging this gives us σ^2_{ww}:

$$\sigma^2_{ww} = \frac{1}{I}\int_0^I \frac{1}{M_i - 1}\sum_{m=1}^{M_i}\left(\omega^w_{i,m} - \frac{\sum_{m'=1}^{M_i}\omega^w_{i,m'}}{M_i}\right)^2 di. \tag{4}$$

Next, observe directly from their definitions that the variance of worker types satisfies $\sigma^2_{worker} = \sigma^2 - \sigma^2_{ww}$. 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 1

$$\sigma_{\text{worker}} = \left(\frac{1}{I}\int_0^I \frac{1}{M_i}\sum_{m=1}^{M_i}\left(\omega^w_{i,m} - \frac{1}{I}\int_0^I \frac{\sum_{m'=1}^{M_{i'}}\omega^w_{i',m'}}{M_{i'}}di'\right)^2 di\right.$$
$$\left. - \frac{1}{I}\int_0^I \frac{1}{M_i - 1}\sum_{m=1}^{M_i}\left(\omega^w_{i,m} - \frac{\sum_{m'=1}^{M_i}\omega^w_{i,m'}}{M_i}\right)^2 di\right)^{1/2} \tag{5}$$

measures the standard deviation of worker types.

The proof follows immediately from the previous definitions and a bit of algebra.

2.6 Measuring the Standard Deviation of Job Types σ_{job}

Our approach to measuring σ_{job}, the standard deviation of job types, is similar, but we need to reweight our data to reflect the fact that different workers have different numbers of jobs. To see why, note that another measure of the average log wage is

$$\bar{w} = \frac{\int_0^J \sum_{n=1}^{N_j} \frac{\omega^f_{j,n}}{M_{h_{j,n}}} dj}{\int_0^J \sum_{n=1}^{N_j} \frac{1}{M_{h_{j,n}}} dj}.$$

That is, a wage observation for worker i is weighted by $1/M_i$, reflecting the fact that we have multiple observations for each worker. Since $\int_0^J \sum_{n=1}^{N_j} \frac{1}{M_{h_{j,n}}} dj = I$, a convenient representation of this measure is

$$\bar{w} = \frac{1}{I}\int_0^J \sum_{n=1}^{N_j} \frac{\omega^f_{j,n}}{M_{h_{j,n}}} dj.$$

We can verify algebraically that this is the same as the measure of the mean log wage across workers in equation (2). Similarly we can measure the variance of log wages across jobs as

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

Again, this is mathematically identical to the variance of log wages across workers in equation (3).

We turn next to measuring the mean of the variance of log wages across jobs, defined as,

$$\sigma_{wj}^2 \equiv \int_Z \int_Y \int_0^1 (w(y,z,u) - \mu(z))^2 \, du \, d\Psi_z(y) \, d\Phi(z).$$

For any individual firm j, we can represent the log wage observations as $\omega_{j,n}^f = \mu(z_j) + \eta_{j,m}$ where $\eta_{j,m}$ is independently and identically distributed across m with a mean 0 and a variance that may depend on z_j. An unbiased estimator of that variance is

$$\frac{1}{N_j - 1}\sum_{n=1}^{N_j}\left(\omega_{j,n}^f - \frac{\sum_{n'=1}^{N_j}\omega_{j,n'}^f}{N_j}\right)^2.$$

Note that this formula does not weight worker i by $1/M_i$ because all workers are equally informative about the variance of firm j's log wage. If we take a weighted average of this variance, weighting firms by the number of jobs they are responsible for, $\sum_{n=1}^{N_j}\frac{1}{M_{h_{j,n}}}$, we measure the desired quantity:

$$\sigma_{wj}^2 = \frac{1}{I}\int_0^J \left(\sum_{n=1}^{N_j}\frac{1}{M_{h_{j,n}}}\right)\left(\frac{1}{N_j - 1}\sum_{n=1}^{N_j}\left(\omega_{j,n}^f - \frac{\sum_{n'=1}^{N_j}\omega_{j,n'}^f}{N_j}\right)^2\right) dj. \tag{6}$$

Next, we again have $\sigma_{job}^2 = \sigma^2 - \sigma_{wj}^2$. Since we have measures of both terms on the right hand side, we also have a measure of the left hand side:

Lemma 2

$$\sigma_{\text{job}} = \left(\frac{1}{I}\int_0^I \frac{1}{M_i}\sum_{m=1}^{M_i}\left(\omega_{i,m}^w - \frac{1}{I}\int_0^I \frac{\sum_{m'=1}^{M_{i'}}\omega_{i',m'}^w}{M_{i'}} di'\right)^2 di \right.$$
$$\left. - \frac{1}{I}\int_0^J \left(\sum_{n=1}^{N_j}\frac{1}{M_{h_{j,n}}}\right)\left(\frac{1}{N_j - 1}\sum_{n=1}^{N_j}\left(\omega_{j,n}^f - \frac{\sum_{n'=1}^{N_j}\omega_{j,n'}^f}{N_j}\right)^2\right) dj\right)^{1/2} \tag{7}$$

measures the standard deviation of job types.

2.7 Measuring the Correlation of Matched Types ρ

The third step is to find the covariance Σ between λ and μ in matched worker-firm pairs. 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,

$$\frac{1}{M_i}\sum_{i=1}^{M_i}\left(\frac{\sum_{m'\neq m}\omega^w_{i,m'}}{M_i-1}\right)\left(\frac{\sum_{n'\neq e_{i,m}}\omega^f_{k_{i,m},n'}}{N_{k_{i,m}}-1}\right),$$

is a random variable with expected value

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

Subtracting off unconditional means and averaging across workers leads to our measure of the covariance $\Sigma = \int_Y\int_Z(\lambda(y)-\bar{w})(\mu(z)-\bar{w})dG_y(z)dF(y)$:

Lemma 3

$$\Sigma = \frac{1}{I}\int_0^I\left(\frac{1}{M_i}\sum_{m=1}^{M_i}\left(\frac{\sum_{m'\neq m}\omega^w_{i,m'}}{M_i-1}-\frac{1}{I}\int_0^I\frac{\sum_{m'=1}^{M_{i'}}\omega^w_{i',m'}}{M_{i'}}di'\right)\left(\frac{\sum_{n'\neq e_{i,m}}\omega^f_{k_{i,m},n'}}{N_{k_{i,m}}-1}-\frac{1}{I}\int_0^I\frac{\sum_{m'=1}^{M_{i'}}\omega^w_{i',m'}}{M_{i'}}di'\right)\right)di \quad (8)$$

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, and wages in which all workers and firms have at least two conditionally independent wage observations.*

2.8 Estimators

Any real world data set only has a finite number of workers and firms. 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:

$$\hat{\rho} \equiv \frac{\hat{\Sigma}}{\hat{\sigma}_{worker}\hat{\sigma}_{job}}$$

where

$$\hat{\sigma}_{worker} = \left(\frac{1}{I} \sum_{i=1}^{I} \frac{1}{M_i} \sum_{m=1}^{M_i} \left(\omega^w_{i,m} - \hat{\bar{w}}\right)^2 - \frac{1}{I} \sum_{i=1}^{I} \frac{\sum_{m=1}^{M_i} \left(\omega^w_{i,m} - \frac{\sum_{m'=1}^{M_i} \omega^w_{i,m'}}{M_i}\right)^2}{M_i - 1} \right)^{1/2},$$

$$\hat{\sigma}_{job} \equiv \left(\frac{1}{I} \sum_{i=1}^{I} \frac{1}{M_i} \sum_{m=1}^{M_i} \left(\omega^w_{i,m} - \hat{\bar{w}}\right)^2 - \frac{1}{I} \sum_{j=1}^{J} \left(\sum_{n=1}^{N_j} \frac{1}{M_{h_{j,n}}} \right) \frac{\sum_{n=1}^{N_j} \left(\omega^f_{j,n} - \frac{\sum_{n'=1}^{N_j} \omega^f_{j,n'}}{N_j}\right)^2}{N_j - 1} \right)^{1/2},$$

$$\hat{\Sigma} \equiv \frac{1}{I} \sum_{i=1}^{I} \frac{1}{M_i} \sum_{m=1}^{M_i} \left(\frac{\sum_{m' \neq m} \omega^w_{i,m'}}{M_i - 1} - \hat{\bar{w}} \right) \left(\frac{\sum_{n' \neq e_{i,m}} \omega^f_{k_{i,m},n'}}{N_{k_{i,m}} - 1} - \hat{\bar{w}} \right),$$

and $\hat{\bar{w}} \equiv \frac{1}{I} \sum_{i=1}^{I} \frac{\sum_{m=1}^{M_i} \omega^w_{i,m}}{M_i}$.

Each of these objects is readily measured using a data set that contains worker and firm identifiers as well as wages.

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, 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.

3 Data

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,[8] we observe earnings and days worked during the year.[9] We also have some limited demographic information on workers, including their birth year and sex. We construct work experience as the total days actually worked since a worker entered the panel. After 1986, we observe registered unemployment spells, which we use in much of our analysis.

Following Card, Heining and Kline (2013), our analysis focuses 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 focus 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, which captures the effects of aggregate nominal wage growth. In some specifications, we also include controls for the worker's age and realized experience. We next aggregate the residual back up to the level of the worker-firm match. That is, for each matched worker-firm pair, we compute a weighted average of the residual across years, weighting each year by the number of days that the pair were matched in this year.[10] Our motive for aggregating wages to the level of the worker-firm match is our need for multiple conditionally independent wage observations, combined with the fact that wages are autocorrelated within the match. From now on, we refer to this average residual log wage observation as the log wage. Thus for each worker-firm pair, we have at most one wage observation.

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

[9]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.

[10]Recalls are common in the Austrian labor market (Pichelmann and Riedel, 1992). Our approach treats all instances where a worker is employed by a firm as a single wage observation.

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 individual. We thus trim our data set by first dropping any worker who only works for a single firm in the data set, then dropping any firm that only employs a single worker in the data set, and then repeating. 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. Fortunately in our case the resulting data set is always nonempty.

Finally, we identify the largest connected set in the matching graph and keep only workers and firms who belong to this set. We note that our approach does not require us to use workers and firms in a connected set. However for comparability with the AKM fixed effects estimates, we estimate our model on the largest connected set. In practice, the largest connected set contains over 99 percent of all observations and so our estimates are very similar with or without this restriction.

For our approach to provide a consistent estimate of the correlation ρ, we need each wage observation to be independent conditional on the worker identifier and conditional on the firm identifier. We approach this in two ways, in both cases motivated by economic theory. First, we use all wage observations. In models like Shimer and Smith (2000), unemployed workers randomly meet firms and decide whether to accept the job. If the job is accepted, the pair bargain on a wage and the relationship continues until a random shock pushes them back into unemployment. There is no impact of one employment relationship on the outcome of other employment relationships, and so the independence assumption is satisfied.

We also recognize, however, that workers often voluntarily move from one job to another, as in models of on-the-job search (Burdett and Mortensen, 1998; Postel-Vinay and Robin, 2002; Christensen, Lentz, Mortensen, Neumann and Werwatz, 2005). Towards that end, we define two firms for a given worker as belonging to different employment spells if there is at least one day of registered unemployment between the last day at one firm and the first day at another.[11] We keep only one firm for each employment spell for a given worker, the one where the worker worked the most days. In models with on-the-job search, the wage observations in different employment spells should be independent conditional on the worker's characteristic. We then further trim the data set so as to obtain the largest one where each worker and firm has at least two observations, and we look at the largest connected set, as described above. Since registered unemployment data are only available starting in 1986, we lose almost half of our data with this restriction.

[11]Together with our treatment of recalls, this definition effectively drops all registered unemployment spells where a worker is later recalled to a pre-spell employer.

Estimated Correlation and Variances: Men

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
correlation of matched types $\hat{\rho}$	0.476	0.530	0.438	0.416	0.423	0.384	0.418
variance of log wages $\hat{\sigma}^2$	0.184	0.199	0.159	0.149	0.143	0.133	0.134
variance of worker types $\hat{\sigma}^2_{worker}$	0.064	0.081	0.049	0.044	0.046	0.037	0.037
variance of job types $\hat{\sigma}^2_{job}$	0.080	0.093	0.069	0.067	0.063	0.053	0.059
number of workers (thousands)	2,810	2,065	1,100	1,100	1,100	1,100	1,100
number of firms (thousands)	498	372	301	234	234	234	234
number of observations (thousands)	16,129	10,572	7,577	4,376	4,376	4,376	4,376
share of observations top-coded	0.134	0.123	0.079	0.078	0.078	0.078	0.144
first year of the sample	1972	1986	1986	1986	1986	1986	1986
one match per employment spell	no	no	no	yes	yes	yes	yes
year dummies	yes	yes	yes	yes	yes	yes	yes
age dummies	no	no	no	no	yes	yes	no
quartic in experience	no	no	no	no	no	yes	no

Table 1: Estimates of variances and correlation between matched workers' and firms' types for men. All columns consider workers and firms with at least two matches over the analyzed time period who belong to the largest connected set. Column (1) considers all matches between such firms and workers over the years 1972–2007. Column (2) restricts the data to the years 1986–2007. Column (3) restricts the sample of workers from (2) further by considering only those who experienced a spell of unemployment. Column (4) considers the same set of workers as (3) but includes only the longest job for each of worker's employment spell. The set of workers and firms in columns (5)–(7) is the same as in (4) but the data differ in the construction of residual wages. In columns (5), we include age as additional controls. In column (6), we also add a quartic polynomial in experience. In column (7), we decrease the top-code in every year by 15 percent and truncate wages at the new top-code. See the main text for additional details on sample contruction.

4 Results

4.1 Main Results

Tables 1 and 2 display our main results for men and women, respectively. Column (1) uses our entire 36 year sample to estimate the total, within, and between variances of wages at the worker and firm level. We also estimate the covariance between the average wage of a worker at other jobs and the average wage a firm pays to other workers. Combining this information, we get two estimates of the correlation between types, one for men ($\hat{\rho} = 0.476$) and a somewhat lower one for women ($\hat{\rho} = 0.375$).

We recognize that due to job-to-job movements, this sample might not have independent wage observations for each worker conditional on type. Hence we focus on results from a

Estimated Correlation and Variances: Women

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
correlation of matched types $\hat{\rho}$	0.375	0.401	0.406	0.392	0.391	0.390	0.389
variance of log wages $\hat{\sigma}^2$	0.249	0.256	0.215	0.206	0.206	0.199	0.198
variance of worker types $\hat{\sigma}^2_{worker}$	0.075	0.086	0.063	0.059	0.059	0.055	0.055
variance of job types $\hat{\sigma}^2_{job}$	0.102	0.103	0.083	0.084	0.083	0.078	0.081
number of workers (thousands)	2,358	1,767	951	951	951	951	951
number of firms (thousands)	522	385	302	238	238	238	238
number of observations (thousands)	11,101	7,665	5,206	3,190	3,190	3,190	3,190
share of observations top-coded	0.043	0.042	0.028	0.026	0.026	0.026	0.051
first year of the sample	1972	1986	1986	1986	1986	1986	1986
one match per employment spell	no	no	no	yes	yes	yes	yes
year dummies	yes	yes	yes	yes	yes	yes	yes
age dummies	no	no	no	no	yes	yes	no
polynomial in experience	no	no	no	no	no	yes	no

Table 2: Estimates of variances and correlation between matched workers' and firms' types for women. See notes under Table 1 and the main text for more details on construction of the samples in columns (1)–(7).

sample which contains one job per employment spell, reported in columns (4)–(7) of Tables 1 and 2. There are three main differences between the sample in column (1) and the samples in columns (4)–(7), and so before turning to our main results, we seek to understand those differences.

First, we only have unemployment information for a shorter time span, 1986–2007. We therefore replicate our analysis in column (1) using this shorter time period. The measured correlation, reported in column (2) of Tables 1 and 2, is somewhat higher for both men and women on this shorter sample.

Second, many workers never experience a registered unemployment spell and therefore do not appear in our sample in columns (4)–(7). Column (3) of Tables 1 and 2 cuts the sample to include the same people as in columns (4)–(7). We still include all jobs for these workers, unless the corresponding firm is matched with only one worker. This substantially reduces the measured correlation, to 0.438 for men and 0.406 for women. This appears to be a feature of the type of worker who experiences unemployment. Those workers we drop going from column (2) to column (3) earn a wage that is 0.07 standard deviations above the mean. The workers we drop also have many fewer jobs than the workers who experience unemployment. Men (women) who disappear from the sample between columns (2) and (3) have 3.1 (3.0) jobs on average, while those who remain have 6.9 (5.5) jobs.[12] Sorting appears to be stronger among these high wage workers in stable jobs.

Finally, our preferred estimates start in column (4) of Tables 1 and 2, where we only include the longest job during each employment spell. In column (4), we construct the log wage residual by regressing log wages on calendar year dummies but no other individual controls. The estimated correlation is $\hat{\rho} = 0.416$ for men and $\hat{\rho} = 0.392$ for women. These numbers are only slightly lower than the ones reported in column (3) We would expect the correlation to be lower to the extent that wages are correlated across jobs within an employment spell conditional on the worker's type, as predicted by many models of on-the-job search. This pushes us to emphasize the results using one match per employment spell, since it implies the other results are biased.

In columns (5) and (6) of Tables 1 and 2, we add more controls when constructing the log wage residual. This does not affect estimated correlation for women. For men, controlling for age (column 5) changes our estimates very little but controlling for age and actual experience reduces the correlation to $\hat{\rho} = 0.384$ (column 6). We believe this is because experience is correlated with worker type (high type workers accumulate more experience) and hence removing this information from log wages reduces the estimated correlation.

As previously noted, our wage data is top-coded. While this undoubtedly affects our estimates, we believe the impact is small. To show this, we examine what correlation we would have measured had the top-code been even more severe. In column (7), we decrease the top-code in every year by fifteen percent, otherwise keeping the specification unchanged from our baseline in column (4). This roughly doubles the share of top-coded observations[13] from 7.8 percent to 14.4 percent for men and from 2.6 percent to 5.1 percent for women. While the more severe top-coding unsurprisingly reduces the total variance of log wages as well as both between variances, it scarcely affects the estimated correlation $\hat{\rho}$. Appendix C shows that this finding is robust to other thresholds for the top-code.[14]

Even though our method does not require it, in our estimation we include only firms and workers belonging to the largest connected set. Our results in columns (4)–(7) are unchanged when we use all workers and jobs with multiple observations. This is because the largest

[12] The sample in column (3) differs from the sample in columns (4)–(7) because it also includes more firms. We have also tried restricting the sample in column (3) to the same workers and firms, while still including all matches, rather than one per employment spell. This delivers the same correlation as reported in column (3).

[13] 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.

[14] 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.

connected set contains 99.99 percent of all observations.

In our preferred estimate (column 4), the standard deviation of worker types is 0.210 for men. The associated standard deviation of firm types is somewhat higher, 0.258. This means that men with one standard deviation higher average log wage typically work at firms that pay about a third of a standard deviation higher average log wage ($\hat{\rho}\hat{\sigma}_{worker}/\hat{\sigma}_{job} = 0.338$). For women, the variability is greater than for men, 0.242 for workers and 0.291 for firms, but the bottom line is similar. Women with one standard deviation higher average log wage work at firms that on average pay 0.327 standard deviations higher average log wage. High wage workers work for high wage firms.

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, and the distribution of the number of matches per worker and firm. See Appendix B 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 compare the actual correlation between types, ρ_b, to the correlation estimated using our approach, $\hat{\rho}_b$, which relies only on wage data and individual identifiers. We construct confidence intervals using the difference $\rho_b - \hat{\rho}_b$. We find that our estimator is both precise and accurate. For example, in column (4), the estimated correlation for men is $\hat{\rho} = 0.4162$, and the 95 percent confidence interval is $[0.4149, 0.4174]$. The estimated correlation for women is $\hat{\rho} = 0.3923$ and the 95 percent confidence interval is $[0.3905, 0.3937]$. 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 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 $I + J$.[15] This is computationally infeasible.[16] 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 -0.0043 and 0.0038. We are confident that our data was not generated from an economy without sorting.

4.3 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. Since age doesn't change (much) within the year, we think it is natural to condition log wages on age fixed effects. Working within years effectively introduces year fixed effects, and so this corresponds to column (5) in Tables 1 and 2.

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 increased slightly for men, from initial 0.37 in 1986 to 0.45 in 1997 where it stayed thereafter. The correlation for women fluctuated over time, peaking at 0.5 in 1992 and then falling to 0.42 in 2006. 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 0.44 for men and 0.47 for women, significantly more than the correlation using the full sample and reported in Tables 1 and 2. We see two possible reasons for this. First, the sample of workers is slightly 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 the correlation.[17] The sample

[15]A natural assumption is that if the correlation between the type of a worker and the type of each of her employers is ρ, the correlation between the type of a worker and the type of each of the workers who share a common employer is ρ^2, the correlation between the type of a worker and the type of each of the other employers of the workers who share a common employer is ρ^3, etc.

[16]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.

[17]In this pooled sample, we regress log wages on both year and age fixed effects and then aggregate all worker-firm-year residual wages back to the worker-firm level by computing an average log wage over years.

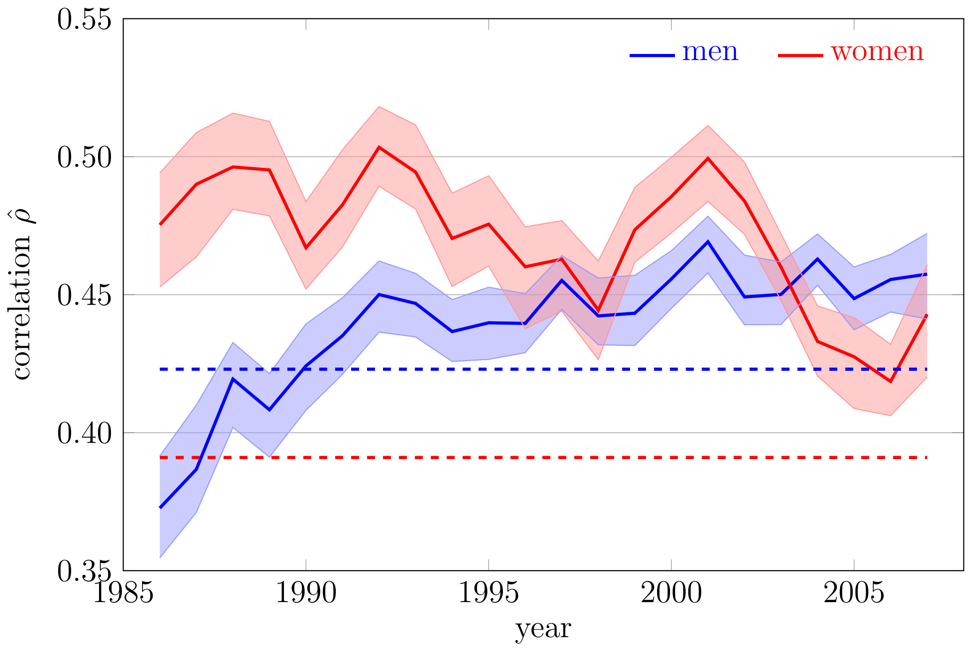

Figure 1: Correlation between worker and firm types, conditional on age fixed effects, using one job per employment spell. Solid lines are computed year-by-year and shaded areas are bootstrapped 95 percent confidence intervals. For each year, the sample considers all workers who switched employers after 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 column (5) of Tables 1 and 2.

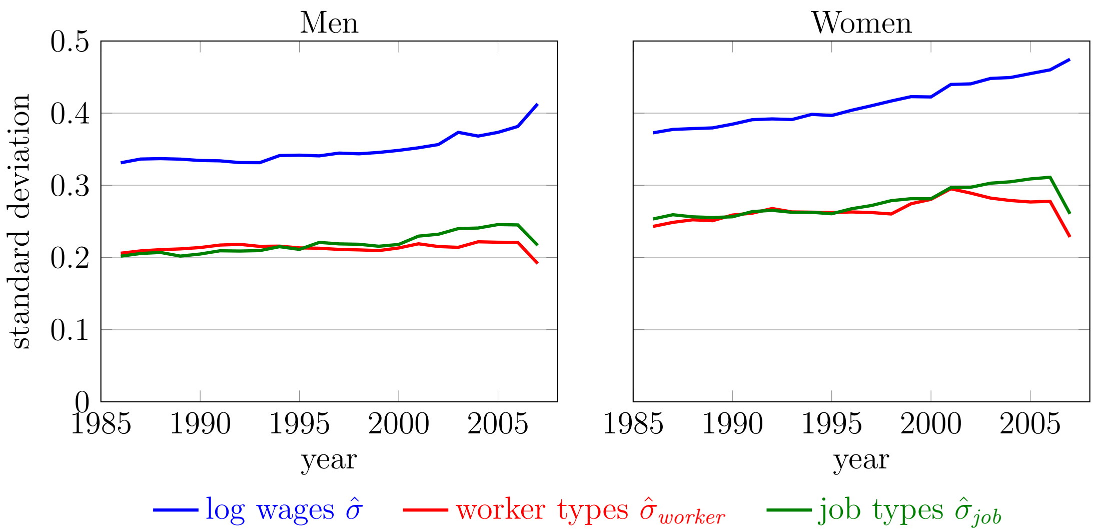

Figure 2: Standard deviations of log wages, worker types, and job types. Each line is computed year-by-year and uses one observation per employment spell and log wages conditional on year and age fixed effects. See the description of Figure 1 and the main text for more details.

contains matches for 624,427 men and 407,907 women, and the estimated correlation is 0.40 for men and 0.41 for women, close to the correlation estimated on the full time period. Sample differences matter, but actually enlarge the gap for men and only close about a quarter of the gap for women.

The second, more interesting possibility is that types gradually change over time, e.g. 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 D 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. A summary of the four panels is that the standard deviation of job types is slightly larger than the standard deviation of firm types. This contrasts with the pooled data in Tables 1 and 2, which show a bigger gap between the two standard deviations. Additionally, all four standard deviations show a modest increase over the sample period, until a sharp reversal in the last year.

We then keep only the longest match in each employment spell.

4.4 Life Cycle

We can also use our approach to create synthetic cohorts and think about how sorting evolves over the life cycle. We again redo our analysis, now using only workers with a particular age a. We regress the average log wage on year fixed effects and so our results again correspond to column (5) in Tables 1 and 2, since we have already effectively controlled for age.

Using only those workers who switch employers after an unemployment spell reduces our sample size considerably, particularly when we focus on older ages. To mitigate this issue, we look at the largest set of workers and firms with at least two observations each and do not restrict attention to the largest connected set. We do not report results for ages where we observe fewer than 100 workers because the results are too noisy and our bootstrap procedure suggests that they are unreliable. This leads us to focus on men age 20–59 and women age 20–55.

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 0.33 at age 20 to 0.76 at age 54; after this, the correlation continues to rise, but the estimates are noisier. 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. The cumulative increase in the correlation for women is only 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 0.46 for men and 0.44. 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, which includes 465,399 men and 289,862 women.[18] We find a correlation of 0.38 for men and 0.41 for women. Sample differences again enlarge the gap for men and close a quarter of the gap for women. The rest can be attributed to two explanations. First, worker types vary over the lifecycle, 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

[18]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. However, 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.

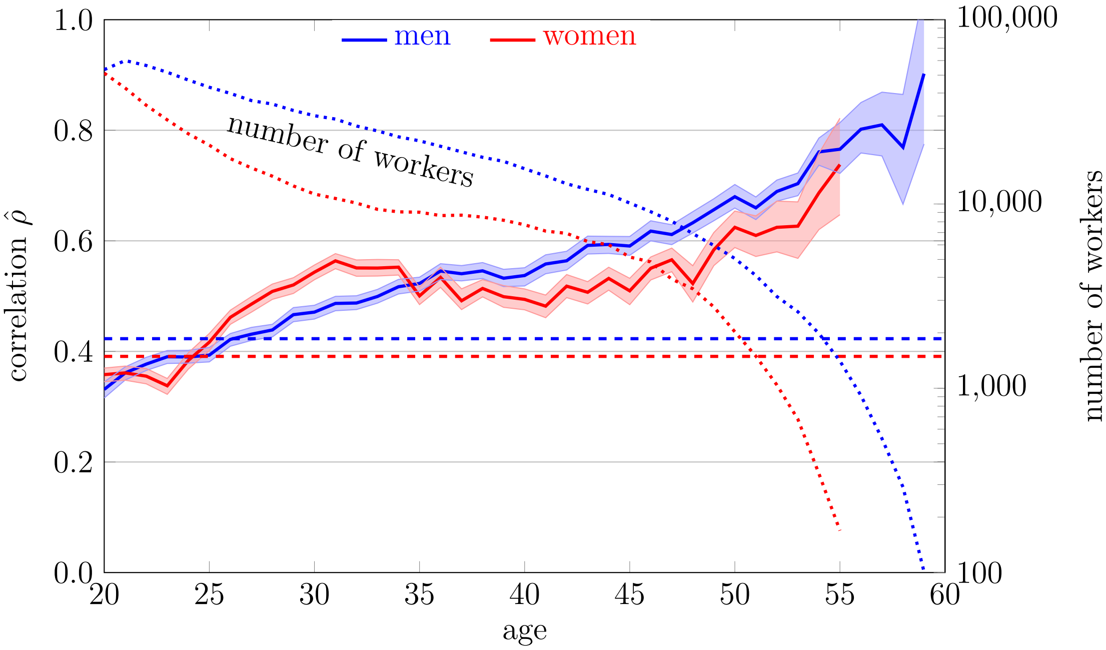

Figure 3: Correlation between worker and firm types by age, conditional on year fixed effects, using one job per employment spell. Solid lines are computed age-by-age and shaded ares are boostrapped 95 percent confidence intervals. For each age, the sample considers all workers who switched employers 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. We use the largest set of workers and jobs with multiple matches, not only those in the largest connected set. 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 where we observe at least 100 workers switching employers after an unemployment spell. Dashed lines are computed using the full sample, reported in column (5) of Tables 1 and 2.

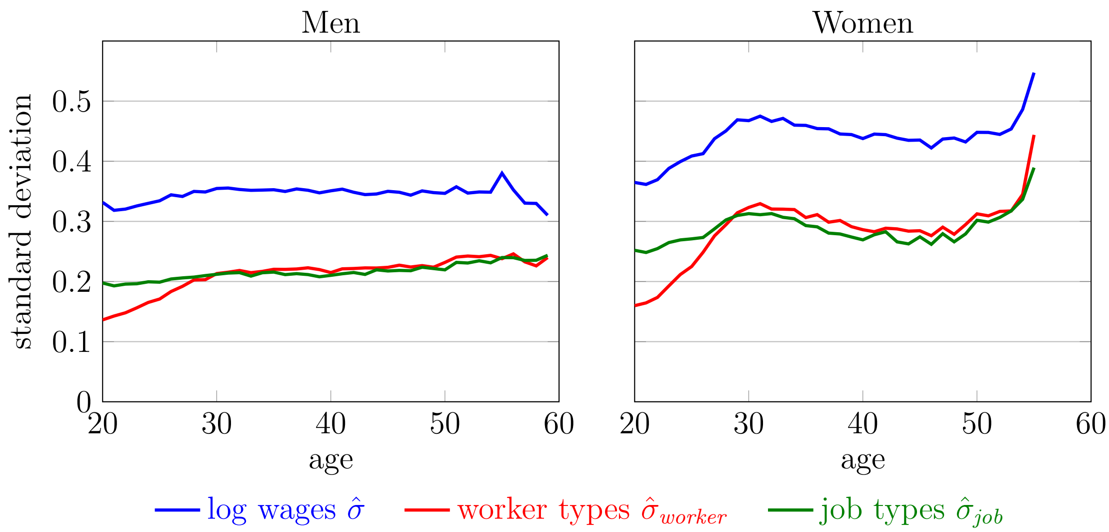

Figure 4: Standard deviations of log wages, worker types, and job types. Each line is computed age-by-age and uses one observation per employment spell and log wages conditional on year fixed effects. See the description of Figure 3 and the main text for more details.

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, particularly for men. The variance of worker 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. It is worth emphasizing then 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.

One interpretation of the 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). Of course, we need to augment this with a theory of sorting between workers and jobs to match all of the data.

A difficulty in interpreting this results is sample selection. The dotted line in Figure 3 shows the number of individuals who fit our sample criterion at each age. For men, this declines from 53,574 at age 20 to 101 at age 59. For women, the decline is from 51,291 at age 20 to 169 at age 55. We have already discussed the concern that individuals who lose their job, become unemployed, and return to work are unusual; compare columns (2) and (3) in Tables 1 and 2. Those workers who do this near the end of their career may be even more unusual.

To address this concern, we extend our approach to allow for individual fixed effects; see Appendix E for details. When we use individual fixed effects, we identify 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 primary analysis, i.e. who have two or more years when they work both before and after a registered unemployment spell. Only 38 percent of men and 27 women satisfy this additional selection criterion.

Although this subsample is even more selected than the sample depicted in Figure 3, we find that the life cycle pattern of the correlation is virtually the same as in the full sample in Figure 3. For example, in the full sample, a linear projection of the correlation on age gives a slope of 0.0100 per year for men between age 20 and 50. In the subsample with two or more spells, the slope is 0.0104 if we do not control for individual fixed effects. Controlling for individual fixed effects moderates the increase to 0.0063, so about forty percent of the increase in the correlation is accounted for by selection and the remainder represents a true life cycle component for these men.

The results for young women are somewhat different. While the correlation increases sharply for women under the age of 30 in our full sample and in the subsample with two or more spells, controlling for individual fixed effects reverses this. That is, life cycle patterns actually push down the correlation for younger women. For women over 30, however, controlling for individual fixed effects scarcely changes our results.

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,

$$\omega^w_{i,m} = x'_{i,m}\beta + \alpha_i + \psi_{k_{i,m}} + v_{i,m}, \tag{9}$$

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 3 and 4 (again for men and women) verify that this finding holds in our data as well. In column (1), we use a standard methodology to estimate the AKM correlation. This methodology uses a slightly different data set than ours. It treats each worker-firm-year wage as a distinct observation. It also only retains at most one job per year for each worker, the one with the highest earnings during the year. In terms of equation (9), this implies m is a year and $k_{i,m}$ is the employer that pays i the most money during year m. The AKM methodology delivers essentially zero correlation between the worker and firm fixed effects, 0.02 for men and −0.03 for women.[19]

In column (2), we collapse the data set to include only one observation per worker-firm match, the average log wage during the match. According to equation (9), the additional observations should not provide any additional information. On the other hand, we also include all jobs, not just the one with the highest annual earnings. This modestly increases the estimated correlation, to 0.05 for men and 0.03 for women. That reflects a combination of the additional jobs and a change in the weights that comes from treating a match rather than a year as the relevant observation.

Column (3) drops all workers and jobs with one observation, i.e. it uses the same data set as in column (1) of Tables 1 and 2. This raises the estimates for both men and women to 0.07, far below the numbers we estimate using our methodology. The remaining columns in Tables 3 and 4 correspond to the data sets and covariates used in the other columns of Tables 1 and 2, respectively. Using the fixed effects approach, the estimated correlation lies between −0.03 and 0.10. Across the six different specifications, the fixed effects correlation is about 0.39 below our estimate of the correlation for men and 0.34 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 already argued in Section 2.3 that the two measures are closely connected and indeed the two correlations should be the

[19]Gruetter and Lalive (2009) find a 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)	(8)	(9)
$\hat{\rho}$	—	—	0.476	0.530	0.438	0.416	0.423	0.384	0.418
$\hat{\rho}_{AKM}$	0.024	0.047	0.070	0.100	0.096	0.047	0.033	-0.030	0.040
$\hat{\rho} - \hat{\rho}_{AKM}$	—	—	0.406	0.430	0.342	0.369	0.390	0.414	0.378
number of workers (thousands)	4,128	4,138	2,810	2,065	1,100	1,100	1,100	1,100	1,100
number of firms (thousands)	659	750	498	372	301	234	234	234	234
number of observations (thousands)	52,873	17,707	16,129	10,572	7,577	4,376	4,376	4,376	4,376
share of observations top-coded	0.208	0.142	0.134	0.123	0.079	0.078	0.078	0.078	0.144
one observation per match	no	yes	yes	yes	yes	yes	yes	yes	yes
drop if $M_i = 1$ or $N_j = 1$	no	no	yes	yes	yes	yes	yes	yes	yes
first year of the sample	1972	1972	1972	1986	1986	1986	1986	1986	1986
one match per spell	no	no	no	no	no	yes	yes	yes	yes
year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
age dummies	no	no	no	no	no	no	yes	yes	no
polynomial in experience	no	no	no	no	no	no	no	yes	no

Table 3: Comparison of our estimates of correlation and AKM fixed effects estimates for men. Column (1) considers all workers and firms in the largest connected set. For each worker, the sample includes at most one job per year, the one which gives worker the highest earnings in that year. Column (2) includes only one observation per worker-firm pair, the average log wage during the match, and considers only workers and firms in the largest connected set. The samples in columns (3)–(9) correspond to the samples in columns (1)–(7) of Table 1; see notes under Table 1 and the main text for more details.

Comparison with AKM: Women

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\hat{\rho}$	—	—	0.375	0.401	0.405	0.392	0.391	0.390	0.389
$\hat{\rho}_{AKM}$	-0.030	0.030	0.065	0.068	0.102	0.040	0.037	0.016	0.037
$\hat{\rho} - \hat{\rho}_{AKM}$	—	—	0.309	0.333	0.304	0.352	0.354	0.374	0.352
number of workers (thousands)	3,376	3,386	2,358	1,767	951	951	951	951	951
number of firms (thousands)	738	821	522	385	302	238	238	238	238
number of observations (thousands)	39,911	12,429	11,101	7,665	5,206	3,190	3,190	3,190	3,190
share of observations top-coded	0.054	0.047	0.043	0.042	0.028	0.026	0.026	0.026	0.051
one observation per match	no	yes	yes	yes	yes	yes	yes	yes	yes
drop if $M_i = 1$ or $N_j = 1$	no	no	yes	yes	yes	yes	yes	yes	yes
first year of the sample	1972	1972	1972	1986	1986	1986	1986	1986	1986
one match per spell	no	no	no	no	no	yes	yes	yes	yes
year dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
age dummies	no	no	no	no	no	no	yes	yes	no
polynomial in experience	no	no	no	no	no	no	no	yes	no

Table 4: Comparison of our estimates of correlation and AKM fixed effects estimates for women. See description of Table 3 for more details on the sample construction.

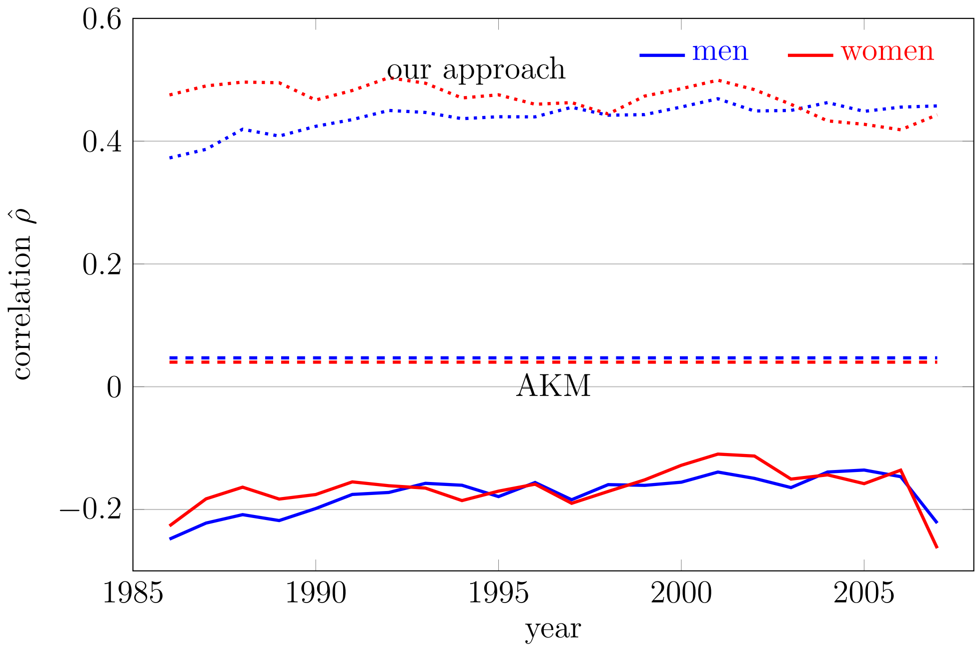

Figure 5: AKM correlation between worker and firm types, conditional on age and year fixed effects, using one job per spell. Solid lines are computed year-by-year. For each year, the sample considers all workers who switched employers after an unemployment spell within that year, and includes one job for each employment spell of these workers. The sample includes wage observation for that job from the considered year even if the job existed in other years. Dashed lines show the AKM correlation using the full sample, reported in column (6) of Tables 3 and 4. Dotted lines show the estimates using our approach, reported in column (4) of Tables 1 and 2.

same if the joint density of the AKM effects is elliptical (Proposition 1). We argue in the next subsection that much of the difference reflects an incidental parameter problem which causes a bias in the measurement of the AKM effects.

It is also worth noting another possible source of bias. The AKM approach is consistent in the limit as the number of observations per worker and firm goes to infinity only if 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, 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. Our approach does not require this orthogonality assumption.

5.2 Finite Sample Bias in Estimated Correlation

The most important difference between the AKM approach and ours is that the AKM approach only yields an unbiased estimate of the correlation between the fixed effects in the limit where we observe each worker at infinitely many different firms and each firm with infinitely many different workers (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.

The basic problem is that, although the estimates of the fixed effects are unbiased when each worker has a finite number of jobs and each firm has a finite number of employees, they are noisy. 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 (9) 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 (9), is correctly specified. Second, we assume that 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 (9), 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.[20] 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 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, then the correlation between λ and μ, our measure of type, is also equal to ρ; see Proposition 1.

We are interested in understanding what happens if estimate equation (9) given 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* (9) *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*

$$\begin{pmatrix} \sigma_\alpha^2 + \frac{N(M-1)\sigma_v^2}{M(MN-M-N)} & \rho\sigma_\alpha\sigma_\psi - \frac{\sigma_v^2}{MN-M-N} \\ \rho\sigma_\alpha\sigma_\psi - \frac{\sigma_v^2}{MN-M-N} & \sigma_\psi^2 + \frac{M(N-1)\sigma_v^2}{N(MN-M-N)} \end{pmatrix}. \tag{10}$$

[20]The matching graph is a set of nodes and links. Nodes represent workers and firms. There is a link between a firm and a worker node if the firm ever employed the worker.

If $\rho \geq 0$, then the correlation between $\hat{\alpha}$ and $\hat{\psi}$ in matched pairs is smaller than ρ, 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 strict exogenous. We call the difference $\hat{\alpha}_i - \alpha_i$ and $\hat{\psi}_j - \psi_j$ the AKM residuals. In terms of the variance-covariance matrix (10), 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 (10) converge to zero and the correlation between $\hat{\alpha}$ and $\hat{\psi}$ converges to ρ, 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_{worker}, the standard deviation of job types σ_{job}, and the correlation between worker and job types in matched pairs ρ. Under the assumption of an elliptical distribution of worker and firm types, the arguments in the proof of Proposition 1 imply

$$\sigma_\alpha = \frac{\sigma_{worker} - \rho\sigma_{job}}{1-\rho^2} \text{ and } \sigma_\psi = \frac{\sigma_{worker} - \rho\sigma_{job}}{1-\rho^2}.$$

Then the variance of the residual in equation (9) satisfies

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

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 columns (4)–(7) of Table 1, 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 (10) does a good job of predicting the AKM correlations in columns (6)–(9) of Tables 3 and 4. For example, the equation predicts an AKM correlation of 0.013 for men and -0.074 for women, compared to 0.047 and 0.040 in the data reported in in column (6).

On the other hand, using the mean numbers for M_i and N_j yields more modest biases, an AKM correlation of 0.230 for men and 0.168 for women. Interestingly, these are close to what we find on the artificial data sets that we use for our bootstrap procedure (see Appendix B). 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 0.229 for men and 0.164 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 in the noise 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. Our 95 percent confidence interval for the correlation is $[-0.0418, -0.0395]$. 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 0.5 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 0.5 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 two 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; and workers' types change over time, arguably more dramatically during a spell of registered unemployment (Ljungqvist and Sargent, 1998). A third reason our approach might understate the correlation is that firms are a collection 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 most production processes and hierarchies require a mix of skills (Garicano, 2000). One way to mitigate this concern might be to measure firm types within narrowly defined occupations. While our data set does not have occupational codes for most workers, they are available in some other countries' administrative data sets.

References

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

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

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

Andrews, Martyn J., Len 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.

_ , Leonard Gill, Thorsten Schank, and Richard Upward, "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.

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.

Kotlarski, Ignacy, "On Characterizing the Gamma and the Normal Distribution," *Pacific Journal of Mathematics*, 1967, *20* (1), 69–76.

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.

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 $\tau(\alpha_1, \psi) = \tau(\alpha_2, \psi)$. Using this, the conditional expected value satisfies

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

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, $\tau(\alpha, \psi) = \tau(2(\bar{\alpha} - \zeta(\psi - \bar{\psi})) - \alpha, \psi)$, 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.

The logic in the body of the paper then implies $\mu = \theta_0 + (1 + \theta_1)\psi$.

A symmetric proof implies that the expected value of ψ conditional on α is $\bar{\psi} + \frac{\rho\sigma_\psi}{\sigma_\alpha}(\alpha - \bar{\alpha}) = \kappa_0 + \kappa_1\alpha$ and hence $\lambda = \kappa_0 + (1 + \kappa_1)\alpha$. The common correlation coefficient follows immediately from linearity. ■

Proof of Proposition 3. The first order condition from minimizing the sum of squared residuals in equation (9) is equivalent to two moment conditions:

$$\hat{\alpha}_i = \frac{1}{M}\sum_{m=1}^{M}(\omega_{i,m}^w - \hat{\psi}_{k_{i,m}}) \text{ for all } i, \tag{11}$$

$$\hat{\psi}_j = \frac{1}{N}\sum_{n=1}^{N}(\omega_{j,n}^f - \hat{\alpha}_{h_{j,n}}) \text{ for all } j, \tag{12}$$

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.[21]

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

[21] 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.

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 (9) by 1 for the m^{th} job of some worker i. This directly affects the fixed effects estimates $\hat{\alpha}_i$ 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 (11) and (12), these satisfy

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

where $\Delta\alpha_1$ and $\Delta\psi_1$ denotes 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

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

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

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

The variance-covariance 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 (10) follows immediately.

Now if $\rho \geq 0$, 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. Alternatively, the covariance is positive but smaller than $\rho\sigma_\alpha\sigma_\psi$. 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. ■

B Bootstrap

B.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_{worker} and σ_{job}, 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 and the number of matches per firm N_j. We draw these from our estimates, e.g. in Tables 1 and 2.

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, the number firms a worker works for. For each $j \in \{1, \ldots, \tilde{J}\}$, we draw the number of employees N_j. We use the distribution of M and 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 will 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_{worker} (σ^2_{job}).

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 $\frac{\lambda_i \rho \sigma_{job}}{\sigma_{worker}}$ and variance $\sigma^2_{job}(1-\rho^2)$. 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. We find the largest connected set and keep only workers and firms in this set.

7. We measure correlation ρ_b using types λ and μ.

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

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

and the variance of the log wage shock satisfies

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

9. We estimate $\hat{\rho}_b$ using our approach (as described in the text) as well as $\hat{\rho}_{AKM,b}$ following AKM methodology.

10. 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_{worker} and σ_{job}.

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.

B.2 Properties of the Artificial Data

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

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

2. The expected value of $\chi_{i,m}^2$ conditional on λ_i is the conditional variance plus the square of the mean, $\sigma_{job}^2(1-\rho^2)+\frac{\lambda_i^2\rho^2\sigma_{job}^2}{\sigma_{worker}^2}$. Thus the unconditional expectation of $\chi_{i,m}^2$ is

$$\sigma_{job}^2(1-\rho^2)+\rho^2\sigma_{job}^2=\sigma_{job}^2.$$

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

3. The expected value of $\lambda\mu$ conditional on λ is $\lambda^2\rho\sigma_{job}/\sigma_{worker}$. Thus the unconditional expected value is

$$\rho\sigma_{job}\sigma_{worker}.$$

   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.

C 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 dataset. Starting from the wage cap as in the data, we decrease it gradually by 2 percent, 4 percent,... , and up to 50 percent. We then censor wages at the wage cap, construct data 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 men, the correlation varies very mildly, not dropping below 0.4 even when 65 percent of observations are top-coded.

Top-coding matters more for women. Setting the maximum wage to 50 percent of what it is in Austria increases the share of top-coded observations from 2.6 percent to 32 percent, and results in a decline of the correlation from 0.392 to 0.337.

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 men is similar and the correlation among high-wage women is stronger than the rest.

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 $\hat{\sigma}$, $\hat{\sigma}_{worker}$, and $\hat{\sigma}_{job}$ 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.

D Time-Varying Types

Consider the following variant of the model. Both workers' and firms' types changes over time, and hence across matches. The time-varying is indeed the expected wage at that point in time, $\omega^w_{i,m} = \lambda_{i,m} + \epsilon_{i,m}$ and $\omega^f_{j,n} = \mu_{j,n} + \eta_{j,n}$, where ϵ and η are independent, mean zero shocks with variances that possibly depend on the time-varying type.

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 shocks with fixed variances σ^2_υ and σ^2_ν, respectively. The cross-sectional distribution of λ and μ is invariant across matches, and so their variances satisfy $\sigma^2_\lambda = \sigma^2_\upsilon/(1-r^2)$ and $\sigma^2_\mu = \sigma^2_\nu/(1-s^2)$.[22]

[22] We can also reverse the order of matches and write $\lambda_{i,m} = r\lambda_{i,m+1} + \tilde{\upsilon}_{i,m}$ and similarly for μ.

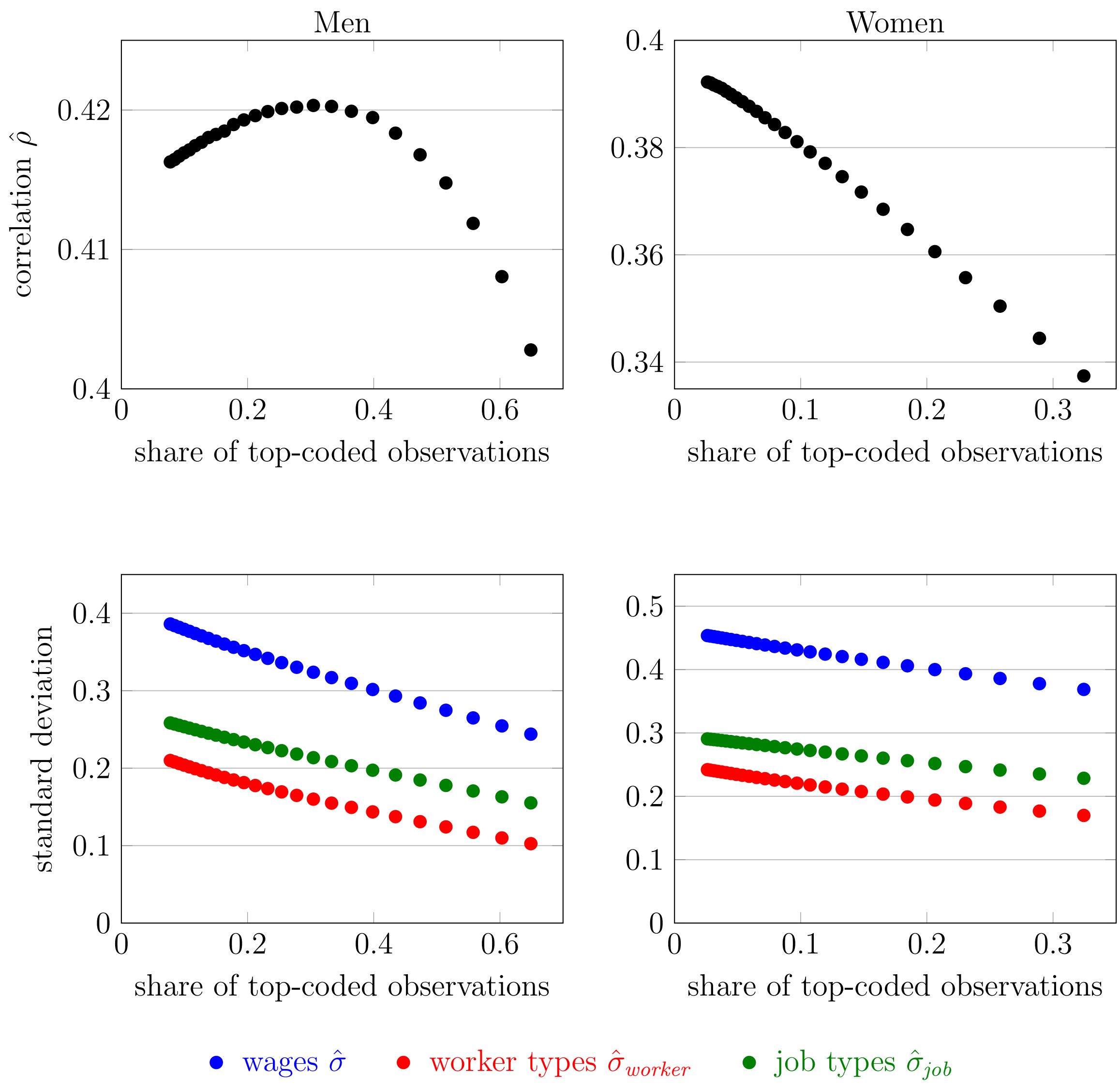

Figure 6: Impact of top-coding on estimated correlation and standard deviation of wages for men and women. Each dot corresponds to a sample where we decreased the top-code by $0, 2, 4, \ldots 50$ percent every year and truncated all wages at this new top-code. The sample of firms and workers corresponds to the sample in 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.

Moreover, if $k_{i,m} = j$ and $h_{j,n} = i$ (so i and j are matched together), the correlation between their types is ρ, as in the standard model.

We are interested in understanding what our estimator would measure in this environment. For algebraic simplicity, we assume all workers and firms have 2 matches. Then using the definition in Lemma 1 and a bit of algebra, we have

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

Similarly, $\sigma^2_{job} = s\sigma^2_\mu$. Finally, the covariance satisfies

$$\begin{aligned}
\Sigma &= \frac{1}{I}\int_0^1 \frac{\omega^w_{i,2}\omega^f_{k_{i,1},2}+\omega^w_{i,1}\omega^f_{k_{i,2},1}}{2}di - \left(\frac{1}{I}\int_0^I \frac{\omega^w_{i,1}+\omega^w_{i,2}}{2}di\right)\left(\frac{1}{I}\int_0^I \frac{\omega^f_{k_{i,1},2}+\omega^f_{k_{i,2},1}}{2}di\right) \\
&= \frac{1}{I}\int_0^1 (r\lambda_{i,1}+\upsilon_{i,2}+\upsilon_{i,2})(s\mu_{k_{i,1},1}+\nu_{k_{i,1},2}+\eta_{k_{i,1},2})di = rs\rho\sigma_\lambda\sigma_\mu.
\end{aligned}$$

The second equation simplifies the first term by recognizing the symmetry of the two matches (see footnote 22) and by dropping the second term, which is zero.

Combining these results, the estimated correlation would by $\Sigma/\sigma_{worker}\sigma_{job} = \rho\sqrt{rs} < \rho$. Thus to the extent that types vary over time, our approach underestimates the correlation between types at a point in time.

We believe 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.

E 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 $\hat{\Sigma}$ can be obtained as a coefficient from a linear regression. Consider the formula for $\hat{\Sigma}$ in Section 2.8. For each worker i in the sample, we can construct the summand,

$$\Sigma_i = \frac{1}{M_i}\sum_{m=1}^{M_i}\left(\frac{\sum_{m'\neq m}\omega^w_{i,m'}}{M_i-1} - \hat{\bar{w}}\right)\left(\frac{\sum_{n'\neq e_{i,m}}\omega^f_{k_{i,m},n'}}{N_{k_{i,m}}-1} - \hat{\bar{w}}\right),$$

where

$$\hat{\bar{w}} = \frac{1}{I} \sum_{i'=1}^{I} \frac{\sum_{m=1}^{M_{i'}} \omega_{i',m}^{w}}{M_{i'}},$$

and regress it on a constant. The coefficient in this regression is the desired $\hat{\Sigma}$. We can similarly obtain the variance of worker types by constructing

$$\sigma_{i,worker}^2 = \frac{1}{M_i} \sum_{m=1}^{M_i} \left(\omega_{i,m}^{w} - \hat{\bar{w}}\right)^2 - \frac{\sum_{m=1}^{M_i} \left(\omega_{i,m}^{w} - \frac{\sum_{m'=1}^{M_i} \omega_{i,m'}^{w}}{M_i}\right)^2}{M_i - 1}$$

and regressing this on a constant to obtain its average value.

We next use a linear regression to generate the results from our life cycle analysis. For the covariance, we construct $\Sigma_{i,a}$ 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 $\hat{\bar{w}}$. We then regress this on a full set of age dummy variables. Again, the coefficients are the average covariance, $\hat{\Sigma}$, 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.

Unfortunately, it is impossible to estimate the variance of job types controlling for worker fixed effects. This is 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 (controlling for worker fixed effects) and the age-specific standard deviation of job types (not controlling for worker fixed effects).

Figure 7 presents results for men (left panel) and women (right panel). We start with the same data set as we used in Section 4.4, pooling together workers of all ages. This gives us an initial sample of 465,399 men and 289,862 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 177,638 men and 78,465 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

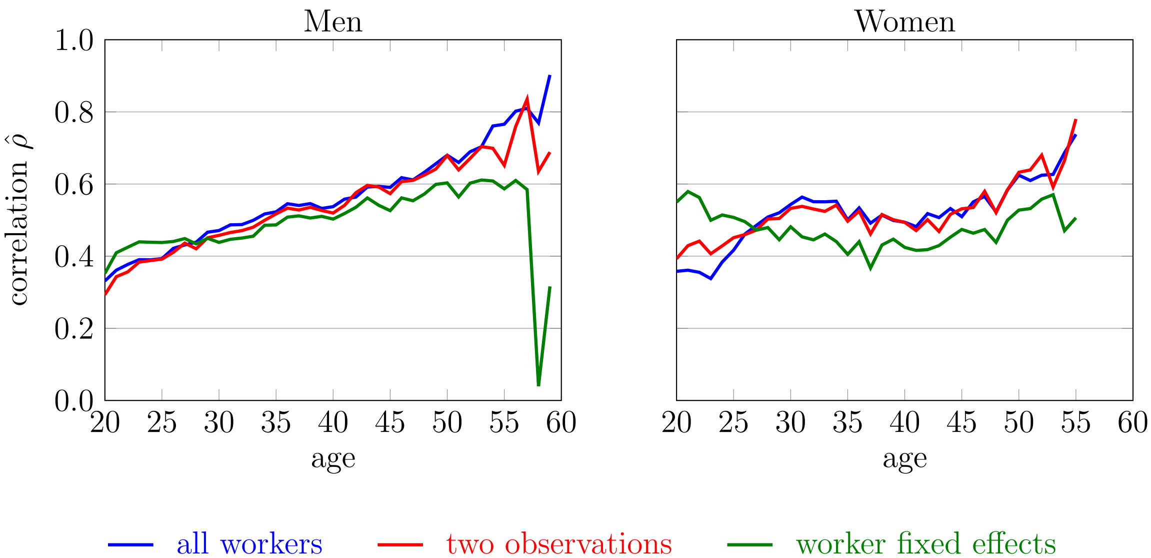

Figure 7: Correlation between worker and firm types by age, conditional on year fixed effects, using one job per employment spell. The blue lines are computed age-by-age using the sample from Section 4.4; see notes under Figure 3 and the main text for more details. The red lines restrict the sample to workers who appear at least twice in the life cycle analysis, that is, they are satisfy the relevant selection criteria at two or more ages. The green lines show the measure of correlation which controls for worker composition.

episodes. To address this, we look at age-specific correlations for this subset of workers, depicted in the red lines. The difference between the red and blue lines is small, mitigating our concern that this sample selection issue is important.

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.[23] Still, the figure shows a dramatic increase in the correlation, from around 0.4 for the youngest workers to 0.6 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.

[23] The estimated correlation drops dramatically at ages 58 and 59, but these results are based on very small samples.