Economics
Discussion Paper

EDP-0615

High wage workers and low wage firms: Negative assortative matching or statistical artefact?

M. Andrews, L. Gill, T. Schank and R. Upward

June 2006

Correspondence email: martyn.andrews@manchester.ac.uk

School of Social Sciences,
The University of Manchester
Oxford Road
Manchester   M13 9PL
United Kingdom

Combining the strengths of UMIST and
The Victoria University of Manchester

High wage workers and low wage firms: negative assortative matching or statistical artefact?[*]

M J Andrews

University of Manchester

L Gill

University of Manchester

T Schank

Universität Erlangen-Nürnberg

R Upward

University of Nottingham

Abstract

In the empirical literature on the estimation of firm and worker heterogeneity using linked employer-employee data, it appears that unobserved worker quality appears to be *negatively* correlated with unobserved firm quality. Following a suggestion made by Barth & Dale-Olsen (2003) and Abowd, Kramarz, Lengermann & Perez-Duarte (2004), we investigate the possibility that this is simply caused by standard estimation error. We develop formulae that show that the estimated correlation is biased downwards if there is true positive assortative matching and when any conditioning covariates are uncorrelated with the firm- and worker- fixed-effects. This result applies to any two-way (or higher) error-components model estimated by fixed-effects methods. We apply these bias corrections to a large German linked employer-employee dataset. We find that although the biases can be considerable, they are not sufficiently large to remove the negative correlation entirely.

Address for Correspondence:

Dr. M.J. Andrews
School of Social Sciences
University of Manchester
Manchester, M13 9PL

Email: `martyn.andrews@manchester.ac.uk`

1　Introduction

There is a rapidly-growing empirical literature which uses linked employer-employee data to estimate the contribution of worker and firm heterogeneity to outcomes in the labour market. Much of this literature stems from Abowd, Kramarz & Margolis (1999) (henceforth AKM) and related papers.[1] An important issue in the literature is the relationship between the unobserved worker- and firm-components of wages. Models of assignment imply positive assortative matching and therefore a positive correlation between worker and firm productivities. In the words of AKM: "high-wage workers and high-wage firms" match together.

However, a puzzle has emerged, in that the unobserved component of workers' wages appears to be *negatively* correlated with the unobserved component of firms' average wages. Apart from AKM's original study, which reported a positive correlation, all subsequent work has reported negative correlations. Abowd, Creecy & Kramarz (2002) report correlations of −0.28 for French data and −0.03 for data from Washington State, whereas Goux & Maurin (1999), using French different data, find a correlation ranging from +0.01 to −0.32 depending on the time period chosen.[2] Gruetter & Lalive (2004) report a correlation of −0.27 for Austrian data. All of these are weaker than Barth & Dale-Olsen's (2003) correlations of between −0.47 and −0.55 for Norwegian data. In other words, when focussing on unobserved components, low wage workers tend to work in high wage firms, and *vice versa*. This seems counter-intuitive in the light of theories of assortative matching.

There are two possible explanations for this emerging stylised fact. The first, suggested by Barth & Dale-Olsen (2003) and Abowd et al. (2004), is that the observed negative correlation is simply the result of using standard econometric techniques. Because the estimates of the worker and firm dummies are estimated with error, it is possible that the estimated correlation between them is biased downwards. It is not immediately obvious why this is so, but an over-estimate of a worker effect leads to, on average, to an under-estimate of a firm effect. The second explanation focuses on whether there any genuine economic explanations for why there might be negative assortative matching. Again, see Abowd et al. (2004).

In this paper, we focus on the first explanation. We derive formulae for the bias

---

[1] See also Abowd & Kramarz (1999) and Haltiwanger, Lane, Spletzer, Theeuwes & Troske (1999) for early surveys of the wide range of issues covered in this literature.

[2] AKM originally used an approximation to the LSDV estimator. Abowd et al. (2002) re-estimated these models using the exact solution they developed subsequently. This is why AKM's results look like outliers.

in the sampling distribution of the covariance between the unobserved worker- and firm-components of wages, and the biases in the variances of both components. When there are no conditioning covariates in the model, or when these covariates are not correlated with the worker and firm dummies, we show that the bias in the correlation is unambiguously negative when there is positive assortative matching. However, it is possible, but unlikely, that the bias can become positive when there is a strong correlation between the observed covariates and the worker and firm dummies. Subject to possible size constraints, the bias can be computed for any given dataset.

We also show that the extent of the bias depends on how much worker mobility each firm experiences, which itself depends on the key features of a given dataset. These include the length of the panel, the average size of firms (more generally, the firm-size distribution), and the error variance of the model. To analyse the impact of these features, we simulate a data generation process which creates an artificial linked employer-employee dataset which exhibits positive assortative matching; with this we estimate the parameters of the model using standard methods, and compute the biases using the formulae we have developed.

Ultimately, the size of the bias is an empirical issue, and should be computed for every application of linked employee-employer data. More importantly, this result applies to any two-way (or higher) error-components model estimated by fixed-effects methods. For example, an estimate of a true positive correlation between unobservably good schools and unobservably good pupils would be biased downwards.

Because it is possible that all of the negative estimates obtained thus far in the literature are consistent with positive assortative matching, we give an example using German linked data, from the Institut für Arbeitsmarkt– und Berufsforschung, Nürnberg (hereafter IAB).[3] It turns out that our bias correction moves the estimate of the correlation from -0.19 to -0.15, and so the econometric explanation—the statistical artefact of the title—is not sufficient to explain negative assortative matching on its own. We then find that the choice of sample is also important, namely whether small plants are excluded from the analysis and whether movers and analysed separately from non-movers. Then our bias-corrected estimate of the correlation is 0.23.

The structure of the paper is as follows. In Section 2 we outline the generic model

---

[3]Like us, Abowd et al. (2004) investigate the same issue, but conclude that the zero or negative correlation between person and firm effects is not explained by estimation biases due to lack of mobility in their data. This is probably because their data have many movers or because they assume that the true correlation is zero.

used in most of the literature and we explain the methods that are used to estimate the parameters of this model. In Section 3 we derive expressions for the biases in the correlation. In Section 4, we generate simulated data that are used to determine those features of the data which cause the bias; Section 5 presents the results of these simulations. In Section 6 we report what happens with our example using German linked data, and Section 7 concludes.

2   The generic model

Consider a model of wages with both employer and employee unobserved heterogeneity and employer and employee observed covariates:

$$ y_{it} = \mu + x_{it}\beta_1 + w_{jt}\beta_2 + u_i\eta + q_j\rho + \alpha_i + \phi_j + \varepsilon_{it}. \tag{1} $$

There are $i = 1, \ldots, N$ workers, $j = 1, \ldots, J$ firms and $t = 1 \ldots, T$ years. y_{it} is the dependent variable (in this case wages); x_{it} and u_i are vectors of observable i-level covariates; w_{jt} and q_j are vectors of observable j-level covariates.[4] α_i and ϕ_j are (scalar) unobserved heterogeneities. It is usual to assume that both are correlated with the observable components of wages. Models of positive assortative matching would also imply that they are positively correlated with each other. Note that both α_i and u_i are variables that are time-invariant for workers. Similarly, ϕ_j and q_j are fixed over time for firms. x_{it}, on the other hand, varies across i and t, and w_{jt} varies across j and t. Equation (1) therefore contains all four possible types of information which a researcher might have about workers and firms. There are K observed covariates in total.

Both workers and firms are assumed to enter and exit the panel, which means we have an unbalanced panel with T_i observations per worker. There are $N^* = \sum_{i=1}^{N} T_i$ observations (worker-years) in total. Individuals also change firms. This is crucial, as the parameters in fixed-effects models are identified by changers. Throughout we assume strict exogeneity, namely that

$$ \mathrm{E}(\varepsilon_{it}|x_{i1}, \ldots, x_{iT}, w_{j1}, \ldots, w_{jT}, \alpha_i, \phi_j) = 0. \tag{2} $$

This implies that workers' mobility decisions are independent of ε_{it}. However, it is

---

[4]A more precise notation would be to write $w_{j(it)t}$, where the function $j(it)$ maps worker i at time t to firm j. This emphasises the point that the unit of observation is an worker/year, but it is more cumbersome.

worth noting that mobility may be a function of the observables α_i and ϕ_j. Indeed, positive assortative matching *requires* that worker mobility is non-random with respect to α_i and ϕ_j.

As shown by AKM, in the presence of any correlations across the two sides of the market, that is correlations between unobserved/observed worker characteristics and unobserved/observed firm characteristics, there are omitted-variables biases which arise when estimating Equation (1) using data from only one side of the market. It is usual to assume that the heterogeneity terms α_i and ϕ_j are correlated with the observables from the same side of the market. This means that random effects methods are biased and inconsistent, and so fixed effects methods are needed to estimate the parameters of interest. This means that $[\eta, \rho]$, the parameter vector associated with the time-invariant variables, is not identified. Rather than dropping $[u_i, q_j]$, it is usual to define $\theta_i \equiv \alpha_i + u_i\eta$ and $\psi_j \equiv \phi_j + q_j\rho$ giving

$$y_{it} = \mu + x_{it}\beta_1 + w_{jt}\beta_2 + \theta_i + \psi_j + \varepsilon_{it}. \tag{3}$$

This is because estimates of $[\eta, \rho]$ can be recovered by making the additional random effects assumptions if the investigator so requires (as AKM do). Equation (3) is the generic model that represents most of the existing literature. The particular focus of this paper is on the estimation of the worker and firm fixed effects, θ_i and ψ_j, and their correlation with each other.

We now write the model in matrix notation:

$$y = Z\gamma + D\theta + F\psi + \varepsilon \tag{4}$$

where y and ε are $N^* \times 1$ vectors, D is a $N^* \times N$ matrix of worker dummies, F is a $N^* \times J$ matrix of firm dummies, and $Z = [X, W]$, where X represents worker covariates and W represents firm covariates. Z is a $N^* \times K$ matrix. θ is a $N \times 1$ parameter vector, ψ is a $J \times 1$ parameter vector, and γ is a $K \times 1$ parameter vector. Because one firm dummy is dropped, J is redefined accordingly, and note that Z does not contain a constant.

If one is not interested in the estimates of θ_i and ψ_j themselves, a consistent estimate of γ from Equation (4) is straightforward to obtain by time-demeaning within each unique worker-firm combination (or "spell"). This is because for each spell of a worker within a firm neither θ_i nor ψ_j vary, and so time-demeaning removes both terms. However, we are interested in the estimates of θ_i and ψ_j themselves, so this solution is not useful because it allows us to recover only the sum $\theta_i + \psi_j$ after

estimation, and not the worker components (see AKM). It is worth noting, however, that for many researchers this "spell fixed effects" (Spell FE) method is a practical and simple solution which does not present any computational difficulty.

As noted by AKM, the Least Squares Dummy Variable (LSDV) estimator of Equation (3) requires the estimation of N worker effects and (approximately) J firm effects. N is often in the order of millions, and J is often in the order of thousands, or tens of thousands. For most realistic values of N and J this is not a practical solution. In standard linear panel data models—that is, where the firm effects are absent—the LSDV estimator gives identical results to models where the heterogeneity is removed algebraically, by taking deviations from the mean of all variables in Equation (4). However, there appears to be no algebraic transformation of the observables that sweeps away both firm and worker effects, nor which allows them to be recovered subsequently. This is because of the lack of patterning between workers and their employers.[5]

To circumvent this problem, AKM note that explicitly including dummy variables for the firm heterogeneity, but sweeping out the worker heterogeneity algebraically, gives exactly the same solution as the LSDV estimator. In other words, Equation (4) is transformed by sweeping out the matrix of worker dummies D using $M_D \equiv I_{N^*} - D(D^T D)^{-1} D^T$:

$$M_D y = M_D Z \gamma + M_D F \psi + M_D \varepsilon. \tag{5}$$

In words, $y_{it} - \bar{y}_i$ is regressed on the vector of covariates $z_{it} - \bar{z}_i$ *and* on J mean-deviated firm dummies $F_{it}^j - \bar{F}_i^j$, where F_{it}^j is the j-th column of F, and $\bar{r}_i = (\sum_t r_{it})/T_i$ for any variable r.

To distinguish this estimator from the standard LSDV estimator, hereafter we label this estimator as "FEiLSDVj". They are identical estimators, but differ in how they are computed. The covariance matrix for FEiLSDVj needs the standard degrees-of-freedom adjustment.

To obtain estimates of the worker heterogeneity, note that

$$D\hat{\theta} = P_D y - P_D Z \hat{\gamma} - P_D F \hat{\psi}, \tag{6}$$

where $P_D \equiv D(D^T D)^{-1} D^T$. This equation gives the intuition as to why there is an observed negative correlation between $\hat{\theta}$ and $\hat{\psi}$ (as noted by Barth & Dale-Olsen (2003) and Abowd et al. (2004)). To see this, write out Equation (6) explicitly for

---

[5]More precisely, sort the data by workers, and the firm dummies are unpatterned; sort the data by firms, and the worker dummies are unpatterned.

each worker:

$$\hat{\theta}_i = \bar{y}_i - \overline{\hat{\psi}}_i - \bar{z}_i\hat{\gamma}, \tag{7}$$

where $\overline{\hat{\psi}}_i$ averages $\hat{\psi}_{j(it)}$ over t. As the ψ_j are estimated by LSDV, they are subject to the usual sampling variation (the firm dummies are no different from any other observed covariate). Once estimated, each $\hat{\psi}_j$ generates a number of $\hat{\theta}_i$, via Equation (7). If a ψ_j is over-estimated, then, on average, the corresponding θ_i are under-estimated, and *vice versa*. This implies that the estimated correlation between θ_i and ψ_j is biased downwards. An expression for this bias is formulated in the next section.

3 The bias

After estimation, one computes the sample variance over all N^* estimates of θ_i (N of which are distinct), the sample variance over all N^* estimates of ψ_j (J of which are distinct, if all are identified)[6], and the sample covariance between these two unobserved components:

$$S_{\hat{\psi}\hat{\psi}} = \frac{1}{N^* - 1} \sum_{it} (\hat{\psi}_j - \overline{\hat{\psi}})^2 = \frac{1}{N^* - 1} \hat{\psi}^T F^T A F \hat{\psi}, \tag{8}$$

$$S_{\hat{\theta}\hat{\theta}} = \frac{1}{N^* - 1} \sum_{it} (\hat{\theta}_i - \overline{\hat{\theta}})^2 = \frac{1}{N^* - 1} \hat{\theta}^T D^T A D \hat{\theta} \tag{9}$$

$$S_{\hat{\theta}\hat{\psi}} = \frac{1}{N^* - 1} \sum_{it} (\hat{\theta}_i - \overline{\hat{\theta}})(\hat{\psi}_j - \overline{\hat{\psi}}) = \frac{1}{N^* - 1} \hat{\theta}^T D^T A F \hat{\psi}. \tag{10}$$

The it-th element of the $N^* \times 1$ vector $D\hat{\theta}$ comprises $\hat{\theta}_i$ and the it-th element of the $N^* \times 1$ vector $F\hat{\psi}$ comprises $\hat{\psi}_j$. $\overline{\hat{\theta}}$ averages $\hat{\theta}_i$ over all of worker i's observations and similarly $\overline{\hat{\psi}}$ averages $\hat{\psi}_j$ over all of firm j's observations. Because these averages are non-zero, this gives rise to A in these expressions, where $A = I_{N^*} - \mathbf{1} \left(\mathbf{1}^T\mathbf{1}\right)^{-1} \mathbf{1}^T$ is the projection matrix producing mean deviations, and $\mathbf{1}$ is a $N^* \times 1$ vector of ones. We emphasise that each of $S_{\hat{\psi}\hat{\psi}}$, $S_{\hat{\theta}\hat{\theta}}$ and $S_{\hat{\psi}\hat{\theta}}$ is computed over N^* observations, that is, a given $\hat{\theta}_i$ is summed over T_i observations and a given $\hat{\psi}_j$ is summed over as many worker-periods the firm is observed in the data. These could be computed over N worker-level observations or J firm-level observations, but it seems sensible to use weighted averages, and so we do not develop these formulae here.

The vectors $\hat{\theta}$ and $\hat{\psi}$ suffer standard least-squares estimation error, and so we com-

---

[6]A ψ_j for a firm with no movement is not identified.

pare the means of the sampling distributions of $S_{\hat{\psi}\hat{\psi}}$, $S_{\hat{\theta}\hat{\theta}}$ and $S_{\hat{\psi}\hat{\theta}}$ with their respective true values $S_{\psi\psi}$, $S_{\theta\theta}$ and $S_{\psi\theta}$:

$$S_{\psi\psi} = \frac{1}{N^* - 1} \psi^T F^T A F \psi, \quad S_{\theta\theta} = \frac{1}{N^* - 1} \theta^T D^T A D \theta, \quad S_{\theta\psi} = \frac{1}{N^* - 1} \theta^T D^T A F \psi.$$

In the Appendix, we show that the resulting biases are as follows:

$$\text{Bias}[S_{\hat{\psi}\hat{\psi}}] = \frac{\sigma_\varepsilon^2}{N^* - 1} \text{tr} \left\{ F^T A F \left[F^T M_V F \right]^{-1} \right\} \tag{11}$$

$$\text{Bias}[S_{\hat{\theta}\hat{\theta}}] = \frac{\sigma_\varepsilon^2}{N^* - 1} \text{tr} \left\{ D^T A D \left[D^T M_{[Z,F]} D \right]^{-1} \right\} \tag{12}$$

$$\text{Bias}[S_{\hat{\theta}\hat{\psi}}] = -\frac{\sigma_\varepsilon^2}{N^* - 1} \text{tr} \left\{ F^T M_Z D \left[D^T M_Z D \right]^{-1} D^T A F \left[F^T M_V F \right]^{-1} \right\}. \tag{13}$$

where σ_ε^2 is the variance of ε_{it}, and $V = (Z, D)$.

We also show that, when the columns of Z are orthogonal to $[D, F]$, each trace can be unambiguously signed as positive. Thus, both $S_{\hat{\theta}}$ and $S_{\hat{\psi}}$ are overestimated whereas, as expected at the end of Section 2, $S_{\hat{\theta}\hat{\psi}}$ is underestimated. It is well-known that $S_{\hat{\psi}}$, in the absence of worker dummies, is biased upwards (Krueger & Summers 1988).[7] Our analysis here emphasises the downwards bias in the covariance. In other words, if the true covariance is positive, that is, there is positive assortative matching, the estimated correlation will always be too small, and could be negative. On the other hand, if the true covariance is negative, the estimated correlation could either be more or less negative.

It is difficult to make clear-cut predictions about what happens when the columns of Z are not orthogonal to $[D, F]$. However, as a particular column of Z becomes less orthogonal to $[D, F]$, loosely speaking, the smaller the bias becomes, but, at the same time, the influence of that variable becomes less important. Ultimately, the sign and the size of the bias is an empirical issue, using the formulae presented immediately above.

The estimated correlation between $\hat{\psi}$ and $\hat{\theta}$ is given by

$$R_{\hat{\theta}\hat{\psi}} = \frac{\hat{\theta}^T D^T A F \hat{\psi}}{\sqrt{\hat{\psi}^T F^T A F \hat{\psi}} \sqrt{\hat{\theta}^T D^T A D \hat{\theta}}}. \tag{14}$$

---

[7]Goux & Maurin (1999) give expressions for these biases, all of which are positive. This is because they use Spell FE, which is only an approximation to LSDV, and does not separately identify $\hat{\theta}$ and $\hat{\psi}$.

All three biases can be estimated, since each depends on only $\sigma_\varepsilon^2/(N^* - 1)$ and the data matrices X, D and F. Thus one can adjust the estimates of the three components by using estimates of the bias, and recompute the correlation. As already noted, linked employee-employer datasets can be very large. As the software has already computed $(F^T M_D F)^{-1}$ to produce LSDV estimates, the number of firms is not an issue. The only potential computational problem is that the expression for Bias$[S_{\hat{\theta}\hat{\theta}}]$ involves inverting the $N \times N$ matrix $[D^T M_{[Z,F]} D]$. The number of workers N might to be too large for the software at hand, in which case one has to assume that Z is orthogonal to D and F and use the formulae given in Appendix A.2. See also Appendix A.3 for further details on how to compute the biases.

We now need to establish some properties of these three bias terms, especially for the covariance. This is easier if one assumes there are no covariates Z. Intuition suggests that the three biases, in absolute terms, are a (complicated) decreasing function of the number of movers between firms, a property of the matrix $F^T D$, which appears a number of times in Equation (A.6). In particular, $F^T D$ is a $J \times N$ matrix that records the number of periods worker i is employed at firm j. In the next section, we use simulated data to show how large these biases might be for the type of datasets used in this literature. In particular, we attempt to uncover the non-linear relationship that links the bias in the correlation (or its components) to the number of movers and other features of these datasets, such as the number of firms and the number of time periods.

4   The simulation design

The simulated data mimics the generic model outlined in Section 2. J firms are created indexed $j = 1, \ldots, J$, each with a random number of employees N_j drawn from a Uniform distribution with mean μ_N. In this section, and the next, we have a balanced panel where each employee is observed for T periods. Each firm is given a realisation of w_{jt} and ψ_j; each worker is given a realisation of x_{it} and θ_i.[8] These realisations are drawn from a joint Normal distribution with the following means and covariance structure for any period t:

---

[8]We use one variable of each type, hence w_{jt} and x_{it} are scalars rather than vectors as in Equation (3).

$$
\begin{bmatrix} \psi_j \\ w_{jt} \\ \theta_i \\ x_{it} \end{bmatrix} \sim N \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} ; \begin{bmatrix} S_{\psi\psi} & & & \\ S_{w\psi} & S_{ww} & & \\ S_{\theta\psi} & S_{\theta w} & S_{\theta\theta} & \\ S_{x\psi} & 0 & S_{x\theta} & S_{xx} \end{bmatrix} \tag{15}
$$

The structure above focuses on the correlation between the unobservables and the observables, and the correlation between the unobservables themselves.[9] We assume that the observed firm and worker effects (w_{jt} and x_{it}) are uncorrelated with each other, but we allow for non-zero covariance between the unobserved components ($S_{\theta\psi} \neq 0$), as well as between the unobserved components and both firm and worker time-varying effects.

The draw of $[\psi_j, w_{jt}, \theta_i, x_{it}]$ initially ensures that workers with certain characteristics are matched with firms with certain characteristics. For example, if $S_{\theta\psi} > 0$ then high wage workers tend, on average, to be matched with high wage firms. This gives the distribution of workers across firms in period $t = 1$.

We now generate the movement of workers between firms. As noted, this is crucial for the identification of the fixed effects. For each worker we draw a potential new firm j' from the list of currently existing firms. This new firm has its own set of characteristics $[\psi_{j'}, w_{j't}]$.[10]

The probability of movement from j to j', denoted m_{it}^*, is determined by a random draw from a Normal distribution. A move occurs if m^* is greater than some critical percentile of the distribution of m^*, denoted m_c^*, such that the probability of movement $p \equiv \Pr(m^* > m_c^*)$ is set at 0.1. Altering p allows us to alter the number of workers who move each period. If a move occurs, the value of j' is copied to j in that period and for all future periods, as are $\psi_{j'}$, $q_{j'}$ and $w_{j't}$. The potential matching of workers and firms occurs once per period t. The number of periods T can be varied to mimic real data. Typically T is small because linked data are recorded annually, and have become available only recently.

It is important to emphasise that the assumption of random mobility is innocuous. So long as Equation (2) holds, any model of mobility will generate simulations with

___

[9]For clarity, we write out the correlation structure at time t. In addition, there are correlations across periods. Both variables x_{it} and w_{jt} are autocorrelated, with parameter 0.9. All x_{it} and w_{jt} pairs are uncorrelated.

[10]In order to ensure that a new match is drawn with a probability proportional to firm size, the list of new firms is weighted by the size of the firm.

similar properties. We choose random mobility because it means that we do not have to choose specific models about how movement occurs; for example, whether matches are "experience" or "search" goods.

Once the identity of each firm is established for every worker in all T rows of the data, the dependent variable y_{it} is generated according to Equation (3). As already noted, the resulting dataset is balanced for workers, unlike real data. It is not however necessarily balanced in terms of firms, because small firms who experience worker exits may disappear.

5 Simulation Results

5.1 Baseline simulation

We now repeatedly generate a synthetic dataset using the methods outlined in Section 4. Table 1 reports the baseline values chosen for the synthetic data and summarises the outcomes of the key parameters for 100 replications.

Table 1: Baseline parameter values and realisations: random mobility

	Population	Realisation (100 reps.)	
		Mean	s.d.
Number of firms J	100	100	–
Number of time periods T	5	5	–
Average number of workers per firm μ_N	50	50.401	1.683
Total number of observations N^*	25,000	24,907.55	1,594.87
Probability of movement per period p	0.1	0.1	–
Number of movers M		1997.18	138.04
Total number of groups G		1.66	0.844
Number of observations in largest group		24,902.05	1,595.29
Variance of idiosyncratic error σ_ε^2	1	1.001	0.0099
Parameter on x_{it}, β_1	0	0	–
Parameter on w_{jt}, β_2	0	0	–
Variance of worker effects $S_{\theta\theta}$	0.3	0.309	0.0087
Variance of firm effects $S_{\psi\psi}$	0.3	0.295	0.0490
Covariance firm and worker effects $S_{\theta\psi}$	0.0737	0.0730	0.0133
Correlation firm and worker effects $R_{\theta\psi}$	0.246	0.241	0.0244

Because the number of workers per firm, N_j, is drawn randomly from a Uniform

distribution with mean μ_N, this varies across simulations, as does the total number of workers who change firm each period, denoted M. The total number of observations, $T\sum_{j=1}^{J} N_j$, varies across simulations for the same reason, even though the number of firms remains fixed. (The population number of observations is $TJ\mu_N = 25,000$.) Each replication involves a completely new set of worker movements from firm to firm, and so the number of groups G (and hence the number of estimable effects) varies slightly between replications.[11] In fact, in about half the replications there is only one group (all workers and firms are connected); moreover, the size of the largest group is only slightly smaller than the total sample size. This is the usual finding in real linked data (Abowd et al. 2002). It is important to emphasise that in the base simulation, the parameters J, T, μ_N, p and σ_ε^2 are held fixed, but will vary when we make departures from the base simulation.

The crucial parameter is the correlation between θ and ψ, which is chosen to be positive ($R_{\theta\psi} = 0.246$): unobservably high wage workers work for unobservably high wage firms. We also assume positive correlation between each unobservable and both time-varying observables: the other four correlations in Equation (15) are $R_{\theta x} = 0.295$, $R_{\theta w} = 0.160$, $R_{\psi x} = 0.082$, and $R_{\psi w} = 0.299$. High wage workers work for firms with observably better characteristics, and high wage firms employ workers with observably better characteristics. The latter assumption is supported by much evidence from real linked employer-employee data.

For each dataset we estimate Equation (3) using FEiLSDVj. Note that we include x_{it} and w_{jt} in the regression, even though $\beta_1 = \beta_2 = 0$ in the data generation process. We then compute $\hat{\theta}$ using Equation (7), from which we compute $S_{\hat{\theta}\hat{\theta}}$, $S_{\hat{\psi}\hat{\psi}}$, $S_{\hat{\theta}\hat{\psi}}$, using Equations (8–10), and $R_{\hat{\theta}\hat{\psi}}$ using Equation (14). In Table 2 we report the baseline estimation results. The reader is reminded that Table 1 reports true simulated values of ψ and θ, whereas Table 2 reports estimated values $\hat{\psi}$ and $\hat{\theta}$.

First note, as expected, that the FEiLSDVj method produces unbiased estimates of $\beta_1 = 0$ and $\beta_2 = 0$: $\beta_1 = 0$ lies within two s.d.s about the mean value of $\widehat{\beta_1} = -0.00128$, and the same is true for $\beta_2 = 0$. However, the resulting estimate of the correlation of the worker and firm effects is significantly downwards biased; the 95% confidence interval about the mean estimate of 0.118 does not contain the true value of 0.246. This result illustrates the key finding of this paper. In fact, as shown algebraically in Section 3, all three components of the correlation are biased, when the observed covariates x_{it} and w_{jt} are absent from the model. The

---

[11]Identification of firm effects is only possible within a group, where a group is defined by the movement of workers between firms (Abowd et al. 2002).

Table 2: Baseline parameter estimates, 100 reps., random mobility

	Population	Simulation Mean	s.d.
Parameter on x_{it}, β_1	0	-0.00128	0.00822
Parameter on w_{jt}, β_2	0	0.00022	0.0074
Variance of worker effects $S_{\hat{\theta}\hat{\theta}}$	0.3	0.534	0.0148
Variance of firm effects $S_{\hat{\psi}\hat{\psi}}$	0.3	0.323	0.0572
Covariance firm and worker effects $S_{\hat{\theta}\hat{\psi}}$	0.0737	0.0492	0.0157
Correlation firm and worker effects $R_{\hat{\theta}\hat{\psi}}$	0.246	0.118	0.0317

variance of the estimated worker unobservables is almost twice as big as the variance of the true worker unobservables: $S_{\hat{\psi}\hat{\psi}} = 0.534$ whereas $S_{\psi\psi} = 0.3$. However, the variance of the estimated firm unobservables is not biased by much: $S_{\hat{\theta}\hat{\theta}} = 0.323$ whereas $S_{\theta\theta} = 0.3$. Finally, the covariance is biased downwards, thereby estimating a positive covariance too close to zero. Here $S_{\hat{\theta}\hat{\psi}} = 0.0492$, whereas $S_{\theta\psi} = 0.0737$. As we know from Section 3, these three biases, taken together, imply that a true positive correlation is always biased downwards. This is clearly being illustrated here.

5.2 Departures from the baseline simulation

We now vary the simulation in single dimensions away from the baseline. We then compute the three biases for each replication. Note that we use Equations (A.4–A.6) because we know that the true model does not contain observable covariates. This allows us to examine the cause of the bias, that is, estimation error in $\hat{\psi}$ and $\hat{\theta}$ in isolation of the estimation error in the parameters on the covariates. In what follows, we seek to quantify the extent of the bias as a function of the characteristics of particular data. In other words, we vary one of the parameters J, T, μ_N, p and σ_ε^2, but keep the others fixed.

Varying the probability of movement. The easiest way to illustrate the basic relationship between the bias in the covariance term, given in Equation (A.6), and the number of movers M is to vary the probability of a match dissolving (p). Simulations for three departures, for $p = 0.05$, $p = 0.15$ and $p = 0.20$ are plotted in Figure 1, together with the baseline replication $p = 0.10$. It is quite clear that the bias in the covariance tends to zero as the number of movers

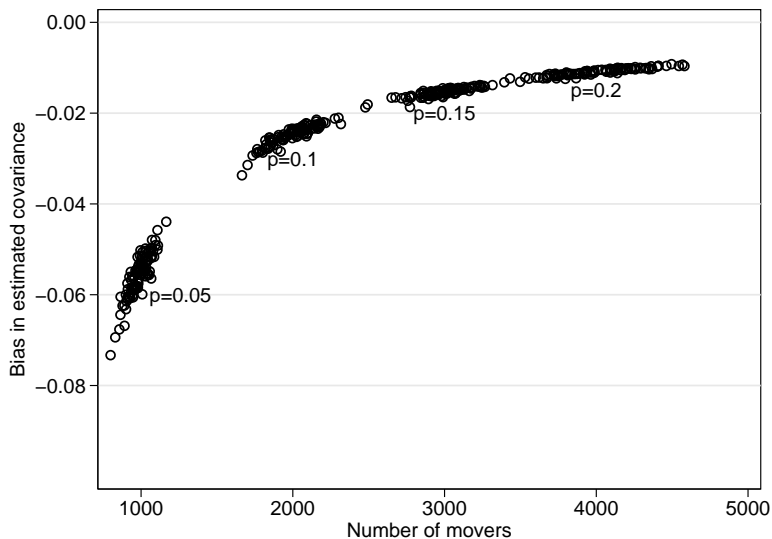endogenously increases. This basic result recurs throughout.

Figure 1: Varying p: bias in covariance

One cannot write down an algebraic expression for the bias in the correlation. All we can do is calculate the estimated correlation for each replication and subtract from it the true correlation. In Figure 2, we plot this difference against M for the same four sets of replications. The plot can be used to assess the probable bias in the correlation for a real dataset that has the same features as our simulated data. There is much more vertical variation in clusters compared with the first plot because the two variance terms in the denominator are also biased. In other words, for the *same* bias in the covariance, there are lots of possible biases in the product of the variances, each giving a different bias in the correlation.

Varying average firm size. Larger firms tend to have more workers joining and leaving them, and so varying μ_N provides another way of endogenously varying the number of movers M. We simulated various datasets for $\mu_N = 25, 50, 75$, with $p = 0.10$. All that happens is that each cluster of 100 replications lies on the curve plotted in Figure 1, with low values of μ_N located to the left (not reported). The same happens if we use different values of p.

Varying the number of time periods. The third dimension over which the number of movers can be endogenously increased is to lengthen the panel. The longer the panel, the more accurately ψ can be estimated because, once again, each firm has on average more movers, and so the bias in the covariance/correlation
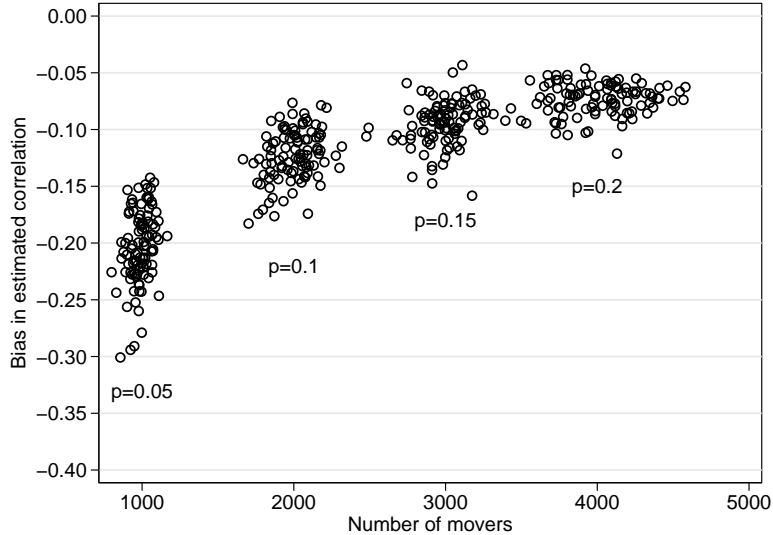
Figure 2: Varying p: bias in correlation

should lessen. Also, the bias in the correlation will get smaller as the bias in the two variances falls as T goes up. This is confirmed in Figure 3. We first re-plot the four clouds in Figure 1 (labelled $T = 5$). Below them are four clouds for which $p = 0.05, 0.10, 0.15, 0.20$, but now $T = 3$. One can see that holding p constant, but reducing T from 5 to 3, increases the bias in the estimated covariance (as predicted) by shifting the relationship in Figure 1 downwards. We finally plot a cluster $p = 0.1$ and $T = 7$. Thus the reader can see the effect of holding p constant at 0.10, and letting $T = 3$, $T = 5$ and $T = 7$. The number of movers gets bigger (there are more periods in which to move) and the bias gets smaller.

Varying the number of firms. In contrast, varying the number of firms for a given p has no effect on the bias of the estimated correlation (nor the true correlation). This is because every new firm requires a new estimated parameter ψ, and no improvement in sampling variability. Figure 4 illustrates this result, where one can see three clusters for $p = 0.10$, for $J = 50, 100, 150$, which lie to left and right of each other.

Varying overall error variance. In Figure 5 we illustrate the effect of increasing the overall error variance of Equation (3). As σ_ε^2 increases the sampling variability of $\hat{\psi}$ increases, which decreases the estimated correlation of ψ and θ, and therefore the absolute value of the bias increases. This plot is different from the others because altering σ_ε^2 has no effect on the number of movers.
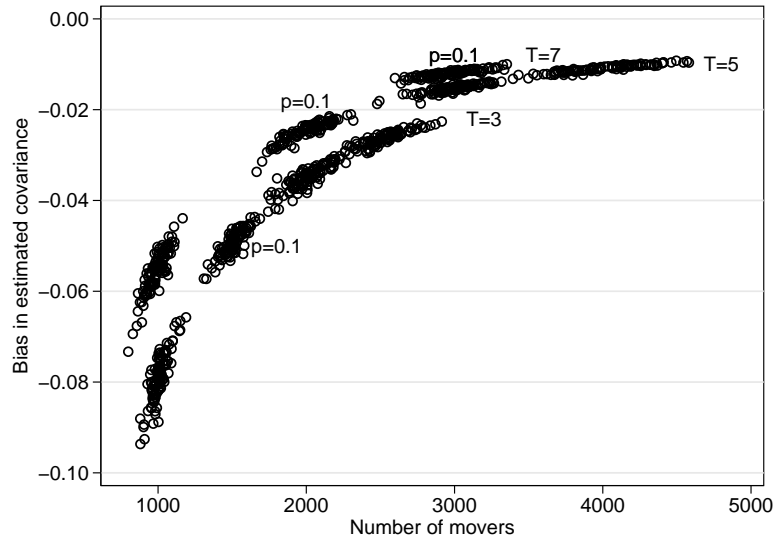
14

Figure 3: Varying T: bias in covariance
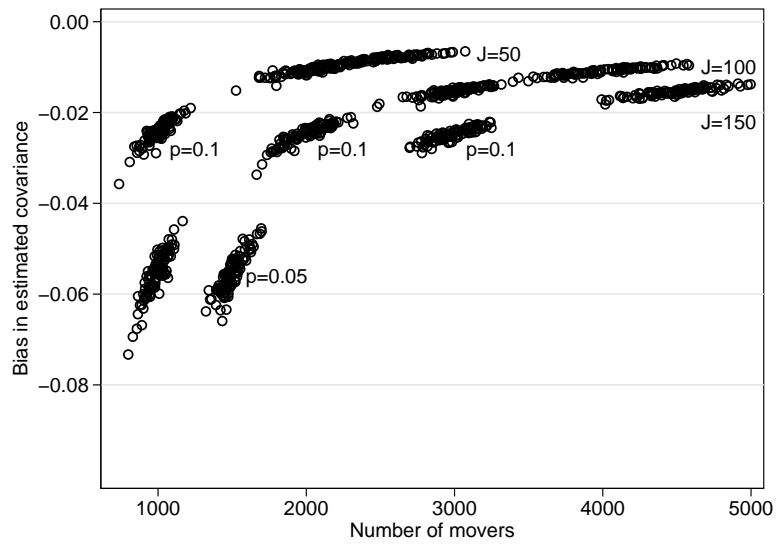
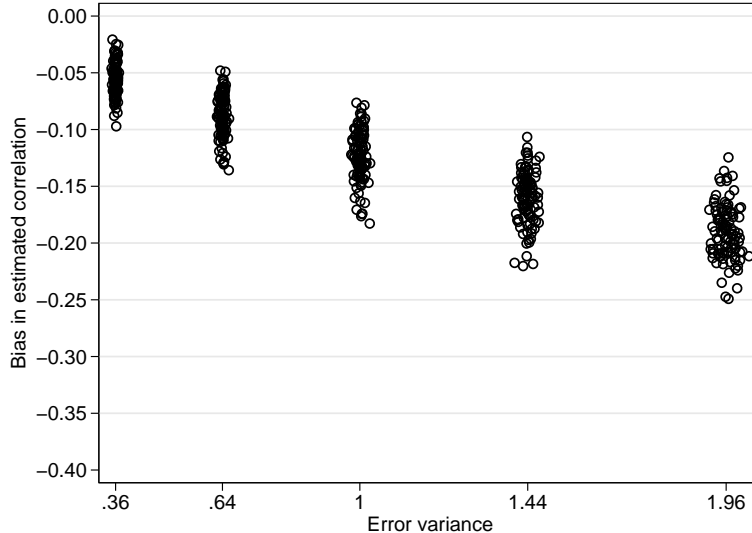Figure 4: Varying J: bias in covariance

15

Figure 5: Varying σ_ε^2: bias in correlation

5.3 Conclusions

The relationship between the bias in the covariance and the number of movers M is very clear, being negative and asymptoting towards zero as M increases. All combinations of μ_N and p lie on this same 'curve'. The curve shifts upwards towards zero as T increases, J decreases and σ_ε^2 decreases. The bias in the correlation, which is also affected by positive biases in both variances, shows the same basic pattern, but is much more affected by the noise in the data generation process, from simulation to simulation. Averaging over this noise, we can conclude that the bias in the correlation is decreasing in T, μ_N, and p, increasing in σ_ε^2, and is unaffected by J, because all of these parameters can be thought of as exogenously altering the number of movers in any given dataset.

Notice that Figures 2 and 5 show that the bias in the correlation can be quite substantial when the numbers of movers is relatively low. In fact, very occasionally in the simulations, $\text{Bias}[R_{\hat{\theta}\hat{\psi}}] < -0.246$, showing that it is possible for there to be negative estimates of the correlation even when there is positive assortative matching.

6   An example using German linked data

To illustrate how a downwards-biased estimate of $R_{\hat{\theta}\hat{\psi}}$ can be corrected, we use data from a linked worker-firm dataset made available by the IAB.[12] The firm data comprise a panel of 4,376 establishments (or "plants") from the former West Germany observed over the period 1993–1997. The worker data comprise a panel of 1,930,260 workers who are employed in these plants. A common establishment identifier is available in both datasets, allowing them to be linked.[13] After eliminating observations with missing or incomplete information, the resulting linked dataset has 5,145,098 worker-year observations (the *it* level). For each row in the data the identity j of the plant is recorded.

Firm effects are identified by the number of movers in each plant; most plants in the IAB data have few or no movers between other plants in the data. This is because the plant data is a survey, and because the dataset is relatively small in the T dimension. There are 1,821 plants (out of the total of 4,376) who have positive turnover.

Notice that N is approximately two million, and so we cannot compute the biases given in Equations (11–13) because of having to invert $D^T M_{[Z,F]} D$ in Equation (12). We therefore must assume that Z is orthogonal to D and F, and instead compute the biases given in Equations (A.4–A.6).

We estimate a standard earnings equation with $K = 53$ covariates, including marital status, age, education thresholds, occupation, union recognition, investment, concentration, plant size, age of plant, and profitability. Because we estimate Equation (3), not Equation (1), time-invariant covariates cannot be included (for example gender and industry). The model is estimated by a Classical Minimum Distance method that very closely approximates FEiLSDVj (see Andrews, Schank & Upward (2006) for further details and how the method is implemented in Stata). This model is reported in full in Andrews, Schank & Upward (2005); here we are only concerned with the estimated correlation between θ_i and ψ_j.

When the model is estimated with a full set of plant dummies, ie for the 1821 plants who have turnover, the estimated correlation between θ_i and ψ_j is -0.191 (see the first column of Table 3). This is consistent with the existing literature (see the Introduction). Applying the bias correction, the correlation moves to -0.148,

---

[12]Hereafter we refer to the data as LIAB: Linked IAB data.

[13]Kölling (2000) provides more information on the IAB establishment panel, Bender, Haas & Klose (2000) has details on the worker data and Alda, Bender & Gartner (2005) has details on the linked data.

Table 3: Bias correction, wage regressions, LIAB data

	All plants		High turnover plants	
	Whole sample	Movers sub-sample	Whole sample	Movers sub-sample
No. observations N^*	$4,883,331$	$72,253$	$5,145,098$	$62,668$
No. workers N	$1,816,368$	$23,393$	$1,930,260$	$20,313$
No. plants J	$1,821$	$1,821$	212	212
No. movers M	$23,393$	$23,393$	$20,313$	$20,313$
Error variance σ_ε^2	0.00459	0.00720	0.00461	0.00742
Uncorrected estimates				
Variance of worker effects $S_{\hat\theta\hat\theta}$	0.05381	0.05747	0.10231	0.20250
Variance of plant effects $S_{\hat\psi\hat\psi}$	0.01339	0.01513	0.00290	0.00562
Cov. plant/worker effects $S_{\hat\theta\hat\psi}$	-0.00512	-0.00389	-0.00030	0.00597
Corrn. plant/worker effects $R_{\hat\theta\hat\psi}$	-0.191	-0.132	-0.017	0.177
Correction to bias				
Bias$[S_{\hat\theta\hat\theta}]$ (Equation (A.5))	0.00320	0.00450	0.00180	0.00330
Bias$[S_{\hat\psi\hat\psi}]$ (Equation (A.4))	0.00149	0.00235	0.00008	0.00092
Bias$[S_{\hat\theta\hat\psi}]$ (Equation (A.6))	-0.00149	-0.00217	-0.00008	-0.00089
Corrected estimates				
Variance of worker effects $S_{\hat\theta\hat\theta}$	0.05061	0.05297	0.10050	0.19921
Variance of plant effects $S_{\hat\psi\hat\psi}$	0.01190	0.01278	0.00283	0.00470
Cov. plant/worker effects $S_{\hat\theta\hat\psi}$	-0.00363	-0.00171	-0.00022	0.00686
Corrn. plant/worker effects $R_{\hat\theta\hat\psi}$	-0.148	-0.066	-0.013	0.224
Correction to bias				
Bias$[R_{\hat\theta\hat\theta}]$	-0.043	-0.066	-0.004	-0.047

primarily because the covariance term moves from -0.00512 to -0.00363. Of the two explanations discussed in the Introduction, clearly the econometric explanation, on its own, does not explain why there is not positive assortative matching. Nonetheless, a 25% movement in the correlation represents a sizeable bias. The actual correction to the bias, namely 0.043, is given in the bottom row of the table.

This is the main message of the paper. However, we still need to investigate two modelling issues that recur in these analyses. The first concerns the size of the bias, and whether it can be ameliorated by pooling "small" plants into a single small "super plant". This often happens in the literature because the number of plants can be too many for the FEiLSDVj estimator. The second is whether we should model movers and non-movers separately.

One possible explanation for why there is a large bias is that the estimates of ψ_j are noisy for plants that experience low turnover. Equation (7) suggests that the more imprecise the estimates of ψ_j, the more biased is the correlation. Of the 1,821 plants who experience turnover, only 211 plants have 30 or more workers who move to other plants in the sample. In what follows, we group together all plants who have fewer than 30 movers into one super-plant, and estimate a model with just 212 identifiable plant effects.

When we re-estimate the model with only 212 plant effects (column 3), the estimated correlation increases to -0.017 and the bias-corrected estimate is -0.013. The absolute size of the bias in the estimated correlation therefore falls substantially from column 1 to column 3 (bottom row), which is what we would expect if the bias is caused by noisy estimates of ψ. However, there may be another reason for the fall in the bias, which is that we are restricting more than 3 million rows of the dataset (about 60% of the sample) to have the same value of $\hat{\psi}$. One should also note that in this case the restriction implied by moving from column 1 to column 3 is easily rejected (the standard F-test is 10.5).

The second issue that recurs with any type of fixed-effects model is that the sub-sample of movers (who effectively identify the parameters of the model) may be a non-random sub-sample. Workers and plants who choose to separate for whatever reason are not necessarily the same as those worker-pairs who tend to stay together. In particular, the correlation of worker and firm effects may not be the same for movers and non-movers. In column (2) we therefore report estimates separately for movers. That is, we use the 72,253 observations for those workers who move between the 1,821 plants. An F-test of parameter equality between movers and non-movers sub-samples rejects the null hypothesis easily (p-value zero). There is

also evidence that movers have a different degree of assortative matching than non-movers. The bias-corrected correlation of plant and worker effects increases from -0.148 to -0.066.

Since the separation of movers and non-movers appears to be important (column 2), and since the pooling of low-turnover plants also reduces the bias (column 3), it seems logical to look at the results for movers in high turnover plants (column 4). When we do this we actually estimate a positive correlation of plant and worker effects (0.224, bias corrected). As before, the pooling of low-turnover plants reduces the size of the bias (compare column 4 with column 2). However, once again, we reject the implied restriction (the standard F-test is 6.6). And also, as before, the correlation for movers is larger than for the whole sample (compare column 4 with column 3).

The lesson from all this is that estimates of the correlation of worker and plant effects are sensitive to modelling decisions as well as the statistical bias highlighted in Sections 3 and 4. The bias may be as large as 50% of the size of the uncorrected correlation. But in our example, looking at movers and non-movers separately, resulted in even larger movements in the correlation. Finally, our preferred estimate of the correlation of -0.066 (column 2) is still negative though somewhat closer to zero than others in the literature. This estimate is much closer to zero than our uncorrected estimate (-0.191), partly because of the bias correction, and partly because the correlation is less negative for movers than non-movers.

7   Conclusion

In this paper, we show that estimates of the correlation between firm- and worker-fixed-effects are biased downwards if there is true positive assortative matching and when any conditioning covariates are uncorrelated with the firm- and worker- fixed-effects. We develop formulae for the biases for the components of the estimated correlation. Ultimately, the size of the bias is an empirical issue, and should be computed for every application of linked employee-employer data. More importantly, this result applies to any two-way (or higher) error-components model estimated by fixed-effects methods.

Using simulations, we show that the extent of the bias depends on how much worker mobility each firm experiences, which itself depends on the propensity to move, the length of the panel, the average size of firms (more generally, the firm-size

distribution) and the error variance of the model. It is, however, unaffected by the number of firms. We apply these bias corrections to a large German linked employer-employee dataset. We find that although the biases can be considerable, they are not sufficiently large to remove the negative correlation entirely. We also show that modelling choices regarding the separation of movers and non-movers and the grouping of small plants can have significant impacts on the estimated correlation.

Appendix A   Algebraic details

A.1   Deriving the three biases

From Equation (4) of the main text,

$$y = Z\gamma + D\theta + F\psi + \varepsilon,$$

by combining Z and D into a matrix V, the Frisch-Waugh (FW hereafter) argument can be used to calculate $\hat{\psi}$ and $\hat{\theta}$ as

$$\hat{\psi} = \left[F^T M_V F\right]^{-1} F^T M_V y,$$

$$\hat{\theta} = \left[D^T M_Z D\right]^{-1} D^T M_Z (y - F\hat{\psi}).$$

The sampling errors of $\hat{\theta}$ and $\hat{\psi}$ are calculated from substitution of y:

$$\hat{\psi} = \psi + \left[F^T M_V F\right]^{-1} F^T M_V \varepsilon, \tag{A.1}$$

and similarly,

$$\hat{\theta} = \theta + \left[D^T M_Z D\right]^{-1} D^T M_Z \left[\varepsilon - F(\hat{\psi} - \psi)\right]. \tag{A.2}$$

Using an alternative organisation of the regressors, the FW argument also gives:

$$\hat{\theta} = \left[D^T M_{[Z,F]} D\right]^{-1} D^T M_{[Z,F]} y$$
$$= \theta + \left[D^T M_{[Z,F]} D\right]^{-1} D^T M_{[Z,F]} \varepsilon.$$

The sample variance of the elements of $\hat{\psi}$ is

$$S_{\hat{\psi}\hat{\psi}} = \frac{1}{N^* - 1} \hat{\psi}^T F^T A F \hat{\psi},$$

where $A = I_{N^*} - \mathbf{1} \left(\mathbf{1}^T \mathbf{1}\right)^{-1} \mathbf{1}^T$ is the projection matrix producing mean deviations, and $\mathbf{1}$ is a $N^* \times 1$ vector of ones.

The expected value of $S_{\hat{\psi}\hat{\psi}}$ is

$$
\begin{aligned}
E[S_{\hat{\psi}\hat{\psi}}] &= \frac{1}{N^* - 1} \left\{\psi + \left[F^T M_V F\right]^{-1} F^T M_V \varepsilon\right\}^T F^T A F \left\{\psi + \left[F^T M_V F\right]^{-1} F^T M_V \varepsilon\right\} \\
&= \frac{1}{N^* - 1} \left(\psi^T F^T A F \psi + E\left\{\varepsilon^T M_V F \left[F^T M_V F\right]^{-1} F^T A F \left[F^T M_V F\right]^{-1} F^T M_V \varepsilon\right\}\right) \\
&= \frac{1}{N^* - 1} \left(\psi^T F^T A F \psi + \sigma_\varepsilon^2 \operatorname{tr}\left\{M_V F \left[F^T M_V F\right]^{-1} F^T A F \left[F^T M_V F\right]^{-1} F^T M_V\right\}\right) \\
&= \frac{1}{N^* - 1} \left(\psi^T F^T A F \psi + \sigma_\varepsilon^2 \operatorname{tr}\left\{F^T A F \left[F^T M_V F\right]^{-1}\right\}\right) \\
&= S_\psi + \frac{\sigma_\varepsilon^2}{N^* - 1} \operatorname{tr}\left\{F^T A F \left[F^T M_V F\right]^{-1}\right\}.
\end{aligned}
$$

The penultimate line comes from the cyclical property of traces.

Thus the bias in estimating $S_{\hat{\psi}\hat{\psi}}$ is

$$
\operatorname{Bias}[S_{\hat{\psi}\hat{\psi}}] = \frac{\sigma_\varepsilon^2}{N^* - 1} \operatorname{tr}\left\{F^T A F \left[F^T M_V F\right]^{-1}\right\}
$$

This is Equation (11) of the main text. Because both A and M_V are positive semidefinite, the matrices $F^T A F$ and $F^T M_V F$ are positive semi-definite, and will be positive definite in practice. The result that $\operatorname{tr}[AB^{-1}] > 0$ if both A and B are positive definite completes the proof that the bias is unambiguously positive.[14] This is because each ψ_j is estimated with error, the square of which is added into the expression for the variance.

For $\hat{\theta}$, the sample variance is

$$
S_{\hat{\theta}\hat{\theta}} = \frac{1}{N^* - 1} \hat{\theta}^T D^T A D \hat{\theta} \tag{A.3}
$$

Using the symmetry between D and F in Equation (3) of the main text,

$$
\operatorname{Bias}[S_{\hat{\theta}\hat{\theta}}] = \frac{\sigma_\varepsilon^2}{N^* - 1} \operatorname{tr}\left\{D^T A D \left[D^T M_{[Z,F]} D\right]^{-1}\right\}.
$$

This is Equation (12) of the main text. Again, this bias is unambiguously positive, for the same reasons as for $\operatorname{Bias}[S_{\hat{\psi}\hat{\psi}}]$.

---

[14]If A and B are symmetric and psd, write both matrices in terms of their symmetric positive square root matrices: $\operatorname{tr}(AB) = \operatorname{tr}[(A^{1/2} A^{1/2})(B^{1/2} B^{1/2})] = \operatorname{tr}[(A^{1/2} B^{1/2})(B^{1/2} A^{1/2})] = \operatorname{tr}(C^T C) \geq 0$.

For the sample covariance between $\hat{\theta}$ and $\hat{\psi}$, we get

$$E[S_{\hat{\theta}\hat{\psi}}] = S_{\theta\psi} + \frac{\sigma_\varepsilon^2}{N^* - 1} \text{tr}\left\{ M_{[Z,F]} D \left[D^T M_{[Z,F]} D \right]^{-1} D^T A F \left[F^T M_V F \right]^{-1} F^T M_V \right\}.$$

There is a "well-known" projection identity (Baltagi 2005, Eqns(9.29, 9.30)) which says that

$$P_V = P_Z + M_Z D \left[D^T M_Z D \right]^{-1} D^T M_Z$$

which translates into an identity for M_V :

$$M_V = M_Z - M_Z D \left[D^T M_Z D \right]^{-1} D^T M_Z,$$

with a corresponding result for

$$M_{[Z,F]} = M_Z - M_Z F \left[F^T M_Z F \right]^{-1} F^T M_Z.$$

Also note that

$$
\begin{aligned}
M_V M_{[Z,F]} &= \left\{ M_Z - M_Z D \left[D^T M_Z D \right]^{-1} D^T M_Z \right\} M_{[Z,F]} \\
&= M_{[Z,F]} - M_Z D \left[D^T M_Z D \right]^{-1} D^T M_{[Z,F]}
\end{aligned}
$$

since $M_Z M_{[Z,F]} = M_{[Z,F]}$.

Plugging this result into the bias expression for $E[S_{\hat{\theta}\hat{\psi}}]$, one obtains

$$
\begin{aligned}
&\text{tr}\left\{ D^T A F \left[F^T M_V F \right]^{-1} F^T M_V M_{[Z,F]} D \left[D^T M_{[Z,F]} D \right]^{-1} \right\} \\
=\ &\text{tr}\left\{ D^T A F \left[F^T M_V F \right]^{-1} F^T M_{[Z,F]} D \left[D^T M_{[Z,F]} D \right]^{-1} \right. \\
&\left. - D^T A F \left[F^T M_V F \right]^{-1} F^T M_Z D \left[D^T M_Z D \right]^{-1} D^T M_{[Z,F]} D \left[D^T M_{[Z,F]} D \right]^{-1} \right\} \\
=\ &-\text{tr}\left\{ D^T A F \left[F^T M_V F \right]^{-1} F^T M_Z D \left[D^T M_Z D \right]^{-1} \right\}
\end{aligned}
$$

since $F^T M_{[Z,F]} = 0$. This is Equation (13) of the main text.

Signing the trace requires that $G^T A \equiv M_Z D \left[D^T M_Z D \right]^{-1} D^T A$ is symmetric. In general this is not so, but is symmetric when Z is absent, or when Z is orthogonal to D and F. In this case, it can shown that the trace is positive, an algebraic proof of which is given in the following subsection. It can be shown, using numerical examples, that the less orthogonal each column of Z is to D, F, (in the sense that a regression of a column of Z on D, F has a higher R^2), then the less symmetric $G^T A$ becomes, and the less positive the trace becomes. In pathological cases, the

trace can become negative, for example when a column of Z is collinear with D, F; however, this case is no interest as D, F contain no extra information, and so the parameter on this column of Z would not be identified.

In short, the essence of why there is a negative bias between worker and firm unobservables is seen in the case where Z is orthogonal to D and F, or, equivalently, when Z is absent. However, given that this is not going to happen in practice, ultimately computing the bias is an empirical issue, using the formulae presented immediately above.

A.2   What happens with no Zs?

When Z is absent, substitute $M_Z = I_{N^*}$ and $M_V = M_D$:

$$
\text{Bias}[S_{\hat{\psi}\hat{\psi}}] = \frac{\sigma_\varepsilon^2}{N^* - 1} \operatorname{tr}\left\{ F^T A F \left[F^T M_D F \right]^{-1} \right\}. \tag{A.4}
$$

$$
\text{Bias}[S_{\hat{\theta}\hat{\theta}}] = \frac{\sigma_\varepsilon^2}{N^* - 1} \left(\operatorname{tr}\{P_D A\} + \operatorname{tr}\left\{ F^T P_D A P_D F \left[F^T M_D F \right]^{-1} \right\} \right)
$$

$$
= \frac{\sigma_\varepsilon^2}{N^* - 1} \left(N - 1 + \operatorname{tr}\left\{ F^T A P_D F \left[F^T M_D F \right]^{-1} \right\} \right). \tag{A.5}
$$

$$
\text{Bias}[S_{\hat{\theta}\hat{\psi}}] = -\frac{\sigma_\varepsilon^2}{N^* - 1} \operatorname{tr}\left\{ F^T D \left[D^T D \right]^{-1} D^T A F \left[F^T M_D F \right]^{-1} \right\}
$$

$$
= -\frac{\sigma_\varepsilon^2}{N^* - 1} \operatorname{tr}\left\{ F^T P_D A F \left[F^T M_D F \right]^{-1} \right\}. \tag{A.6}
$$

The first line of (A.5) comes from substituting (A.1) into (A.2), and using $M_V = M_D$ and $D^T M_Z = D^T$:

$$
\hat{\theta} = \theta + \left[D^T D \right]^{-1} D^T \left(I_{N^*} - F \left[F^T M_D F \right]^{-1} F^T M_D \right) \varepsilon.
$$

Substituting into (A.3), and taking expectations as with $E[S_{\hat{\theta}\hat{\theta}}]$ above gives the expression shown in the second line of (A.5).

Exactly the same results occur when Z orthogonal to D and F.

To show that the trace in Equation (A.6) is positive requires showing that the double projection matrix $P_D A$ is positive semi-definite. Three properties of the matrix D are used: (a) that the rows sum to unity, $D\mathbf{i} = \mathbf{1}$; (b) that the columns sum to T and (c) that $D^T D = \operatorname{diag}\{T_i\}$. (b) and (c) imply that

$$
(D^T D)^{-1} D^T \mathbf{1} = \mathbf{i}
$$

where $\mathbf{i}$ is a $N \times 1$ vector of ones. Hence

$$
\begin{aligned}
P_D A &= P_D(I_{N^*} - \mathbf{1}\left(\mathbf{1}^T\mathbf{1}\right)^{-1}\mathbf{1}^T) \\
&= P_D - D(D^T D)^{-1} D^T \mathbf{1}\left(\mathbf{1}^T\mathbf{1}\right)^{-1}\mathbf{1}^T \\
&= P_D - D\mathbf{i}\left(\mathbf{1}^T\mathbf{1}\right)^{-1}\mathbf{1}^T \quad \text{using (b,c)} \\
&= P_D - \mathbf{1}\left(\mathbf{1}^T\mathbf{1}\right)^{-1}\mathbf{1}^T \quad \text{using (a)} \\
&= P_D - P_{\mathbf{1}}
\end{aligned}
$$

Hence $P_D A$ is psd, with $\text{trace}(P_D A) = \text{trace}(P_D) - \text{trace}(P_{\mathbf{1}}) = N - 1$.

Also note that $F^T M_D F$ is a strictly positive definite matrix except when $M_D F = 0$. This can only occur when there is no movement between firms, in which the firm dummy effects ψ cannot be identified.

Without the covariates, the intuition as to why there is a negative bias can be easily seen. Substitute $M_Z = I_{N^*}$ into Equation (A.2), and writing out for worker i:

$$
\hat{\theta}_i - \theta_i = -(\overline{\hat{\psi}}_i - \overline{\psi}_i) + \overline{\varepsilon}_i,
$$

where $\overline{\hat{\psi}}_i$ averages $\hat{\psi}_{j(it)}$ over t, where $\overline{\psi}_i$ averages $\psi_{j(it)}$ over t, and $\overline{\varepsilon}_i$ averages ε_{it} over t. This is the equation that shows that, on average, an under-estimate of ψ_j leads to an over-estimate of θ_i, and vice versa. This is the cause of the downwards bias that this paper seeks to establish.

Finally note that $S_{\hat{\theta}}$ is over-estimated for the same reason that $S_{\hat{\theta}\hat{\psi}}$ is underestimated, as both have the same bias term. There is an extra bias term in $S_{\hat{\theta}}$, which is $\approx (N/N^*)\sigma_\varepsilon^2$, or σ_ε^2/T in a balanced panel. This term comes about because each worker-effect θ_i is estimated on T observations, a bias effect that disappears as T goes to infinity. All the other bias terms disappear as N^* goes to infinity. It is this term that Krueger & Summers (1988) and Haisken-DeNew & Schmidt (1997) use to adjust the variance of the estimates of a set of industry dummies in their analysis of inter-industry wage differentials. If one drops $D\theta$ from Equation (4), it is easy to show, using the same properties of D above applied to F, that the bias of $S_{\hat{\psi}}$ is $\sigma_\varepsilon^2\frac{J-1}{N^*-1}$. This is the same as Haisken-Denew & Schmidt, who use a different parameterisation.

A.3 Computation and size constraints

Linked employee-employer datasets can be very large. N is often in the order of millions, and J is often in the order of thousands, or tens of thousands. In the text, where we discuss estimation of the generic model, it is assumed that the software can invert $J \times J$ matrices, but not $N \times N$ or $N^* \times N^*$ matrices. With these constraints, one cannot compute $\text{Bias}[S_{\hat{\theta}\hat{\theta}}]$ in Equation (12), which requires inverting $\left[D^T M_{[Z,F]} D \right]$. As with the existing literature (Krueger & Summers 1988, Haisken-DeNew & Schmidt 1997), one is therefore forced to ignore the fact that most models will be estimated with observable (worker and firm) covariates, ie assume that Z is orthogonal to D and F. All the other traces can be computed by running auxiliary regressions that do not involve inverting matrices larger than $J \times J$. It is also useful to have software, such as Stata, that "accumulates" data matrices with N^* rows into cross-product matrices of dimension $J \times J$.[15]

In Equations (11–13), suppose that one can invert $\left[D^T M_{[Z,F]} D \right]$. There are two more inversions required in the three biases. The first is $\left[F^T M_V F \right]^{-1}$. Here one takes each column of F, denoted f_j, form mean-deviations for worker i, and repeat for all the columns of Z. Regress f_j in mean-deviations on Z in mean-deviations, and form residuals. Denote this as Regression (R_j). After $j = 1, \ldots, J$ loops, stack the J vectors of residuals, form the inner product, and invert.

The second inversion is $F^T M_Z D \left[D^T M_Z D \right]^{-1}$. Consider the j-th regression

$$f_j = Z\beta_{1j} + D\beta_{2j} + u_j.$$

Using FW, $\hat{\beta}_{2j}$ can be computed in 2 ways:

$$\hat{\beta}_{2j} = [D^T M_Z D]^{-1} D^T M_Z f_j \tag{A.7}$$

or

$$\hat{\beta}_{2j} = [D^T D]^{-1} D^T (f_j - Z\hat{\beta}_{1j}) \qquad \text{with } \hat{\beta}_{1j} = [Z^T M_D Z]^{-1} Z^T M_D f_j. \tag{A.8}$$

Equation (A.7) is what is required; Equation (A.8) gives how to compute it without inverting $N \times N$ matrices. In other words, run Regression (R_j) above and form "residuals" $f_j - Z\hat{\beta}_{1j}$. Take the average for each worker i and save as a $N \times 1$ vector. After looping over $j = 1, \ldots, J$ firms, form the $N \times J$ matrix, as required.

---

[15]Stata code that computes all the biases given above is available on request.

References

Abowd, J., Creecy, R. & Kramarz, F. (2002), Computing person and firm effects using linked longitudinal employer-employee data, Technical Paper 2002-06, U.S. Census Bureau.

Abowd, J. & Kramarz, F. (1999), The analysis of labor markets using matched employer-employee data, *in* O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', Vol. 3B, Elsevier, Amsterdam, chapter 40, pp. 2567–627.

Abowd, J., Kramarz, F., Lengermann, P. & Perez-Duarte, S. (2004), Are good workers employed by good firms? A test of a simple assortative matching model for France and the United States, mimeo.

Abowd, J., Kramarz, F. & Margolis, D. (1999), 'High wage workers and high wage firms', *Econometrica* **67**, 251–333.

Alda, H., Bender, S. & Gartner, H. (2005), The linked employer-employee dataset of the IAB (LIAB), Discussion Paper No. 06/2005, IAB.

Andrews, M., Schank, T. & Upward, R. (2005), Practical estimation methods for linked employer-employee data, Discussion Paper No. 29, University of Erlangen-Nürnburg.

Andrews, M., Schank, T. & Upward, R. (2006), 'Practical fixed effects estimation methods for the three-way error components model'. Forthcoming *Stata Journal*.

Baltagi, B. (2005), *Econometric Analysis of Panel Data*, third edn, Wiley.

Barth, E. & Dale-Olsen, H. (2003), Assortative matching in the labour market? Stylised facts about workers and plants, mimeo, Institute for Social Research, Oslo.

Bender, S., Haas, A. & Klose, C. (2000), The IAB employment subsample 1975-1995: Opportunities for analysis provided by the anonymised sample, Discussion Paper No. 117, IZA.

Goux, D. & Maurin, E. (1999), 'Persistence of interindustry wage differentials: a reexamination using matched worker-firm panel data', *Journal of Labor Economics* **17**, 492–533.

Gruetter, M. & Lalive, R. (2004), The importance of firms in wage determination, Discussion Paper No. 1367, IZA.

Haisken-DeNew, J. & Schmidt, C. (1997), 'Interindustry and interregion differentials: mechanics and interpretation', *Review of Economics and Statistics* **79**, 516–21.

Haltiwanger, J., Lane, J., Spletzer, J., Theeuwes, J. & Troske, K., eds (1999), *The creation and analysis of employer-employee matched data*, North-Holland.

Kölling, A. (2000), 'The IAB establishment panel', *Schmollers Jahrbuch: Zeitschrift für Wirtschafts- und Sozialwissenschaften* **120**, 291–300.

Krueger, A. & Summers, L. (1988), 'Efficiency wages and the inter-industry wage structure', *Econometrica* **56**, 259–93.