High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias?

Author(s): M. J. Andrews, L. Gill, T. Schank and R. Upward

Source: *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 2008, Vol. 171, No. 3 (2008), pp. 673-697

Published by: Wiley for the Royal Statistical Society

Stable URL: https://www.jstor.org/stable/30135090

J. R. Statist. Soc. A (2008)
171, *Part* 3, *pp.* 673–697

High wage workers and low wage firms: negative assortative matching or limited mobility bias?

M. J. Andrews and L. Gill,

University of Manchester, UK

T. Schank

Universität Erlangen-Nürnberg, Germany

and R. Upward

University of Nottingham, UK

[Received June 2006. Final revision November 2007]

Summary. In the empirical literature on assortative matching using linked employer–employee data, unobserved worker quality appears to be *negatively* correlated with unobserved firm quality. We show that this can be 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. We show that this bias is bigger the fewer movers there are in the data, which is 'limited mobility bias'. This result applies to any two-way (or higher) error components model that is estimated by fixed effects methods. We apply these bias corrections to a large German linked employer–employee data set. We find that, although the biases can be considerable, they are not sufficiently large to remove the negative correlation entirely.

Keywords: Biases; Fixed effects; Limited mobility bias; Linked employer–employee panel data

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 *et al.* (1999) and related references. (See also Abowd and Kramarz (1999) and Haltiwanger *et al.* (1999) for early surveys of the wide range of issues that are covered in this literature.) An important issue in the literature is the relationship between the unobserved worker and firm components of wages. Both economic theory and common sense suggest that there should be a positive correlation between worker and firm productivities. In the words of Abowd *et al.* (1999): 'high-wage workers and high-wage firms' match together. This is known as positive 'assortative matching' in the economics literature (see Atakan (2006), for example, and references within). 'Assortative matching' means that the matching is non-random.

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

Address for correspondence: M. J. Andrews, School of Social Sciences, University of Manchester, Oxford Road, Manchester, M13 9PL, UK.
E-mail: martyn.andrews@manchester.ac.uk

 0964–1998/08/171673

from the original study of Abowd *et al.* (1999), which reports a positive sample correlation, all subsequent work has reported negative correlations. Abowd *et al.* (2004) re-estimated these models by using the exact solution that they developed subsequently and reported correlations of -0.24 for French data and 0.02 for data from Washington State, whereas Goux and Maurin (1999), using different French data, found a correlation ranging from -0.32 to 0.01 depending on the time period that was chosen. Gruetter and Lalive (2004) reported a correlation of -0.27 for Austrian data. All of these are weaker than Barth and Dale-Olsen's (2003) correlations of between -0.47 and -0.55 for Norwegian data. In other words, when focusing on unobserved components, low wage workers tend to work in high wage firms, and vice versa. This seems counterintuitive in the light of theories of assortative matching.

There are two possible explanations for this emerging stylized fact. The first, which was suggested by Barth and Dale-Olsen (2003) and Abowd *et al.* (2004) is that the observed negative sample correlation is simply the result of using standard econometric techniques. Because the estimates of the worker and firm dummy variables 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 overestimate of a worker effect leads to, on average, an underestimate of a firm effect. The second explanation focuses on whether there are any genuine economic explanations for why there might be negative assortative matching. Again, see Abowd *et al.* (2004).

In this paper, we analyse the first explanation. In what we label a fixed effects data generation process, we consider the standard fixed effects estimator of the model; we then derive formulae for the bias in the sampling distribution of the sample covariance between the unobserved worker and firm components of wages, and the biases in the sample 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 dummy variables, we show that the bias in the sample 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 dummy variables. We also show that correcting the biases in the sample covariance and two sample variance terms leads to consistent estimators and, by Slutsky's theorem, the correlation can also be consistently estimated. By simulating the data generation process, we show that the corresponding bias-corrected estimator of the correlation is also approximately unbiased. Subject to possible size constraints, the bias corrections can be computed for any given data set.

The fixed effects data generation process treats the worker and firm heterogeneity terms as population parameters that can be estimated. The more modern treatment is to assume that the heterogeneity terms are unobserved random components that are correlated with the observed covariates. We refer to this as the random-effects data generation process. In this paper we rely on simulation results to show that exactly the same formulae for the biases are applicable to the random-effects data generation process.

Abowd *et al.* (2004) suggested that the bias in the estimated correlation is bigger when there are fewer movers in the data, which they labelled 'limited mobility bias'. We supply a formula that establishes this proposition in a simple, stylized data set. For more realistic data sets, we must simulate the random-effects data generation process to investigate this further.

Ultimately, the size of the bias is an empirical issue and should be computed for every application of linked employer–employee data. More importantly, this result applies to any two-way (or higher) error components model that is estimated by fixed effects methods. Another example occurs if there is a true positive correlation between unobservably good schools and unobservably good pupils, and these error components were correlated with the observed covariates. Again the fixed effects estimator of the correlation would be biased.

Because it is possible that all of the negative estimates that have been 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 (IAB), Nürnberg. It turns out that our bias correction moves the estimate of the correlation from -0.19 to -0.15, and so the econometric explanation 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 are 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 that is used in most of the literature and we explain the methods that are used to estimate the parameters of this model. In Section 3, for the fixed effects data generation process, we derive expressions for the biases in the sample correlation and its three components and discuss limited mobility bias. In Section 4, we further extend our analysis by generating simulated data for both data generation processes. In Section 5 we report what happens with an example using German linked data, and Section 6 concludes.

2. The generic model

Consider a model of wages with both firm and worker unobserved heterogeneity and firm and worker observed covariates:

$$y_{it} = \mu + \mathbf{x}_{it}\boldsymbol{\beta}_1 + \mathbf{w}_{jt}\boldsymbol{\beta}_2 + \mathbf{u}_i\boldsymbol{\eta} + \mathbf{q}_j\boldsymbol{\rho} + \theta_i + \psi_j + \varepsilon_{it}. \tag{1}$$

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); $\mathbf{x}_{it}$ and $\mathbf{u}_i$ are vectors of observable i-level covariates; $\mathbf{w}_{jt}$ and $\mathbf{q}_j$ are vectors of observable j-level covariates. θ_i and ψ_j are (scalar) unobserved heterogeneities. It is usual to assume that both θ_i and ψ_j 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 θ_i and $\mathbf{u}_i$ are variables that are time invariant for workers. Similarly, ψ_j and $\mathbf{q}_j$ are fixed over time for firms. $\mathbf{x}_{it}$, in contrast, varies across i and t, and $\mathbf{w}_{jt}$ 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.

There are two views about how one interprets θ_i and ψ_j. The traditional view is that θ_i and ψ_j are population parameters to be estimated, and so the associated worker and firm dummy variables are treated no differently from other components of $\mathbf{x}_{it}$ and $\mathbf{w}_{jt}$. The more modern treatment is that they are unobserved random components, which are drawn from the population just like ε_{it}.

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

For deriving properties of the estimators of the model above, it is standard to assume strict exogeneity. For the random-effects data generation process, this is

$$E(\varepsilon_{it}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, \mathbf{w}_{j1}, \ldots, \mathbf{w}_{jT}, \mathbf{u}_i, \mathbf{q}_j, \theta_i, \psi_j) = 0. \tag{2}$$

This implies that workers' mobility decisions are independent of ε_{it}. However, it is worth noting that mobility may be a function of the unobservables θ_i and ψ_j. There is a corresponding assumption for the fixed effects data generation process.

It is usual to assume that the heterogeneity terms θ_i and ψ_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. Consequently, the parameter vector $(\boldsymbol{\eta}, \boldsymbol{\rho})$ that is associated with the time invariant variables is not identified. Thus $\mathbf{u}_i$ and $\mathbf{q}_j$ are dropped from the model, giving

$$y_{it} = \mu + \mathbf{x}_{it}\boldsymbol{\beta}_1 + \mathbf{w}_{jt}\boldsymbol{\beta}_2 + \theta_i + \psi_j + \varepsilon_{it}. \tag{3}$$

The same problem arises in the fixed effects data generation process because $\mathbf{u}_i$ is perfectly collinear with the worker dummy variables and $\mathbf{q}_j$ is perfectly collinear with the firm dummy variables. 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:

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

where $\mathbf{y}$ and ε are $N^* \times 1$ vectors, $\mathbf{D}$ is an $N^* \times N$ matrix of worker dummy variables, $\mathbf{F}$ is an $N^* \times J$ matrix of firm dummy variables and $\mathbf{Z} = (\mathbf{X}, \mathbf{W})$, where $\mathbf{X}$ represents worker covariates and $\mathbf{W}$ represents firm covariates. $\mathbf{Z}$ is an $N^* \times K$ matrix. $\boldsymbol{\theta}$ is an $N \times 1$ parameter vector, $\boldsymbol{\psi}$ is a $J \times 1$ parameter vector and γ is a $K \times 1$ parameter vector. Because one firm dummy variable is dropped, J is redefined accordingly, and note that $\mathbf{Z}$ does not contain a constant. Another reason why J might need to be redefined is that ψ_j for a firm with no movement is not identified.

As noted by Abowd *et al.* (1999), the least squares dummy variable (LSDV) estimator of equation (3) requires the estimation of N worker effects and J firm effects. N is often of the order of millions, and J is often of 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—i.e. where the firm effects are absent—the LSDV estimator gives identical results to models where the heterogeneity is removed algebraically, by forming within-worker mean deviations for all the variables in equation (4). In the standard two-way fixed effects model with worker and time dummy variables we can form the familiar 'double-mean' deviations. This fails in our model because creating firm mean deviations destroys the patterning in $\mathbf{D}$ and creating worker mean deviations destroys the patterning in $\mathbf{F}$. More precisely, sort the data by workers, and the firm dummy variables are unpatterned; sort the data by firms, and the worker dummy variables are unpatterned. In the standard two-way fixed effects model the patterning is preserved.

To circumvent this problem, Abowd *et al.* (1999) noted 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 dummy variables $\mathbf{D}$ by using $\mathbf{M_D} \equiv \mathbf{I}_{N^*} - \mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}$:

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

In words, $y_{it} - \bar{y}_i$ is regressed on the vector of covariates $\mathbf{z}_{it} - \bar{\mathbf{z}}_i$ *and* on J mean-deviated firm dummy variables $F_{it}^j - \bar{F}_i^j$, where F_{it}^j are elements of the jth column of $\mathbf{F}$, and $\bar{r}_i = (\Sigma_t r_{it})/T_i$ for any variable r. We label this estimator 'FEiLSDVj'. The covariance matrix for FEiLSDVj needs the standard degrees-of-freedom adjustment. This means implicitly accounting for the parameters in $\boldsymbol{\theta}$ in the model which might be ignored by standard regression software that is used for estimating equation (5).

To obtain estimates of the worker heterogeneity, note that

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

where $\mathbf{P_D} \equiv \mathbf{D}(\mathbf{D}^\mathrm{T}\mathbf{D})^{-1}\mathbf{D}^\mathrm{T}$. This equation gives intuition about why there is an observed negative correlation between $\hat{\theta}_i$ and $\hat{\psi}_j$ (as noted by Barth and Dale-Olsen (2003) and Abowd *et al.* (2004)). To see this, write out equation (6) explicitly for each worker:

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

The matrix $\mathbf{P_D}$ forms averages for each worker i, and so the typical element of $\mathbf{P_D}\mathbf{F}\hat{\boldsymbol{\psi}}$ is $\bar{\hat{\psi}}_i$, which forms a weighted average of $\hat{\psi}_{j(it)}$ over t. As the ψ_j are estimated by LSDV, they are subject to the usual sampling variation (the firm dummy variables are no different from any other observed covariate). Once estimated, each $\hat{\psi}_j$ generates a number of $\hat{\theta}_i$, via equation (7). It can also be shown that

$$\hat{\theta}_i - \theta_i = -\bar{\mathbf{z}}_i(\hat{\gamma} - \gamma) - (\bar{\hat{\psi}}_i - \bar{\psi}_i) + \bar{\varepsilon}_i,$$

where $\bar{\varepsilon}_i$ averages ε_{it} over t. Conditional on the observed covariates, if a ψ_j is overestimated, then, on average, the corresponding θ_i are underestimated, and vice versa. This implies that the estimated correlation between θ_i and ψ_j is biased downwards. For the fixed effects data generation process, it is possible to derive an expression for this bias; this is done in the next section, together with an analysis of whether the bias disappears asymptotically. For the random-effects data generation process, we use simulation methods.

3. Bias and consistency for the fixed effects data generation process

In this section, θ_i and ψ_j are population parameters that are to be estimated. In other words, the associated worker and firm dummy variables $\mathbf{D}$ and $\mathbf{F}$ are treated the same as the remaining covariates $\mathbf{Z}$. After equation (3) has been estimated by fixed effects, because the unit of observation is a worker-year, each $\hat{\theta}_i$ is the itth element of the $N^* \times 1$ vector $\mathbf{D}\hat{\boldsymbol{\theta}}$; only N of these $\hat{\theta}_i$s are distinct. Similarly, the itth element of the $N^* \times 1$ vector $\mathbf{F}\hat{\boldsymbol{\psi}}$ comprises $\hat{\psi}_j$. Only J of these are distinct. The weighted sample variances, covariance and correlation of θ_i and ψ_j are computed as follows:

$$\tilde{\sigma}^2_\theta = \frac{1}{N^*}\sum_{it}(\hat{\theta}_i - \bar{\hat{\theta}})^2 = \frac{\hat{\boldsymbol{\theta}}^\mathrm{T}\mathbf{D}^\mathrm{T}\mathbf{A}\mathbf{D}\hat{\boldsymbol{\theta}}}{N^*}, \tag{8}$$

$$\tilde{\sigma}^2_\psi = \frac{1}{N^*}\sum_{it}(\hat{\psi}_j - \bar{\hat{\psi}})^2 = \frac{\hat{\boldsymbol{\psi}}^\mathrm{T}\mathbf{F}^\mathrm{T}\mathbf{A}\mathbf{F}\hat{\boldsymbol{\psi}}}{N^*}, \tag{9}$$

$$\tilde{\sigma}_{\theta\psi} = \frac{1}{N^*}\sum_{it}(\hat{\theta}_i - \bar{\hat{\theta}})(\hat{\psi}_j - \bar{\hat{\psi}}) = \frac{\hat{\boldsymbol{\theta}}^\mathrm{T}\mathbf{D}^\mathrm{T}\mathbf{A}\mathbf{F}\hat{\boldsymbol{\psi}}}{N^*}, \tag{10}$$

$$\tilde{\rho}_{\theta\psi} = \frac{\tilde{\sigma}_{\theta\psi}}{\surd(\tilde{\sigma}^2_\theta\tilde{\sigma}^2_\psi)}. \tag{11}$$

$\bar{\hat{\theta}}$ averages $\hat{\theta}_i$ over the N^* worker-years and similarly $\bar{\hat{\psi}}$ averages $\hat{\psi}_j$ over N^* worker-years. Because these averages are non-zero, this gives rise to the residual projector for the intercept $\mathbf{A}$ in these expressions, where $\mathbf{A} = \mathbf{I}_{N^*} - \mathbf{1}(\mathbf{1}^\mathrm{T}\mathbf{1})^{-1}\mathbf{1}^\mathrm{T}$ and $\mathbf{1}$ is an $N^* \times 1$ vector of 1s. Because each of $\tilde{\sigma}^2_\psi$, $\tilde{\sigma}^2_\theta$ and $\tilde{\sigma}_{\theta\psi}$ is computed over N^* observations, we note that all three are weighted estimators.

The vectors $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\psi}}$ suffer standard least squares estimation error, and so we compare the means of the sampling distributions of $\tilde{\sigma}^2_\theta$, $\tilde{\sigma}^2_\psi$ and $\tilde{\sigma}_{\theta\psi}$ with their respective population quantities by replacing $\hat{\boldsymbol{\theta}}$ by $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\psi}}$ by $\boldsymbol{\psi}$:

$$\sigma^2_{\theta N} = \frac{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\boldsymbol{\theta}}{N^*}, \qquad \sigma^2_{\psi N} = \frac{\boldsymbol{\psi}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}\boldsymbol{\psi}}{N^*}, \qquad \sigma_{\theta\psi N} = \frac{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{F}\boldsymbol{\psi}}{N^*}. \tag{12}$$

We assume that as $N^* \to \infty$ these 'finite sample' population quantities converge to limits that are denoted σ^2_θ, σ^2_ψ and $\sigma_{\theta\psi}$ respectively. A corresponding population correlation is defined by

$$\rho_{\theta\psi N} = \frac{\sigma_{\theta\psi N}}{\surd(\sigma^2_{\theta N}\sigma^2_{\psi N})}, \tag{13}$$

which is assumed to converge to a limit $\rho_{\theta\psi} = \sigma_{\theta\psi}/\sigma_\theta\sigma_\psi$, as $N^* \to \infty$. The distinction between these two concepts will be useful in the simulations later.

Our strategy in what follows is to show that each of $\tilde{\sigma}^2_\theta$, $\tilde{\sigma}^2_\psi$ and $\tilde{\sigma}_{\theta\psi}$ is a biased estimator of $\sigma^2_{\theta N}$, $\sigma^2_{\psi N}$ and $\sigma_{\theta\psi N}$ respectively. Each bias is a function of $\mathbf{Z}$, $\mathbf{D}$, $\mathbf{F}$ and the error variance σ^2_ε. From these we define infeasible bias-adjusted estimators $\hat{\sigma}_{\theta\psi}$, $\hat{\sigma}^2_\theta$ and $\hat{\sigma}^2_\psi$. These infeasible estimators can be estimated by using a consistent estimate of σ^2_ε. Given $\mathbf{Z}$, the conditional expectation of each of the bias-corrected estimators is independent of $\mathbf{Z}$ and is thus an unconditional expectation as well. We show that each uncorrected estimator has a bias which does not disappear as $N \to \infty$. Thus we can then examine whether the bias-corrected estimators are consistent by seeing whether the variance of each estimator goes to 0 as N grows larger. (This is why we use subscript N in equation (12).) There are two distinct cases. In the first, we hold both the number of firms J and the number of time periods T fixed; in the second we let $J \to \infty$ as well as $N \to \infty$ so that $J/N \to k$, where k is 1/(average firm size). Probably, the latter is more appealing conceptually. In what follows, we summarize the key results; all the formal analysis is presented in Appendix A. Note that stronger assumptions are required to establish consistency for $\tilde{\sigma}^2_\theta$, $\tilde{\sigma}^2_\psi$ and $\tilde{\sigma}_{\theta\psi}$ than the strict exogeneity assumption (which was noted in Section 2) for consistency of $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\psi}}$.

To derive the properties of $\tilde{\sigma}^2_\theta$ and $\tilde{\sigma}^2_\psi$, we need convenient expressions for the LSDV estimators $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\psi}}$. We first eliminate $\mathbf{Z}$ from equation (4) by using the transformation $\mathbf{M_Z} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}$ to produce

$$\mathbf{M_Z y} = \mathbf{M_Z D}\boldsymbol{\theta} + \mathbf{M_Z F}\boldsymbol{\psi} + \mathbf{M_Z}\boldsymbol{\varepsilon},$$

and then $\mathbf{M_Z F} = \tilde{\mathbf{F}}$ is eliminated by using the transformation $\tilde{\mathbf{M}}_{\mathbf{F}} = \mathbf{I} - \tilde{\mathbf{F}}(\tilde{\mathbf{F}}^{\mathrm{T}}\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}^{\mathrm{T}}$ to give

$$\tilde{\mathbf{M}}_{\mathbf{F}}\mathbf{M_Z y} = \tilde{\mathbf{M}}_{\mathbf{F}}\mathbf{M_Z D}\boldsymbol{\theta} + \tilde{\mathbf{M}}_{\mathbf{F}}\mathbf{M_Z}\boldsymbol{\varepsilon}.$$

From this, the LSDV estimator of $\boldsymbol{\theta}$ (which is also the ordinary least squares estimator) is given by

$$\hat{\boldsymbol{\theta}} = (\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{y}, \tag{14}$$

where $\tilde{\mathbf{Q}}_{\mathbf{F}} = \mathbf{M}^{\mathrm{T}}_{\mathbf{Z}}\tilde{\mathbf{M}}^{\mathrm{T}}_{\mathbf{F}}\tilde{\mathbf{M}}_{\mathbf{F}}\mathbf{M_Z} = \mathbf{M_Z}\tilde{\mathbf{M}}_{\mathbf{F}}\mathbf{M_Z}$ owing to the symmetry of the $\tilde{\mathbf{M}}_{\mathbf{F}}$ and $\mathbf{M_Z}$ matrices and also their idempotency, e.g. $\tilde{\mathbf{M}}_{\mathbf{F}}\tilde{\mathbf{M}}_{\mathbf{F}} = \tilde{\mathbf{M}}_{\mathbf{F}}$.

Analogous arguments generate an expression for $\hat{\boldsymbol{\psi}}$: project out $\mathbf{Z}$; then project out $\mathbf{M_Z D} = \tilde{\mathbf{D}}$ to give

$$\hat{\boldsymbol{\psi}} = (\mathbf{F}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{D}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{D}}\mathbf{y}, \tag{15}$$

where $\tilde{\mathbf{Q}}_{\mathbf{D}} = \mathbf{M_Z}\tilde{\mathbf{M}}_{\mathbf{D}}\mathbf{M_Z}$.

3.1. Variance of worker heterogeneity

We first show that $\tilde{\sigma}^2_\theta$ is a biased estimator of $\sigma^2_{\theta N}$. From equation (14), $\hat{\boldsymbol{\theta}}$ can be decomposed as

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + (\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\boldsymbol{\varepsilon}.$$

We then find that, conditional on $\mathbf{Z}$,

$$E(\tilde{\sigma}^2_\theta) = E\left(\frac{\hat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\hat{\boldsymbol{\theta}}}{N^*}\right) = \frac{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\boldsymbol{\theta}}{N^*} + \frac{\sigma^2_\varepsilon}{N^*}\mathrm{tr}\{(\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\}, \tag{16}$$

where σ^2_ε is the variance of ε_{it}. The bias is the trace term in equation (16) and, to show that it is unambiguously non-negative, note that

$$\mathrm{tr}\{(\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\} = \mathrm{tr}\{(\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D})^{1/2}(\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D})^{-1}(\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D})^{1/2}\}.$$

This is the trace of a positive semidefinite matrix and is therefore non-negative. The bias arises purely because of the extra variation that is caused by having to estimate the θ_is. The largest eigenvalue of $(\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}$ is bounded above by a finite positive constant $c_{\mathbf{D}}$ that does not depend on N, N^*, T_i or $\mathbf{Z}$, under the assumptions in Appendix A.2. Using the arguments in Appendix A.4, parts (a)–(c), it follows that

$$\frac{1}{N^*}\mathrm{tr}\{(\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\} < \frac{1}{N^*}Nc_{\mathbf{D}} < \infty. \tag{17}$$

The bias does not disappear as $N \to \infty$ because N^*/N converges to the average number of periods observed per worker. This is true for both J fixed and $J \to \infty$.

An infeasible bias-corrected estimator for σ^2_θ is constructed as

$$\hat{\sigma}^2_\theta = \frac{\hat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\hat{\boldsymbol{\theta}} - \sigma^2_\varepsilon\,\mathrm{tr}\{(\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\}}{N^*}, \tag{18}$$

which has expected value $\sigma^2_{\theta N}$. This is the conditional expectation of $\hat{\sigma}^2_\theta$, given $\mathbf{Z}$. As explained, because it is independent of $\mathbf{Z}$, it is also the unconditional expectation.

Under the assumption that $\sigma^2_{\theta N}$ converges to a limit σ^2_θ, the argument in Appendix A.2 and Appendix A.5, part (a), exploits the upper bound $c_{\mathbf{D}}$ to show that the conditional variance of $\tilde{\sigma}^2_\theta$, and hence the unconditional variance of $\hat{\sigma}^2_\theta$, goes to 0 as $N^* \to \infty$, and then it follows that the bias-corrected estimator is consistent:

$$\hat{\sigma}^2_\theta \xrightarrow{\mathrm{p}} \lim_{N\to\infty}\left(\frac{\boldsymbol{\theta}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\boldsymbol{\theta}}{N^*}\right) = \sigma^2_\theta.$$

Again, this is true for both J fixed and $J \to \infty$.

3.2. Variance of firm heterogeneity

Analogous arguments apply to $\tilde{\sigma}^2_\psi$. Using equation (15):

$$\hat{\boldsymbol{\psi}} = \boldsymbol{\psi} + (\mathbf{F}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{D}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{D}}\boldsymbol{\varepsilon},$$

and it follows that, conditional on $\mathbf{Z}$,

$$E(\tilde{\sigma}^2_\psi) = E\left(\frac{\hat{\boldsymbol{\psi}}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}\hat{\boldsymbol{\psi}}}{N^*}\right) = \frac{\boldsymbol{\psi}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}\boldsymbol{\psi}}{N^*} + \frac{\sigma^2_\varepsilon}{N^*}\mathrm{tr}\{(\mathbf{F}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{D}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}\}.$$

By symmetry with the worker effects, this bias is also unambiguously positive. It is well known that $\tilde{\sigma}^2_{\psi}$, in the absence of worker dummy variables, is biased upwards (Krueger and Summers, 1988). Again, the arguments in Appendix A.2 and Appendix A.4, parts (a)–(c), can be applied to show that

$$\frac{1}{N^*}\mathrm{tr}\{(\mathbf{F}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{D}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}\} \leqslant \frac{c_{\mathbf{F}}J}{N^*}, \tag{19}$$

and so the bias goes to 0 if J is fixed, but the bias does not disappear if both J and N go to ∞. $c_{\mathbf{F}}$ is a positive constant that does not depend on N, N^*, T_i or $\mathbf{Z}$. With J fixed, the precision of the estimates of ψ_j improves as N grows bigger, and so the bias falls. The infeasible bias-corrected estimator $\hat{\sigma}^2_{\psi}$ is written

$$\hat{\sigma}^2_{\psi} = \frac{\hat{\psi}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}\hat{\psi} - \sigma^2_{\varepsilon}\,\mathrm{tr}\{(\mathbf{F}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{D}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}\}}{N^*}. \tag{20}$$

Using the same arguments as for worker heterogeneity, we show that, conditionally on $\mathbf{Z}$, $\mathrm{var}(\hat{\sigma}^2_{\psi}) \to 0$ as $N \to \infty$ when J is fixed *and* when $J \to \infty$. When J is fixed, the bias correction is not necessary asymptotically; however, when $J \to \infty$, it is essential.

3.3. Covariance between worker and firm heterogeneity

The estimator $\tilde{\sigma}_{\theta\psi}$ of the covariance $\sigma_{\theta\psi N}$ has expected value

$$E(\tilde{\sigma}_{\theta\psi}) = E\left(\frac{\hat{\theta}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{F}\hat{\psi}}{N^*}\right) = \frac{\theta^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{F}\psi}{N^*} + \frac{\sigma^2_{\varepsilon}}{N^*}\mathrm{tr}\{\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D}(\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{F}(\mathbf{F}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{D}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{D}}\}. \tag{21}$$

Combining the results in Appendix A.4, parts (a)–(c), and the Cauchy–Schwartz inequality for traces that is stated in Appendix A.4, part (a)(ii), shows that the trace term is bounded by $(NJc_{\mathbf{D}}c_{\mathbf{F}})^{1/2}$, as indicated in Appendix A.4, part (d). This means that the bias correction is not needed when J is fixed but is needed when $J \to \infty$. The trace can also be written as

$$-\frac{\sigma^2_{\varepsilon}}{N^*}\mathrm{tr}\{\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{F}(\mathbf{F}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{D}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{M}_{\mathbf{Z}}\mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{M}_{\mathbf{Z}}\mathbf{D})^{-1}\}, \tag{22}$$

because $\mathbf{F}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{D}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D} = -\mathbf{F}^{\mathrm{T}}\mathbf{M}_{\mathbf{Z}}\mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{M}_{\mathbf{Z}}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D}$. This is useful because, when there are no $\mathbf{Z}$s in the model, or when the columns of $\mathbf{Z}$ are orthogonal to $(\mathbf{D}, \mathbf{F})$, the bias becomes

$$-\frac{\sigma^2_{\varepsilon}}{N^*}\mathrm{tr}\{\mathbf{F}^{\mathrm{T}}\mathbf{P}_{\mathbf{D}}\mathbf{A}\mathbf{F}(\mathbf{F}^{\mathrm{T}}\mathbf{M}_{\mathbf{D}}\mathbf{F})^{-1}\}, \tag{23}$$

as shown in Appendix A.4, part (e). In this case, the trace can be unambiguously signed as positive, in which case the estimated covariance is biased downwards. However, with $\mathbf{Z}$s present, as a particular column of $\mathbf{Z}$ becomes less orthogonal to $(\mathbf{D}, \mathbf{F})$, 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 are an empirical issue, but the case where $\mathbf{Z}$ is orthogonal to $(\mathbf{D}, \mathbf{F})$ is a useful benchmark.

The bias-corrected estimator of $\sigma_{\theta\psi}$ is written as

$$\hat{\sigma}_{\theta\psi} = \frac{\hat{\theta}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{F}\hat{\psi} + \sigma^2_{\varepsilon}\,\mathrm{tr}\{\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{F}(\mathbf{F}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{D}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{M}_{\mathbf{Z}}\mathbf{D}(\mathbf{D}^{\mathrm{T}}\mathbf{M}_{\mathbf{Z}}\mathbf{D})^{-1}\}}{N^*}. \tag{24}$$

Because we can show that $\mathrm{var}(\hat{\sigma}^2_{\theta\psi}) \rightarrow 0$ in both J cases, it follows that the bias-corrected estimator of $\sigma_{\theta\psi}$ is consistent for the limit of $\sigma_{\theta\psi N}$ as $N \rightarrow \infty$. Just as with firm heterogeneity, bias correction is necessary for consistency when $J \rightarrow \infty$.

In Sections 3.1 and 3.2 we showed that both $\tilde{\sigma}^2_\theta$ and $\tilde{\sigma}^2_\psi$ are biased upwards, whereas, as argued at the end of Section 2, $\tilde{\sigma}_{\theta\psi}$ is probably biased downwards, and, as just seen, definitely so if $\mathbf{Z}$ is orthogonal to $\mathbf{D}$ and $\mathbf{F}$. Equations (18), (20) and (24) display bias-corrected estimators. By showing that the unconditional variance of these bias-corrected estimators goes to 0 as $N^* \rightarrow \infty$, as in Appendix A.5, their consistency is guaranteed. To make the bias-corrected estimators feasible, a consistent estimator of σ^2_ε is needed. This is the standard estimator $\hat{\sigma}^2_\varepsilon$ that is based on the residuals after estimating by LSDV. Replacing σ^2_ε by $\hat{\sigma}^2_\varepsilon$ in the expressions for the bias-corrected estimators $\hat{\sigma}_{\theta\psi}, \hat{\sigma}^2_\theta$ and $\hat{\sigma}^2_\psi$ does not affect their consistency.

3.4. Consistency of correlation estimators

It is the correlation between the unobserved worker and firm components of $\mathbf{y}$, $\rho_{\theta\psi N}$, that is of interest (see equation (13)). $\rho_{\theta\psi N}$ is a continuous function of $\sigma^2_{\theta N}$, $\sigma^2_{\psi N}$ and $\sigma_{\theta\psi N}$, provided that $\sigma^2_{\theta N}$ and $\sigma^2_{\psi N}$ are non-zero. As a result, an estimator $\hat{\rho}_{\theta\psi}$ of $\rho_{\theta\psi N}$ that is constructed from the bias-corrected and hence consistent estimators $\hat{\sigma}^2_\theta$, $\hat{\sigma}^2_\psi$ and $\hat{\sigma}_{\theta\psi}$ is given by

$$\hat{\rho}_{\theta\psi} = \frac{\hat{\sigma}_{\theta\psi}}{\surd(\hat{\sigma}^2_\theta \hat{\sigma}^2_\psi)}, \tag{25}$$

and converges in probability to $\rho_{\theta\psi}$, provided that σ^2_θ and σ^2_ψ are non-zero. Even though $\hat{\rho}_{\theta\psi}$ cannot be shown to be unbiased, the consistency of $\hat{\rho}_{\theta\psi}$ for $\rho_{\theta\psi}$ shows that the sampling distribution of $\hat{\rho}_{\theta\psi}$ is properly centred at $\rho_{\theta\psi}$, as $N^* \rightarrow \infty$. In practice we might expect the bias in $\hat{\rho}_{\theta\psi}$ to be relatively small. This is established by simulations in Section 4.2 below.

The nature of the inconsistency of the non-bias-corrected correlation estimator $\tilde{\rho}_{\theta\psi}$ depends on whether J is fixed or goes to ∞ as $N^* \rightarrow \infty$. For the fixed J case, we have established that $\tilde{\sigma}^2_\psi$ and $\tilde{\sigma}_{\theta\psi}$ are consistent for σ^2_ψ and $\sigma_{\theta\psi}$ respectively, but that $\tilde{\sigma}^2_\theta$ is inconsistent for σ^2_θ, because the bias correction term does not disappear as $N^* \rightarrow \infty$. When $J \rightarrow \infty$, the sign of the inconsistency is indeterminate, because all three components of $\tilde{\rho}_{\theta\psi}$ are inconsistent.

Finally, we need to establish whether the non-bias-corrected correlation is biased upwards or downwards in finite samples, since most of the evidence that was noted in Section 1 suggests that it is biased downwards. If the covariance term is biased downwards, and if the true covariance is positive (i.e. there is positive assortative matching), then the estimated correlation will always be too small and could be negative. Signing the bias is less clear cut if the true covariance is negative, and the covariance term is biased downwards. Now the estimated correlation could either be smaller or larger than the true (negative) value. (For example, suppose that $\sigma_{\theta\psi} = -0.2$ and $\sigma^2_\theta = \sigma^2_\psi = 1$ and suppose that the estimates are $\hat{\sigma}_{\theta\psi} = -0.3$, $\hat{\sigma}^2_\theta = 1.2$ and $\hat{\sigma}^2_\psi = 1.4$. Then the true correlation of –0.2 is estimated as –0.179, i.e. it has gone the 'wrong' way.) All this is further complicated if we cannot sign the bias in the covariance term, i.e. when the columns of $\mathbf{Z}$ are not orthogonal to $\mathbf{D}$ and $\mathbf{F}$.

3.5. Limited mobility bias

Abowd *et al.* (2004) suggested that the bias in the estimated correlation is bigger when there are fewer movers in the data, which they labelled 'limited mobility bias'. The source of the bias is the covariance term. This happens whether or not there are covariates $\mathbf{Z}$; formulae with $\mathbf{Z}$s just complicate the analysis. From equation (23), we can rewrite the bias as

$$-\frac{\sigma_\varepsilon^2}{N^*}\mathrm{tr}\{\mathbf{F}^{\mathrm{T}}(\mathbf{A}-\mathbf{M_D})\mathbf{F}(\mathbf{F}^{\mathrm{T}}\mathbf{M_D}\mathbf{F})^{-1}\} = -\frac{\sigma_\varepsilon^2}{N^*}\mathrm{tr}\{\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}(\mathbf{F}^{\mathrm{T}}\mathbf{M_D}\mathbf{F})^{-1}\} + \frac{J\sigma_\varepsilon^2}{N^*}, \tag{26}$$

since $\mathbf{P_D}\mathbf{A}=\mathbf{P_D}-\mathbf{P}_1=\mathbf{A}-\mathbf{M_D}$ and $\mathbf{P}_1=\mathbf{1}(\mathbf{1}^{\mathrm{T}}\mathbf{1})^{-1}\mathbf{1}^{\mathrm{T}}$ (see Appendix A.4, part (e)). Recall that J is the number of identifiable firm effects.

To illustrate, consider a balanced panel where all N workers are observed twice. We compare two data sets. In both data sets, only so-called balanced moves take place: this is where one worker moves from firm 1 to firm 2, another from firm 2 to firm 3, ..., and a $(J+1)$th worker from firm $J+1$ to firm 1. (We consider $J+1$ firms because one firm effect is not identified.) In the first data set, only one balanced move happens ($J+1$ actual moves); in the second, two balanced moves happen ($2(J+1)$ actual moves). It is the same $J+1$ firms in the two data sets. The matrices of firm dummy variables are denoted $\mathbf{F}_1$ and $\mathbf{F}_2$ respectively. The reason why we consider balanced moves is because the matrix $\mathbf{A}$ takes mean deviations over all N^* observations, and then it is easy to show that $\mathbf{F}_2^{\mathrm{T}}\mathbf{A}\mathbf{F}_2=\mathbf{F}_1^{\mathrm{T}}\mathbf{A}\mathbf{F}_1$. Although this is unlikely to happen in practice—essentially the distribution of firm size is unaffected by these $J+1$ or $2(J+1)$ moves—it means that the $\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}$ term on the right-hand side of equation (26) is unaffected by the number of moves.

We now consider $\mathbf{F}_1^{\mathrm{T}}\mathbf{M_D}\mathbf{F}_1$ and $\mathbf{F}_2^{\mathrm{T}}\mathbf{M_D}\mathbf{F}_2$. It can be shown that

$$\mathbf{F}_2^{\mathrm{T}}\mathbf{M_D}\mathbf{F}_2=2\mathbf{F}_1^{\mathrm{T}}\mathbf{M_D}\mathbf{F}_1=\mathbf{B},$$

where

$$\mathbf{B}=\begin{pmatrix} 2 & -1 & 0 & \ldots & 0 & 0 & -1 \\ -1 & 2 & -1 & & & 0 & 0 \\ 0 & -1 & 2 & & & & 0 \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ 0 & & & & 2 & -1 & 0 \\ 0 & 0 & & & -1 & 2 & -1 \\ -1 & 0 & 0 & \ldots & 0 & -1 & 2 \end{pmatrix}.$$

$\mathbf{B}$ is a $(J+1)\times(J+1)$ positive semidefinite matrix whose trace records the number of movers and whose off-diagonal elements record which firms are affected by each move. As a result $(\mathbf{F}_1^{\mathrm{T}}\mathbf{M_D}\mathbf{F}_1)^{-1}-(\mathbf{F}_2^{\mathrm{T}}\mathbf{M_D}\mathbf{F}_2)^{-1}$ is a positive semidefinite matrix, and so

$$(\mathbf{F}_1^{\mathrm{T}}\mathbf{A}\mathbf{F}_1)^{1/2}(\mathbf{F}_1^{\mathrm{T}}\mathbf{M_D}\mathbf{F}_1)^{-1}(\mathbf{F}_1^{\mathrm{T}}\mathbf{A}\mathbf{F}_1)^{1/2}-(\mathbf{F}_1^{\mathrm{T}}\mathbf{A}\mathbf{F}_1)^{1/2}(\mathbf{F}_2^{\mathrm{T}}\mathbf{M_D}\mathbf{F}_2)^{-1}(\mathbf{F}_1^{\mathrm{T}}\mathbf{A}\mathbf{F}_1)^{1/2}$$

is also positive semidefinite. It follows that

$$\mathrm{tr}\{\mathbf{F}_1^{\mathrm{T}}\mathbf{A}\mathbf{F}_1(\mathbf{F}_1^{\mathrm{T}}\mathbf{M_D}\mathbf{F}_1)^{-1}\}-\mathrm{tr}\{\mathbf{F}_1^{\mathrm{T}}\mathbf{A}\mathbf{F}_1(\mathbf{F}_2^{\mathrm{T}}\mathbf{M_D}\mathbf{F}_2)^{-1}\}\geqslant 0.$$

Because $\mathbf{F}_2^{\mathrm{T}}\mathbf{M_D}\mathbf{F}_2^{\mathrm{T}}=2\mathbf{F}_1^{\mathrm{T}}\mathbf{M_D}\mathbf{F}_1$, it follows that the J positive roots of $\mathbf{F}_1^{\mathrm{T}}\mathbf{A}\mathbf{F}_1(\mathbf{F}_1^{\mathrm{T}}\mathbf{M_D}\mathbf{F}_1)^{-1}$ are twice those of $\mathbf{F}_1^{\mathrm{T}}\mathbf{A}\mathbf{F}_1(\mathbf{F}_2^{\mathrm{T}}\mathbf{M_D}\mathbf{F}_2)^{-1}$, and so the trace halves. By repeating this comparison, it therefore follows that the trace term $\mathrm{tr}\{\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}(\mathbf{F}^{\mathrm{T}}\mathbf{M_D}\mathbf{F})^{-1}\}$ in equation (26) is of the form k/M, where k is a constant ($k>JM$), and so the overall bias in the covariance is written

$$-\frac{\sigma_\varepsilon^2}{N^*}\left(\frac{k}{M}-J\right), \tag{27}$$

where M is the number of moves. This formula only applies when J is held fixed. Suppose that we

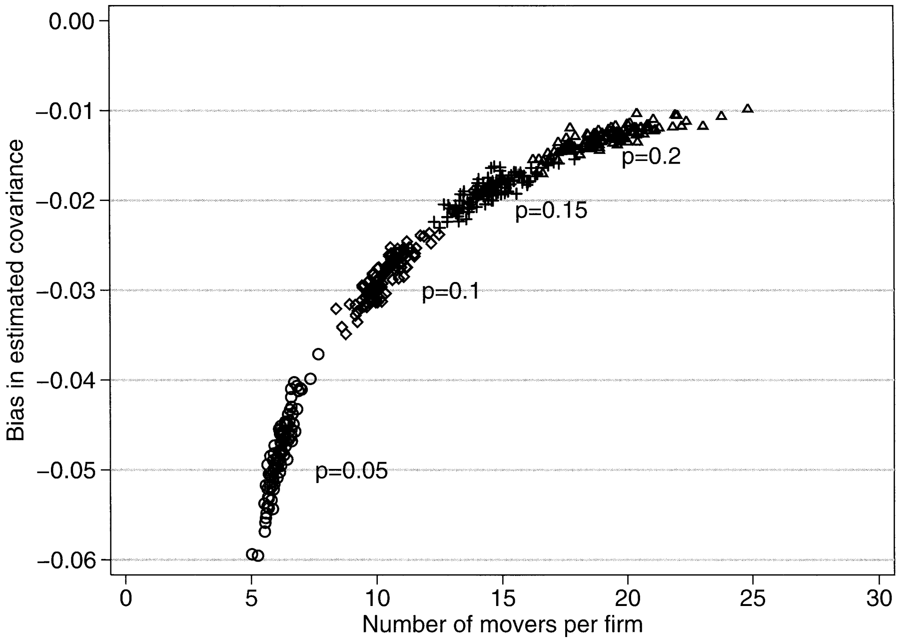

Fig. 1. Varying p: bias in covariance

were to compare one data set involving $J+1$ moves between $J+1$ firms, and another involving an additional $J+1$ moves between an extra $J+1$ firms. It is easy to show that the bias does not necessarily fall. In other words, the formula is not homogeneous of degree 0 in M and J, as might be expected.

This simple example is only illustrative. To investigate limited mobility bias more extensively we also ran simulations of the random-effects data generation process. A summary is given in Appendix B. Fig. 1 is similar to the simple relationship that is given in equation (27).

4. Simulations

The main focus of the paper is to analyse the fixed effects data generation process. In this section, we run simulations to see whether the bias-corrected correlation really is unbiased (Section 3.4 above). In addition, using the same design, we then simulate a random-effects data generation process, by letting θ_i and ψ_j vary over replications, to see whether the same formulae correct the biases when the model is estimated by the usual LSDV estimator.

4.1. The simulation design

The simulated data mimic the generic model that was outlined in Section 2. For both data generation processes, J firms are created indexed $j=1,\ldots,J$, each with a random number of employees N_j drawn from a log-normal distribution with mean μ_N. We have a balanced panel where each employee is observed for T periods. Each firm is given a realization of $\mathbf{w}_{jt}$ and ψ_j; each worker is given a realization of $\mathbf{x}_{it}$ and θ_i. We use one variable of each type; hence w_{jt} and x_{it} are scalars. These realizations are drawn from a joint normal distribution with the following means and covariance structure for any period t:

$$\begin{pmatrix} x_{it} \\ w_{jt} \\ \theta_i \\ \psi_j \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & 0 & \sigma_{x\theta} & \sigma_{x\psi} \\ & \sigma_w^2 & \sigma_{w\theta} & \sigma_{w\psi} \\ & & \sigma_\theta^2 & \sigma_{\theta\psi} \\ & & & \sigma_\psi^2 \end{pmatrix} \right\}. \tag{28}$$

This structure focuses on the correlation between the unobservables and the observables, and the correlation between the unobservables themselves. 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 auto-correlated, with parameter 0.9. 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 ($\sigma_{\theta\psi} \neq 0$), as well as between the unobserved components and both firm and worker time varying effects.

The first draw of $(x_{it}, w_{jt}, \theta_i, \psi_j)$ at $t = 1$ ensures that workers with certain characteristics are matched with firms with certain characteristics. For example, if $\sigma_{\theta\psi} > 0$ 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 $(\psi_{j'}, w_{j't})$. To ensure that a new match is drawn with a probability that is proportional to firm size, the list of new firms is weighted by the size of the firm.

The probability of movement from j to j', which is denoted m^*_{it}, is determined by a random draw from a uniform distribution. A move occurs if m^* is greater than some critical percentile of the distribution of m^*, which is denoted m^*_{c}, such that the probability of movement $p \equiv \Pr(m^* > m^*_{\mathrm{c}})$ is set at, for example, 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 $\psi_{j'}$ 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 emphasize that the assumption of random mobility is innocuous. So long as equation (2) holds, any model of mobility will generate simulations with similar properties. We choose random mobility because it means that we do not have to choose specific models about how movement occurs, namely to specify precisely how movement depends on θ_i and ψ_j.

Once the identity of each firm has been 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 data set 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.

As noted above, there are two different versions of the data generation process. In the simulations, they manifest themselves as follows. For the fixed effects data generation process, in equation (3) ε is drawn $r = 1, \ldots, R$ times, with $\mathbf{Z}$, $\mathbf{D}$, $\mathbf{F}$, $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ held fixed at their $r = 1$ values. In the random-effects data generation process, ε, $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ are drawn R times.

4.2. *Is the bias-corrected correlation unbiased?*

In Section 3, we established the properties of the estimators of interest and how the biases can be corrected. The only thing that we could not establish is whether the bias-corrected correlation $\tilde{\rho}$ is approximately unbiased. To design an appropriate simulation by using the fixed effects data generation process, we hold the number of firms J fixed as $N \to \infty$, rather than let $J \to \infty$. This means that the average number of workers per firm increases with N. To do this, we vary the mean of the underlying log-normal distribution of firm size by conducting five experiments: $\mu_N = 6, 8, 10, 12, 14$.

The parameters of the simulation are given in Table 1. Each column describes an experiment. Keeping J exactly fixed is not possible because we discard all observations that do not belong to the biggest group; it is well known that the 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). As N increases, so does the total number of worker-years N^* (as T is fixed at 5) and the number of movers M (as the probability of moving is fixed at 0.1).

Because this is a fixed effects data generation process, we draw θ_i and ψ_j once per experiment. The true values of the underlying distribution σ^2_θ, σ^2_ψ, $\sigma_{\theta\psi}$ and $\rho_{\theta\psi}$ are also given in Table 1 (see also equation (28)); the actual values $\sigma^2_{\theta N}$, $\sigma^2_{\psi N}$, $\sigma_{\theta\psi N}$ and $\rho_{\theta\psi N}$ vary from experiment to experiment. The crucial parameter is the correlation between θ and ψ, which is chosen to be positive ($\rho_{\theta\psi} = 0.246$): unobservably high wage workers work for unobservably high wage firms. We also assume a positive correlation between each unobservable and both time varying observables: the other four correlations in equation (28) are $\rho_{\theta x} = 0.295$, $\rho_{\theta w} = 0.160$, $\rho_{\psi x} = 0.082$ and $\rho_{\psi w} = 0.299$. 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 evidence from real linked employer–employee data (see Section 1).

For each experiment, ε_{it} is drawn R times (replications). Everything else is held fixed. In the analysis reported in this section, there are no other observed covariates $\mathbf{Z}$ in the data generation process, which means that the covariance is unambiguously biased downwards. For each data set we estimate equation (3) by using FEiLSDVj.

We then compute the four bias-uncorrected estimators $\tilde{\sigma}^2_\theta$, $\tilde{\sigma}^2_\psi$, $\tilde{\sigma}_{\theta\psi}$ and $\tilde{\rho}_{\theta\psi}$, by using equations (9)–(11). The mean of the bias in the sampling distribution for $\tilde{\sigma}^2_\theta$ is computed as

$$\frac{1}{R}\sum_r \tilde{\sigma}^2_{\theta,r} - \sigma^2_{\theta_N} \qquad r = 1, \ldots, R,$$

together with its estimated standard error. There are similar expressions for $\tilde{\sigma}^2_\psi$, $\tilde{\sigma}_{\theta\psi}$ and $\tilde{\rho}_{\theta\psi}$. See the four rows labelled 'Estimates' in Table 1. The four rows labelled 'Mean bias' show that all four estimators are biased in the directions expected, as shown algebraically in Section 3. For example, for $\mu_N = 10$, the estimated variance of the worker unobservables is 0.634 and is almost twice as big as the true variance of the worker unobservables, namely 0.321. However, the estimated variance of the firm unobservables is not biased by as much, 0.525 being 0.131 bigger than 0.394. The covariance is biased downwards, because a true positive covariance of 0.090 is estimated as −0.023. The sign of these three biases, taken together, implies that a true positive correlation is always biased downwards. For example, in the fourth column ($\mu_N = 10$), the bias is −0.293 with a 95% confidence interval that easily excludes 0. In fact, for all five columns, these biases are quite considerable: for all except the rightmost column, a positive correlation of about 0.25 is estimated as negative.

The formulae that bias-correct the four estimators are given in equations (18), (20), (24) and (25). These formulae can only be computed for moderate sample sizes, which is why N does not exceed about 12000. (More realistic sample sizes are used when simulating the random-effects data generation process, as reported in Appendix B.)

The four rows of Table 1 that are labelled 'Mean bias-corrected bias' show that the bias correction formulae are correct as all entries are insignificant and small in size. In particular, Table 1 demonstrates that the resulting bias-corrected correlation $\hat{\rho}_{\theta\psi}$ is approximately unbiased. It is this result that this simulation demonstrates; as it is based on moderately sized samples, it will

Table 1. Is the bias-corrected correlation unbiased?: fixed effects data generation process; $R = 1000$ replications

Log-normal μ_N	6	8	10	12	14
Number of individuals N	4650	6458	8262	10078	12027
Number of firms J	650	724	743	745	778
Number of time periods T	5	5	5	5	5
Average firm size†	7.15	8.91	11.1	13.5	15.45
Total number of observations N^*	23250	32290	41310	50390	60135
Number of movers M	1996	2663	3381	4129	4985
Number of movers per firm M/J	3.071	3.532	4.550	5.542	6.407
Proportion moving each period	0.107	0.103	0.102	0.102	0.103
True values					
$\sigma^2_{\theta N}$ $(\sigma^2_\theta = 0.3)$	0.3125	0.3135	0.3206	0.3019	0.3124
$\sigma^2_{\psi N}$ $(\sigma^2_\psi = 0.3)$	0.3569	0.2671	0.3939	0.2410	0.3568
$\sigma_{\theta\psi N}$ $(\sigma_{\theta\psi} = 0.0737)$	0.0742	0.0717	0.0900	0.0485	0.0883
$\rho_{\theta\psi N}$ $(\rho_{\theta\psi} = 0.2457)$	0.2223	0.2478	0.2533	0.1799	0.2645
Estimates					
$R^{-1}\Sigma_r(\tilde{\sigma}^2_{\theta,r})$	0.6815 (0.0210)	0.6563 (0.0154)	0.6335 (0.0127)	0.5907 (0.0103)	0.5916 (0.0095)
$R^{-1}\Sigma_r(\tilde{\sigma}^2_{\psi,r})$	0.5520 (0.0273)	0.4325 (0.0221)	0.5250 (0.0199)	0.3438 (0.0135)	0.4483 (0.0153)
$R^{-1}\Sigma_r(\tilde{\sigma}_{\theta\psi,r})$	−0.0939 (0.0172)	−0.0716 (0.0125)	−0.0232 (0.0107)	−0.0399 (0.0075)	0.0096 (0.0077)
$R^{-1}\Sigma_r(\tilde{\rho}_{\theta\psi,r})$	−0.1525 (0.0236)	−0.1340 (0.0202)	−0.0398 (0.0178)	−0.0883 (0.0150)	0.0188 (0.0152)
Mean bias					
$R^{-1}\Sigma_r(\tilde{\sigma}^2_{\theta,r}) - \sigma^2_{\theta N}$	0.3690 (0.0210)	0.3428 (0.0154)	0.3129 (0.0127)	0.2888 (0.0103)	0.2792 (0.0095)
$R^{-1}\Sigma_r(\tilde{\sigma}^2_{\psi,r}) - \sigma^2_{\psi N}$	0.1951 (0.0273)	0.1654 (0.0221)	0.1311 (0.0199)	0.1028 (0.0135)	0.0915 (0.0153)
$R^{-1}\Sigma_r(\tilde{\sigma}_{\theta\psi,r}) - \sigma_{\theta\psi N}$	−0.1681 (0.0172)	−0.1433 (0.0125)	−0.1132 (0.0107)	−0.0884 (0.0075)	−0.0787 (0.0077)
$R^{-1}\Sigma_r(\tilde{\rho}_{\theta\psi,r}) - \rho_{\theta\psi N}$	−0.3748 (0.0236)	−0.3818 (0.0202)	−0.2931 (0.0178)	−0.2682 (0.0150)	−0.2457 (0.0152)
Mean bias-corrected bias					
$R^{-1}\Sigma_r(\hat{\sigma}^2_{\theta,r}) - \sigma^2_{\theta N}$	0.00057 (0.0206)	−0.00004 (0.0153)	−0.00023 (0.0125)	0.00029 (0.0102)	0.00011 (0.0095)
$R^{-1}\Sigma_r(\hat{\sigma}^2_{\psi,r}) - \sigma^2_{\psi N}$	−0.00130 (0.0273)	0.00017 (0.0220)	−0.00009 (0.0199)	−0.00050 (0.0134)	−0.00050 (0.0153)
$R^{-1}\Sigma_r(\hat{\sigma}_{\theta\psi,r}) - \sigma_{\theta\psi N}$	0.00037 (0.0171)	−0.00048 (0.0125)	0.00003 (0.0107)	0.00011 (0.0074)	0.00030 (0.0077)
$R^{-1}\Sigma_r(\hat{\rho}_{\theta\psi,r}) - \rho_{\theta\psi N}$	0.00475 (0.0623)	0.00100 (0.0544)	0.00142 (0.0373)	0.00161 (0.0330)	0.00177 (0.0284)

†Average firm size is always an overestimate of μ_N because of the bias that is associated with estimating non-linear functions. This is unimportant as μ_N is simply a device for varying N_j. $\sigma^2_\varepsilon = 1$ for all simulations.

be more generally true for larger data sets. It is also true when we repeat all of the above for simulations where the number of firms gets bigger as N gets bigger.

4.3. Does a random-effects data generation process make any difference?

We cannot develop, in this paper, bias correction formulae for a random-effects data generation process as we do for a fixed effects data generation process in Section 3. However, our intuition is that the same formulae apply, and so in this subsection we demonstrate that this is indeed so by amending the simulations that were reported immediately above.

The only difference between the two data generation processes is that $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ are redrawn each replication in the random-effects data generation process. Note that $\mathbf{D}$ and $\mathbf{F}$ are fixed across replications. The results are given in Table 2. The same data set for N, J, T, $\mathbf{D}$ and $\mathbf{F}$ is being used in both tables. The only difference is that we have one draw of $\sigma^2_{\theta N}$, $\sigma^2_{\psi N}$, $\sigma_{\theta\psi N}$ and $\rho_{\theta\psi N}$ in Table 1, but 1000 draws in Table 2. Thus the mean of these draws is much closer to the true 'superpopulation' value than the one draw in the fixed effects data generation process.

The next four rows of the two tables differ because of comparing estimates that are based on one draw with averages of estimates across 1000 draws. However, we can see that the mean bias is very similar in both tables. In other words, the fixed effects estimator that is applied to both data generation processes gives very similar biases. This is why the bias correction formulae have the same effect in both tables, which is seen by comparing the final blocks. Here, instead of fixing on one draw of θ and ψ, we are averaging over R draws. Again, this is also true when we repeat all of the above for simulations where the number of firms grows increasingly large. The implication here is that the investigator can use the same bias correction formulae without having to worry about how he thinks the data have been generated.

5. An example using German linked data

To illustrate how a downward-biased estimate of $\tilde{\rho}_{\theta\psi}$ can be corrected, we use data from a linked worker–firm data set LIAB, which was made available by the IAB. The firm data comprise a panel of 4376 establishments (or 'plants') from the former West Germany observed over the period 1993–1997. The worker data comprise a panel of $N = 1930260$ workers who are employed in these plants. A common establishment identifier is available in both data sets, allowing them to be linked. (Kölling (2000) provides more information on the IAB establishment panel, Bender *et al.* (2000) has details on the worker data and Alda *et al.* (2005) has details on the linked data.) After eliminating observations with missing or incomplete information, the resulting linked data set has $N^* = 5145098$ worker-year observations. 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 linked IAB data have few or no movers between other plants in the data. This is because the plant data are a survey, and because the data set is relatively small in the T dimension. There are 1821 plants (out of the total of 4376) who have positive turnover.

Note that N is approximately 2 million, and so we cannot compute the bias-corrected estimates because of having to invert $\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D}$ in equation (18) and $\mathbf{D}^{\mathrm{T}}\mathbf{M}_{\mathbf{Z}}\mathbf{D}$ in equation (24). We therefore must asume that $\mathbf{Z}$ is orthogonal to $\mathbf{D}$ and $\mathbf{F}$. Substituting $\tilde{\mathbf{Q}}_{\mathbf{D}} = \mathbf{M}_{\mathbf{D}}$ into the bias expressions in equations (16) and (20) gives

$$\frac{\sigma^2_{\varepsilon}}{N^*}[N - 1 + \mathrm{tr}\{\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{P}_{\mathbf{D}}\mathbf{F}(\mathbf{F}^{\mathrm{T}}\mathbf{M}_{\mathbf{D}}\mathbf{F})^{-1}\}] \tag{29}$$

and

Table 2. Is the bias-corrected correlation unbiased?: random-effects data generation process; $R = 1000$ replications

Log-normal μ_N	6	8	10	12	14
Number of individuals N	4650	6458	8262	10078	12027
Number of firms J	650	724	743	745	778
Number of time periods T	5	5	5	5	5
Average firm size†	7.15	8.91	11.1	13.5	15.45
Total number of observations N^*	23250	32290	41310	50390	60135
Number of movers M	1996	2663	3381	4129	4985
Number of movers per firm M/J	3.071	3.532	4.550	5.542	6.407
Proportion moving each period	0.107	0.103	0.102	0.102	0.103
True values					
$\sigma^2_{\theta N}$ $(\sigma^2_\theta = 0.3)$	0.2990 (0.0077)	0.2994 (0.0071)	0.2996 (0.0065)	0.2997 (0.0058)	0.2997 (0.0053)
$\sigma^2_{\psi N}$ $(\sigma^2_\psi = 0.3)$	0.2944 (0.0527)	0.2959 (0.0532)	0.2957 (0.0503)	0.2963 (0.0449)	0.2982 (0.0433)
$\sigma_{\theta\psi N}$ $(\sigma_{\theta\psi} = 0.0737)$	0.0709 (0.0135)	0.0719 (0.0134)	0.0724 (0.0126)	0.0727 (0.0113)	0.0728 (0.0109)
$\rho_{\theta\psi N}$ $(\rho_{\theta\psi} = 0.2457)$	0.2380 (0.0235)	0.2405 (0.0220)	0.2422 (0.0203)	0.2431 (0.0184)	0.2426 (0.0179)
Estimates					
$R^{-1}\Sigma_r(\tilde{\sigma}^2_{\theta,r})$	0.6688 (0.0215)	0.6416 (0.0172)	0.6136 (0.0140)	0.5883 (0.0117)	0.5788 (0.0101)
$R^{-1}\Sigma_r(\tilde{\sigma}^2_{\psi,r})$	0.4916 (0.0589)	0.4608 (0.0587)	0.4269 (0.0532)	0.4005 (0.0475)	0.3900 (0.0455)
$R^{-1}\Sigma_r(\tilde{\sigma}_{\theta\psi,r})$	−0.0998 (0.0215)	−0.0705 (0.0179)	−0.0410 (0.0164)	−0.0161 (0.0140)	−0.0061 (0.0127)
$R^{-1}\Sigma_r(\tilde{\rho}_{\theta\psi,r})$	−0.1729 (0.0397)	−0.1313 (0.0363)	−0.0817 (0.0341)	−0.0345 (0.0296)	−0.0140 (0.0269)
Mean bias					
$R^{-1}\Sigma_r(\tilde{\sigma}^2_{\theta,r}) - \sigma^2_{\theta N}$	0.3698 (0.0200)	0.3421 (0.0154)	0.3139 (0.0123)	0.2887 (0.0100)	0.2791 (0.0086)
$R^{-1}\Sigma_r(\tilde{\sigma}^2_{\psi,r}) - \sigma^2_{\psi N}$	0.1972 (0.0270)	0.1650 (0.0210)	0.1312 (0.0184)	0.1042 (0.0151)	0.0912 (0.0135)
$R^{-1}\Sigma_r(\tilde{\sigma}_{\theta\psi,r}) - \sigma_{\theta\psi N}$	−0.1694 (0.0169)	−0.1424 (0.0126)	−0.1134 (0.0103)	−0.0888 (0.0080)	−0.0789 (0.0068)
$R^{-1}\Sigma_r(\tilde{\rho}_{\theta\psi,r}) - \rho_{\theta\psi N}$	−0.4109 (0.0275)	−0.3717 (0.0228)	−0.3239 (0.0211)	−0.2776 (0.0180)	−0.2566 (0.0156)
Mean bias-corrected bias					
$R^{-1}\Sigma_r(\hat{\sigma}^2_{\theta,r}) - \sigma^2_{\theta N}$	0.00138 (0.0198)	−0.00054 (0.0152)	0.00070 (0.0121)	0.00018 (0.0098)	−0.00004 (0.0085)
$R^{-1}\Sigma_r(\hat{\sigma}^2_{\psi,r}) - \sigma^2_{\psi N}$	0.00080 (0.0270)	−0.00019 (0.0209)	−0.00006 (0.0184)	0.00086 (0.0151)	−0.00025 (0.0135)
$R^{-1}\Sigma_r(\hat{\sigma}_{\theta\psi,r}) - \sigma_{\theta\psi N}$	−0.00082 (0.0169)	0.00039 (0.0125)	−0.00017 (0.0103)	−0.00027 (0.0080)	0.00022 (0.0068)
$R^{-1}\Sigma_r(\hat{\rho}_{\theta\psi,r}) - \rho_{\theta\psi N}$	0.00082 (0.0710)	0.00446 (0.0534)	0.00089 (0.0433)	−0.00033 (0.0336)	0.00169 (0.0284)

†Average firm size is always an overestimate of μ_N because of the bias that is associated with estimating non-linear functions. This is unimportant as μ_N is simply a device for varying N_j. $\sigma^2_\varepsilon = 1$ for all simulations.

$$\frac{\sigma_\varepsilon^2}{N^*}\mathrm{tr}\{\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}(\mathbf{F}^{\mathrm{T}}\mathbf{M}_{\mathbf{D}}\mathbf{F})^{-1}\} \tag{30}$$

for $\tilde{\sigma}_\theta^2$ and $\tilde{\sigma}_\psi^2$ respectively. (To derive expression (30), we need to use a partitioned inverse identity: see Abadir and Magnus (2005), page 106.) Equation (23) gives the bias for $\tilde{\sigma}_{\theta\psi}$. Now the largest matrix to be inverted is $J \times J$.

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 (e.g. gender and industry). The model is estimated by a classical minimum distance method that very closely approximates FEiLSDVj (see Andrews *et al.* (2006) for further details and how the method is implemented in Stata). This model is reported in full in Andrews *et al.* (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 dummy variables, i.e. for the 1821 plants who have turnover, the estimated correlation between θ_i and ψ_j is -0.191 (see the second column of Table 3). This is consistent with the existing literature (see Section 1). Applying the bias correction, the correlation moves to -0.148, primarily because the covariance term moves from -0.00512 to -0.00363. Of the two explanations that were discussed in Section 1, 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 sizable bias. The actual correction to the bias, namely –0.043, is given in the bottom row of Table 3.

This is the main message of the paper. However, we still need to investigate two modelling issues that recur in these analyses. The first issue concerns the size of the bias, and whether it can be ameliorated by pooling 'small' plants into a single small 'superplant'. This often happens in the literature because the number of plants can be too many for the FEiLSDVj estimator. The second issue 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. The discussion below equation (7) suggests that, the more imprecise the estimates of ψ_j, the more biased is the correlation. Of the 1821 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 that have fewer than 30 movers into one superplant, and we estimate a model with just 212 identifiable plant effects.

When we re-estimate the model with only 212 plant effects (fourth column of Table 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 the second to fourth column (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 data set (about 60% of the sample) to have the same value of $\hat{\psi}_j$. We should also note that in this case the restriction that is implied by moving from the second to fourth column 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 subsample of movers (who effectively identify the parameters of the model) may be a non-random subsample. 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 the third column of Table 3 we therefore report estimates separately for movers, i.e. we use the 72253 observations for those workers who move between the 1821 plants. An F-test of parameter equality between movers

Table 3. Bias correction, wage regressions and LIAB data

	Results for all plants		*Results for high turnover plants*	
	Whole sample	*Movers subsample*	*Whole sample*	*Movers subsample*
Number of observations N^*	4883331	72253	5145098	62668
Number of workers N	1816368	23393	1930260	20313
Number of plants J	1821	1821	212	212
Number of movers M	23393	23393	20313	20313
Error variance σ_ε^2	0.00459	0.00720	0.00461	0.00742
Uncorrected estimates				
Variance of worker effects $\tilde{\sigma}_\theta^2$	0.05381	0.05747	0.10231	0.20250
Variance of plant effects $\tilde{\sigma}_\psi^2$	0.01339	0.01513	0.00290	0.00562
Covarience of worker–plant effects $\tilde{\sigma}_{\theta\psi}$	−0.00512	−0.00389	−0.00030	0.00597
Correlation of worker–plant effects $\tilde{\rho}_{\theta\psi}$	−0.191	−0.132	−0.017	0.177
Correction to bias				
Bias($\tilde{\sigma}_\theta^2$) (equation (29))	0.00320	0.00450	0.00180	0.00330
Bias($\tilde{\sigma}_\psi^2$) (equation (30))	0.00149	0.00235	0.00008	0.00092
Bias($\tilde{\sigma}_{\theta\psi}$) (equation (23))	−0.00149	−0.00217	−0.00008	−0.00089
Corrected estimates				
Variance of worker effects $\hat{\sigma}_\theta^2$	0.05061	0.05297	0.10050	0.19921
Variance of plant effects $\hat{\sigma}_\psi^2$	0.01190	0.01278	0.00283	0.00470
Covarience of worker–plant effects $\hat{\sigma}_{\theta\psi}$	−0.00363	−0.00171	−0.00022	0.00686
Correlation of worker–plant effects $\hat{\rho}_{\theta\psi}$	−0.148	−0.066	−0.013	0.224
Correction to bias				
$\tilde{\rho}_{\theta\psi} - \hat{\rho}_{\theta\psi}$	−0.043	−0.066	−0.004	−0.047

and non-movers subsamples rejects the null hypothesis easily (p-value 0). There is also evidence that movers have a different degree of assortative matching from that of 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 (third column of Table 3), and since the pooling of low turnover plants also reduces the bias (fourth column), it seems logical to look at the results for movers in high turnover plants (fifth column). 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 the fifth with the third column). 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 the fifth with the fourth column).

The lesson from all this is that estimates of the correlations of worker and plant effects are sensitive to modelling decisions as well as the statistical bias that was 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 results in even larger movements in the correlation. Finally, our preferred estimate of the correlation of -0.066 (third column of Table 3) is still negative though somewhat closer to 0 than others in the literature. This estimate is much closer to 0 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.

6. 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. For a fixed effects data generation process, we develop formulae for the biases for the components of the estimated correlation, show that the bias-corrected components are consistent and show that the bias-corrected correlation is approximately unbiased (by using simulations). We also show that the bias in the covariance term becomes more negative as the number of movers decreases: the so-called limited mobility bias.

Using simulations for a random-effects data generation process, we show that exactly the same formulae correct the biases. This means that the investigator can use the same bias correction formulae without having to worry about how he thinks the data have been generated.

Using more simulations for a random-effects data generation process, we further examine limited mobility bias, and we again 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 data set. 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 effects on the estimated correlation.

Acknowledgements

The authors thank the IAB for funding this research: in particular, Lutz Bellmann. We also thank Stefan Bender for helping to prepare the data. The views that are expressed in this paper are solely those of the authors and are not those of the IAB. The comments of participants at various presentations are gratefully acknowledged. These include the IAB, the Institute of Social and Economic Research at Essex, the Rheinisch-Westfälisches Institut für Wirtschaftsforschung in Essen, the 'Symposium of multisource databases' at the Universität Erlangen-Nürnberg in July 2004, the 2004 annual conference of the European Association of Labour Economists (Lisbon), the 2005 annual Work Pensions and Labour Economics Conference (York), the 2005 Conference of the Verein für Socialpolitik (Bonn) and the Department of Economics at Aberdeen, Erlangen-Nürnberg, Kent, Manchester and Warwick. The usual disclaimer applies.

All calculations were performed by using Stata 9/SE and all code is available on request. In the replications, to estimate equation (3) by using FEiLSDVj, we use the `a2reg` package ported to Stata by Amine Ouazad from the original Fortran code written by Robert Creecy. Software can be downloaded from `http://repository.ciser.cornell.edu/viewcvs-public/cg2/branches/stata`.

Appendix A: Algebraic details

A.1. Introduction

In Appendix A, proofs of the results that are stated in Section 3 of the paper are supplied, in particular relating to the consistency of the variance and covariance estimators.

The main problem is that the number of parameters being estimated increases with N and J, the dimensions of the parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ respectively. However, the variance estimators are scalar functions of $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$, and might be expected to be well behaved. The argument for the consistency of the estimators of $\sigma^2_{\theta N}$, $\sigma^2_{\psi N}$ and $\sigma_{\theta\psi N}$ relies on the use of Chebyshev's inequality. This seems natural because of the interest in the finite sample bias of the estimators $\tilde{\sigma}^2_\theta$, $\tilde{\sigma}^2_\psi$ and $\tilde{\sigma}_{\theta\psi}$. Equation (16) may be stated as

$$E(\tilde{\sigma}^2_\theta|\mathbf{Z}) = \sigma^2_{\theta N} + b_N(\mathbf{Z}),$$

where $b_N(\mathbf{Z})$ is the (conditional) bias, so the bias-corrected estimator $\hat{\sigma}^2_\theta$ has a conditional expectation that is independent of $\mathbf{Z}$:

$$E(\hat{\sigma}^2_\theta|\mathbf{Z}) = \sigma^2_{\theta N}.$$

In turn, the conditional variance $\text{var}(\tilde{\sigma}^2_\theta|\mathbf{Z}) = v_N(\mathbf{Z})$, say, is equal to $\text{var}(\hat{\sigma}^2_\theta|\mathbf{Z})$ and, if $v_N(\mathbf{Z}) < d_N$, almost surely, where d_N is a constant that is independent of $\mathbf{Z}$ such that $d_N \rightarrow 0$ as $N^* \rightarrow \infty$, then $E\{v_N(\mathbf{Z})\} = \text{var}(\hat{\sigma}^2_\theta) < d_N \rightarrow 0$ as well. Chebyshev's inequality then guarantees that $\hat{\sigma}^2_\theta - \sigma^2_{\theta N} \rightarrow^{\text{P}} 0$, or that $\hat{\sigma}^2_\theta \rightarrow^{\text{P}} \sigma^2_\theta$. However, the probability limit of $\tilde{\sigma}^2_\theta$ is not $\sigma^2_{\theta N}$, since the bias in this estimator does not vanish as $N^* \rightarrow \infty$.

The same arguments can be applied to $\tilde{\sigma}^2_\psi$ and $\tilde{\sigma}_{\theta\psi}$. Sufficient conditions for the various conditional expectations to exist are that $E(\tilde{\sigma}^2_\theta)^2 < \infty$ and $E(\tilde{\sigma}^2_\psi)^2 < \infty$. The assumptions below will ensure this.

A.2. Assumptions

Denote by $\mathbf{z}^{\text{T}}_{it}$ the row of $\mathbf{Z}$ for the tth year for the ith worker.

Assumption 1. The vectors $\mathbf{z}_{it}$ and ε_{it} are independent and identically distributed over $t = 1, \ldots, T_i$ and $i = 1, \ldots, N$, where $\mathbf{z}_{it}$ and ε_{it} are independent for a given i and t.

Assumption 2. $E(\varepsilon_{it}) = 0$, $E(\varepsilon^2_{it}) = \sigma^2_\varepsilon < \infty$, $E(\varepsilon^3_{it}) = 0$ and $E(\varepsilon^4_{it}) = 3\sigma^4_\varepsilon$. These would follow from assuming that ε_{it} are independent and identically distributed $N(0, \sigma^2_\varepsilon)$, although the normality is not used.

Assumption 3. The data matrix $\mathbf{V} \equiv (\mathbf{Z}, \mathbf{D}, \mathbf{F})$, $N^* \times (K + N + J)$, is full column rank for any combination of N^*, N and J.

Assumption 4. For $\mathbf{P_Z} = \mathbf{Z}(\mathbf{Z}^{\text{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\text{T}}$, $\mathbf{D}^{\text{T}}\mathbf{P_Z}\mathbf{D}$ and $\mathbf{F}^{\text{T}}\mathbf{P_Z}\mathbf{F}$ are positive definite, for every combination of N^*, N and J.

Assumption 5. N and J are allowed to go to ∞, but in such a way that $J/N \rightarrow k$ and $N/N^* \rightarrow l$, where l is the limit of

$$\left(\frac{1}{N}\sum_{i=1}^{N} T_i\right)^{-1}.$$

The term in the parentheses is the 'average number of years'.

Assumption 6. The maximum number of years for a worker satisfies $\max_i(T_i) < T_{\max} < \infty$, and the maximum worker-years at any firm j satisfies $\max_j(N_j) < N_{\max} < \infty$. This latter assumption is required to show that $E(|\tilde{\sigma}^4_\psi|) < \infty$.

Assumption 7. The 'population' variances and covariance $\sigma^2_{\theta N}$, $\sigma^2_{\psi N}$ and $\sigma_{\theta\psi N}$ that are defined in equation (12) converge to limits σ^2_θ, σ^2_ψ and $\sigma_{\theta\psi}$ respectively as $N^* \rightarrow \infty$.

A consequence of assumption 2 is that, if $\boldsymbol{\varepsilon}$ is the $N^* \times 1$ vector with typical element ε_{it}, $E(\boldsymbol{\varepsilon}^{\text{T}}\boldsymbol{\varepsilon})^2 \leqslant N^*$, and $E(\varepsilon^4_{it}) = N^* \times 3\sigma^4_\varepsilon < \infty$, which in turn implies that

$$E(\boldsymbol{\varepsilon}^{\text{T}}\boldsymbol{\varepsilon})^{3/2} < \infty,$$
$$E(\boldsymbol{\varepsilon}^{\text{T}}\boldsymbol{\varepsilon}) < \infty,$$
$$E(\boldsymbol{\varepsilon}^{\text{T}}\boldsymbol{\varepsilon})^{1/2} < \infty,$$

via Holder's inequality.

Assumptions 1–6 imply that $E(\tilde{\sigma}_\theta^2)^2 < \infty$ and $E(\tilde{\sigma}_\psi^2)^2 < \infty$, and, by the Cauchy–Schwartz inequality (see Appendix A.4, part (a)(ii) below), $E(\tilde{\sigma}_{\theta\psi})^2 < \infty$. $E(\tilde{\sigma}_\theta^2)^2 < \infty$ implies that $E(\tilde{\sigma}_\theta^2|\mathbf{Z})$ and $E(\tilde{\sigma}_\theta^4|\mathbf{Z})$ exist almost surely, with corresponding assertions for $\tilde{\sigma}_\psi^2$ and $\tilde{\sigma}_{\theta\psi}$. Proofs are available on request.

A.3. Some notation

To compress long expressions it is convenient to define the matrices

$$\begin{aligned} &\mathbf{R_D} = \mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D}, \qquad \mathbf{S_D} = \mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}, \qquad \mathbf{V_D} = \mathbf{D}^{\mathrm{T}}\mathbf{M_Z}\mathbf{D}, \qquad \mathbf{W_D} = \mathbf{D}^{\mathrm{T}}\mathbf{D}, \\ &\mathbf{R_F} = \mathbf{F}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{D}}\mathbf{F}, \qquad \mathbf{S_F} = \mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}, \qquad \mathbf{V_F} = \mathbf{F}^{\mathrm{T}}\mathbf{M_Z}\mathbf{F}, \qquad \mathbf{W_V} = \mathbf{F}^{\mathrm{T}}\mathbf{F}. \end{aligned} \tag{31}$$

The sampling errors for $\hat{\theta}$ and $\hat{\psi}$ can be compressed by using matrices $\mathbf{G}_\theta$ and $\mathbf{G}_\psi$:

$$\begin{aligned} \hat{\theta} &= \theta + (\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\varepsilon = \theta + \mathbf{G}_\theta\varepsilon, \\ \hat{\psi} &= \psi + (\mathbf{F}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{D}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{D}}\varepsilon = \psi + \mathbf{G}_\psi\varepsilon. \end{aligned} \tag{32}$$

A.4. Bias and consistency

(a) Some matrix inequalities are required.

(i) A basic tool is given by Abadir and Magnus (2005), exercise 8.29, page 222. If $\mathbf{A}$ is a positive semidefinite matrix and $\mathbf{B}$ a positive definite matrix, then the eigenvalues of $(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}$ satisfy $0 \leqslant \lambda_i < 1$. If $\mathbf{A}$ is positive definite, then $\lambda_i > 0$. Rephrased in terms of data matrices $(\mathbf{Y}, \mathbf{X})$, if $\mathbf{A} = \mathbf{Y}^{\mathrm{T}}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} = \mathbf{Y}^{\mathrm{T}}\mathbf{P_X}\mathbf{Y}$ and $\mathbf{B} = \mathbf{Y}^{\mathrm{T}}\mathbf{Y} - \mathbf{Y}^{\mathrm{T}}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} = \mathbf{Y}^{\mathrm{T}}\mathbf{M_X}\mathbf{Y}$, then

$$(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A} = (\mathbf{Y}^{\mathrm{T}}\mathbf{Y})^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{P_X}\mathbf{Y},$$

$$(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B} = (\mathbf{Y}^{\mathrm{T}}\mathbf{Y})^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{M_X}\mathbf{Y},$$

with the latter matrix also having eigenvalues in the interval $(0, 1)$. Provided that $\mathbf{Y}^{\mathrm{T}}\mathbf{P_X}\mathbf{Y}$ and $\mathbf{Y}^{\mathrm{T}}\mathbf{M_X}\mathbf{Y}$ are positive definite, the inverses of $(\mathbf{Y}^{\mathrm{T}}\mathbf{Y})^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{P_X}\mathbf{Y}$ and $(\mathbf{Y}^{\mathrm{T}}\mathbf{Y})^{-1}\mathbf{Y}^{\mathrm{T}}\mathbf{M_X}\mathbf{Y}$ have finite eigenvalues in the interval $(1, \infty)$.

(ii) The Cauchy–Schwartz inequality for n vectors $\mathbf{x}$ and $\mathbf{y}$ states that

$$(\mathbf{x}^{\mathrm{T}}\mathbf{y})^2 \leqslant (\mathbf{x}^{\mathrm{T}}\mathbf{x})(\mathbf{y}^{\mathrm{T}}\mathbf{y})$$

(Abadir and Magnus (2005), exercise 1.9, page 7) and has a form for traces

$$\operatorname{tr}(\mathbf{A}^{\mathrm{T}}\mathbf{B})^2 \leqslant \operatorname{tr}(\mathbf{A}^{\mathrm{T}}\mathbf{A})\operatorname{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{B}),$$

for $\mathbf{A}$ and $\mathbf{B}$ of the same order (Abadir and Magnus (2005), page 325). There is also a version for random variables X and Y (Karr (1992), page 120):

$$E(|XY|) \leqslant \sqrt{E(X^2)}\sqrt{E(Y^2)}.$$

(iii) Suppose that $\mathbf{H}$ is a positive semidefinite $N \times N$ matrix with maximum root $\lambda_{\max}(\mathbf{H})$. Then

$$\operatorname{tr}(\mathbf{H}) \leqslant N\,\lambda_{\max}(\mathbf{H}).$$

(b) The ideas in part (a)(i) must be adapted to deal with expressions like the bias term in equation (16), $\operatorname{tr}\{(\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\} = \operatorname{tr}(\mathbf{R_D}^{-1}\mathbf{S_D})$, where $\tilde{\mathbf{Q}}_{\mathbf{F}} = \mathbf{M_Z}\tilde{\mathbf{M}}_{\mathbf{F}}\mathbf{M_Z}$, $\tilde{\mathbf{F}} = \mathbf{M_Z}\mathbf{F}$ and $\tilde{\mathbf{D}} = \mathbf{M_Z}\mathbf{D}$. Here $\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D} = \tilde{\mathbf{D}}^{\mathrm{T}}\mathbf{M_F}\tilde{\mathbf{D}}$ but $\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}$ cannot be expressed in terms of $\tilde{\mathbf{D}}$. Using the notation from equation (31), the eigenvalues of $\mathbf{R_D}^{-1}\mathbf{S_D}$ are also the eigenvalues of $\mathbf{S_D}^{1/2}\mathbf{R_D}^{-1}\mathbf{S_D}^{1/2}$, which are also the eigenvalues of $\mathbf{S_D}^{1/2}\mathbf{V_D}^{-1/2}(\mathbf{V_D}^{1/2}\mathbf{R_D}^{-1}\mathbf{V_D}^{1/2})\mathbf{V_D}^{-1/2}\mathbf{S_D}^{1/2}$, where positive semidefinite square-root matrices are being used. The result that is given in Magnus and Neudecker (1988), exercise 12, page 237, is as follows. If $\mathbf{V}$ is a positive semidefinite matrix, and $\mathbf{A}$ a rectangular matrix, then

$$\lambda_{\max}(\mathbf{A}\mathbf{V}\mathbf{A}^{\mathrm{T}}) \leqslant \lambda_{\max}(\mathbf{V})\,\lambda_{\max}(\mathbf{A}\mathbf{A}^{\mathrm{T}}),$$

where $\lambda_{\max}(\mathbf{A})$ denotes the largest eigenvalue of $\mathbf{A}$; then

$$\lambda_{\max}\{(\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\} \leqslant \lambda_{\max}(\mathbf{V_D}^{1/2}\mathbf{R_D}^{-1}\mathbf{V_D}^{1/2})\,\lambda_{\max}(\mathbf{S_D}^{1/2}\mathbf{V_D}^{-1}\mathbf{S_D}^{1/2}).$$

It follows from assumptions 3 and 4 that $1 < \lambda_{\max}(\mathbf{V}_\mathbf{D}^{1/2}\mathbf{R}_\mathbf{D}^{-1}\mathbf{V}_\mathbf{D}^{1/2}) < \infty$. The same argument can be used to show that

$$\lambda_{\max}(\mathbf{S}_\mathbf{D}^{1/2}\mathbf{V}_\mathbf{D}^{-1}\mathbf{S}_\mathbf{D}^{1/2}) \leqslant \lambda_{\max}(\mathbf{W}_\mathbf{D}^{1/2}\mathbf{V}_\mathbf{D}^{-1}\mathbf{W}_\mathbf{D}^{1/2})\,\lambda_{\max}(\mathbf{S}_\mathbf{D}^{1/2}\mathbf{W}_\mathbf{D}^{-1}\mathbf{S}_\mathbf{D}^{1/2}).$$

$\lambda_{\max}(\mathbf{S}_\mathbf{D}^{1/2}\mathbf{W}_\mathbf{D}^{-1}\mathbf{S}_\mathbf{D}^{1/2}) = 1$, and $1 < \lambda_{\max}(\mathbf{W}_\mathbf{D}^{1/2}\mathbf{V}_\mathbf{D}^{-1}\mathbf{W}_\mathbf{D}^{1/2}) < \infty$ follows from assumption 4.

(c) Combining parts (a) and (b), $\lambda_{\max}\{(\mathbf{D}^\mathrm{T}\tilde{\mathbf{Q}}_\mathbf{F}\mathbf{D})^{-1}\mathbf{D}^\mathrm{T}\mathbf{A}\mathbf{D}\} < c_\mathbf{D} < \infty$, with an analogous result for $\lambda_{\max}\{(\mathbf{F}^\mathrm{T}\tilde{\mathbf{Q}}_\mathbf{D}\mathbf{F})^{-1}\mathbf{F}^\mathrm{T}\mathbf{A}\mathbf{F}\} < c_\mathbf{F} < \infty$, where $c_\mathbf{D}$ and $c_\mathbf{F}$ are constants that are independent of $\mathbf{Z}$. As a result, $\mathrm{tr}\{(\mathbf{D}^\mathrm{T}\tilde{\mathbf{Q}}_\mathbf{F}\mathbf{D})^{-1}\mathbf{D}^\mathrm{T}\mathbf{A}\mathbf{D}\} \leqslant Nc_\mathbf{D} < \infty$ and

$$\frac{1}{N^*}\mathrm{tr}\{(\mathbf{D}^\mathrm{T}\tilde{\mathbf{Q}}_\mathbf{F}\mathbf{D})^{-1}\mathbf{D}^\mathrm{T}\mathbf{A}\mathbf{D}\} \leqslant \frac{N}{N^*}c_\mathbf{D} < lc_\mathbf{D} < \infty,$$

showing that the bias of $\tilde{\sigma}_\theta^2$ does not vanish as $N^* \to \infty$. Exactly the same argument can be applied to the bias term for $\tilde{\sigma}_\psi^2$—literally all that is needed is to interchange $\mathbf{D}$ and $\mathbf{F}$ above, to replace N by J, and to replace the bound $c_\mathbf{D}$ by $c_\mathbf{F}$:

$$\frac{1}{N^*}\mathrm{tr}\{(\mathbf{F}^\mathrm{T}\tilde{\mathbf{Q}}_\mathbf{D}\mathbf{F})^{-1}\mathbf{F}^\mathrm{T}\mathbf{A}\mathbf{F}\} \leqslant \frac{J}{N^*}c_\mathbf{F} \to klc_\mathbf{F} < \infty.$$

(d) For $\tilde{\sigma}_{\theta\psi}$, the Cauchy–Schwartz inequality in part (a) (ii) shows that the bias term of equation (21) is bounded above by a factor depending on N and J:

$$\begin{aligned}\mathrm{tr}(\tilde{\mathbf{Q}}_\mathbf{F}\mathbf{D}\mathbf{R}_\mathbf{D}^{-1}\mathbf{D}^\mathrm{T}\mathbf{A}\mathbf{A}\mathbf{F}\mathbf{R}_\mathbf{F}^{-1}\mathbf{F}^\mathrm{T}\tilde{\mathbf{Q}}_\mathbf{D}) &\leqslant \mathrm{tr}(\tilde{\mathbf{Q}}_\mathbf{F}\mathbf{D}\mathbf{R}_\mathbf{D}^{-1}\mathbf{S}_\mathbf{D}\mathbf{R}_\mathbf{D}^{-1}\mathbf{D}^\mathrm{T}\tilde{\mathbf{Q}}_\mathbf{F})^{1/2}\,\mathrm{tr}(\tilde{\mathbf{Q}}_\mathbf{D}\mathbf{F}\mathbf{R}_\mathbf{F}^{-1}\mathbf{S}_\mathbf{F}\mathbf{R}_\mathbf{F}^{-1}\mathbf{F}^\mathrm{T}\tilde{\mathbf{Q}}_\mathbf{D})^{1/2}\\ &< (NJc_\mathbf{D}c_\mathbf{F})^{1/2}.\end{aligned}$$

We can see that this trace expression will vanish on division by N^*, in the J fixed case, so that bias correction in estimating $\sigma_{\theta\psi}$ is not required asymptotically. In the case where $J \to \infty$ as $N^* \to \infty$,

$$\frac{1}{N^*}\mathrm{tr}(\tilde{\mathbf{Q}}_\mathbf{F}\mathbf{D}\mathbf{R}_\mathbf{D}^{-1}\mathbf{D}^\mathrm{T}\mathbf{A}\mathbf{A}\mathbf{F}\mathbf{R}_\mathbf{F}^{-1}\mathbf{F}^\mathrm{T}\tilde{\mathbf{Q}}_\mathbf{D}) \leqslant \left(\frac{N}{N^*}\frac{J}{N^*}c_\mathbf{D}c_\mathbf{F}\right)^{1/2}$$

and the bound converges to a limit as $N^* \to \infty$. Thus, bias correction will again be necessary for consistency.

(e) When there are no $\mathbf{Z}$s in the model, the bias expression for $\tilde{\sigma}_{\theta\psi}$ in equation (22) can be simplified. First, $\tilde{\mathbf{Q}}_\mathbf{D} = \mathbf{M}_\mathbf{D}$ and $\mathbf{M}_\mathbf{Z}\mathbf{D}(\mathbf{D}^\mathrm{T}\mathbf{M}_\mathbf{Z}\mathbf{D})^{-1}\mathbf{D}^\mathrm{T} = \mathbf{D}(\mathbf{D}^\mathrm{T}\mathbf{D})^{-1}\mathbf{D}^\mathrm{T} = \mathbf{P}_\mathbf{D}$, producing equation (23). In turn, the diagonal elements of $\mathbf{D}^\mathrm{T}\mathbf{D}$ and the elements of $\mathbf{D}^\mathrm{T}\mathbf{1}$ are the same so $\mathbf{P}_\mathbf{D}\mathbf{1} = \mathbf{1}$, and then $\mathbf{P}_\mathbf{D}\mathbf{A} = \mathbf{P}_\mathbf{D} - \mathbf{1}(\mathbf{1}^\mathrm{T}\mathbf{1})^{-1}\mathbf{1}^\mathrm{T} = \mathbf{A} - \mathbf{M}_\mathbf{D}$, making the bias expression

$$-\mathrm{tr}\{\mathbf{F}^\mathrm{T}(\mathbf{A}-\mathbf{M}_\mathbf{D})\mathbf{F}(\mathbf{F}^\mathrm{T}\mathbf{M}_\mathbf{D}\mathbf{F})^{-1}\} = -\mathrm{tr}\{\mathbf{F}^\mathrm{T}\mathbf{A}\mathbf{F}(\mathbf{F}^\mathrm{T}\mathbf{M}_\mathbf{D}\mathbf{F})^{-1}\} + J.$$

A.5. Variance and covariance estimators

All three estimators, $\tilde{\sigma}_\theta^2$, $\tilde{\sigma}_\psi^2$ and $\tilde{\sigma}_{\theta\psi}$, are quadratic forms in the data vector $\mathbf{y}$, with an asymmetric matrix in the case of $\tilde{\sigma}_{\theta\psi}$. Results for the variances of quadratic forms are needed, allowing for the asymmetry. These results are well known and can be found in a variety of sources (e.g. Magnus and Neudecker (1988), theorem 12, page 251). Writing $\mathbf{y} = \boldsymbol{\mu} + \varepsilon$, then the estimators in question can be written as $\mathbf{y}^\mathrm{T}\mathbf{H}\mathbf{y}$, where $\mathbf{H}$ may be symmetric or asymmetric. In the model of equation (4), given $\mathbf{Z}$, $\boldsymbol{\mu} = \mathbf{Z}\gamma + \mathbf{D}\boldsymbol{\theta} + \mathbf{F}\boldsymbol{\psi}$.

Then, under the assumptions,

$$\mathrm{var}(\mathbf{y}^\mathrm{T}\mathbf{H}\mathbf{y}) = \sigma_\varepsilon^2(2\boldsymbol{\mu}^\mathrm{T}\mathbf{H}^2\boldsymbol{\mu} + \boldsymbol{\mu}^\mathrm{T}\mathbf{H}\mathbf{H}^\mathrm{T}\boldsymbol{\mu} + \boldsymbol{\mu}^\mathrm{T}\mathbf{H}^\mathrm{T}\mathbf{H}\boldsymbol{\mu}) + \sigma_\varepsilon^4\,\mathrm{tr}(\mathbf{H}^2 + \mathbf{H}\mathbf{H}^\mathrm{T}).$$

When $\mathbf{H}$ is symmetric, this collapses to

$$\mathrm{var}(\mathbf{y}^\mathrm{T}\mathbf{H}\mathbf{y}) = 4\sigma_\varepsilon^2\boldsymbol{\mu}^\mathrm{T}\mathbf{H}^2\boldsymbol{\mu} + 2\sigma_\varepsilon^4\,\mathrm{tr}(\mathbf{H}^2). \tag{33}$$

The bias terms for the biased estimators $\tilde{\sigma}_\theta^2$, $\tilde{\sigma}_\psi^2$ and $\tilde{\sigma}_{\theta\psi}$ are non-stochastic, given $\mathbf{Z}$, so the variances of the biased and bias-corrected estimators $\hat{\sigma}_\theta^2$, $\hat{\sigma}_\psi^2$ and $\hat{\sigma}_{\theta\psi}$ are the same, conditional on $\mathbf{Z}$. So, the variance calculations are applied to the biased estimators.

(a) For $\tilde{\sigma}_\theta^2$, the expression $\hat{\boldsymbol{\theta}}^\mathrm{T}\mathbf{D}^\mathrm{T}\mathbf{A}\mathbf{D}\hat{\boldsymbol{\theta}}$ is equal to

$$\hat{\boldsymbol{\theta}}^\mathrm{T}\mathbf{D}^\mathrm{T}\mathbf{A}\mathbf{D}\hat{\boldsymbol{\theta}} = \mathbf{y}^\mathrm{T}\tilde{\mathbf{Q}}_\mathbf{F}\mathbf{D}\mathbf{R}_\mathbf{D}^{-1}\mathbf{S}_\mathbf{D}\mathbf{R}_\mathbf{D}^{-1}\mathbf{D}^\mathrm{T}\tilde{\mathbf{Q}}_\mathbf{F}\mathbf{y} = \mathbf{y}^\mathrm{T}\mathbf{H}\mathbf{y}$$

where the matrix $\mathbf{H}$ is symmetric. It is useful to note that

$$\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\boldsymbol{\mu} = \mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\,(\mathbf{Z}\gamma + \mathbf{D}\boldsymbol{\theta} + \mathbf{F}\boldsymbol{\psi}) = \mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D}\boldsymbol{\theta}$$

since $\mathbf{Z}$ and $\mathbf{F}$ are projected away by $\tilde{\mathbf{Q}}_{\mathbf{F}}$. The terms in equation (33) are

$$\boldsymbol{\mu}^{\mathrm{T}}\mathbf{H}^2\boldsymbol{\mu} = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{S}_{\mathbf{D}}\mathbf{R}_{\mathbf{D}}^{-1}\mathbf{S}_{\mathbf{D}}\boldsymbol{\theta} \leqslant \lambda_{\max}(\mathbf{R}_{\mathbf{D}}^{-1}\mathbf{S}_{\mathbf{D}})\boldsymbol{\theta}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\boldsymbol{\theta},$$

using the result that $\mathbf{x}^{\mathrm{T}}\mathbf{H}\mathbf{x} \leqslant \lambda_{\max}(\mathbf{H})\mathbf{x}^{\mathrm{T}}\mathbf{x}$ if $\mathbf{H}$ is positive semidefinite, and

$$\mathrm{tr}(\mathbf{H}^2) = \mathrm{tr}\{(\mathbf{D}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{F}}\mathbf{D})^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\}^2 \leqslant N\{\lambda_{\max}(\mathbf{R}_{\mathbf{D}}^{-1}\mathbf{S}_{\mathbf{D}})\}^2 \leqslant Nc_{\mathbf{D}}^2.$$

Overall,

$$\mathrm{var}(\tilde{\sigma}_\theta^2|\mathbf{Z}) \leqslant \frac{1}{N^*}\left(4\sigma_\varepsilon^2 c_{\mathbf{D}}\sigma_{\theta N}^2 + 2\sigma_\varepsilon^4\frac{N}{N^*}c_{\mathbf{D}}^2\right) \to 0 \qquad \text{as } N^* \to \infty, \text{ almost surely.}$$

This inequality will also be satisfied by the unconditional variance $\mathrm{var}(\hat{\sigma}_\theta^2)$, so $\hat{\sigma}_\theta^2$ is consistent for σ_θ^2.

(b) Just as in the bias case, the same arguments are needed for $\tilde{\sigma}_\psi^2$, with interchange of $\mathbf{F}$ for $\mathbf{D}$, and J for N, to give

$$\mathrm{var}(\hat{\sigma}_\psi^2|\mathbf{Z}) \leqslant \frac{1}{N^*}\left(4\sigma_\varepsilon^2 c_{\mathbf{F}}\sigma_{\psi N}^2 + 2\sigma_\varepsilon^4\frac{J}{N^*}c_{\mathbf{F}}^2\right),$$

which will go to 0 with N^*, even when $J \to \infty$, provided that assumption 5 holds.

(c) Variance calculations for $\tilde{\sigma}_{\theta\psi}$ are more complicated because the matrix of the quadratic form representing this estimator is asymmetric. The term $\boldsymbol{\mu}^{\mathrm{T}}\mathbf{H}^2\boldsymbol{\mu}$ can be bounded via the Cauchy–Schwartz inequality, so only the terms $\boldsymbol{\mu}^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}\boldsymbol{\mu}$ need to be examined. One can show that

$$\begin{aligned}
\boldsymbol{\mu}^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\mu} &= \boldsymbol{\theta}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{F}(\mathbf{F}^{\mathrm{T}}\tilde{\mathbf{Q}}_{\mathbf{D}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{D}\boldsymbol{\theta} \\
&= \boldsymbol{\theta}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{F}\mathbf{S}_{\mathbf{F}}^{-1/2}(\mathbf{S}_{\mathbf{F}}^{1/2}\mathbf{R}_{\mathbf{F}}^{-1}\mathbf{S}_{\mathbf{F}}^{1/2})\mathbf{S}_{\mathbf{F}}^{-1/2}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{D}\boldsymbol{\theta} \\
&\leqslant \lambda_{\max}(\mathbf{S}_{\mathbf{F}}^{1/2}\mathbf{R}_{\mathbf{F}}^{-1}\mathbf{S}_{\mathbf{F}}^{1/2})\boldsymbol{\theta}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{F}\mathbf{S}_{\mathbf{F}}^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{D}\boldsymbol{\theta} \\
&= \lambda_{\max}(\mathbf{R}_{\mathbf{F}}^{-1}\mathbf{S}_{\mathbf{F}})\boldsymbol{\theta}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{F}\mathbf{S}_{\mathbf{F}}^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{D}\boldsymbol{\theta}.
\end{aligned}$$

The matrix $\mathbf{A}\mathbf{F}\mathbf{S}_{\mathbf{F}}^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{A}$ is symmetric idempotent, with largest root 1, so

$$\boldsymbol{\theta}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{F}\mathbf{S}_{\mathbf{F}}^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{D}\boldsymbol{\theta} \leqslant \boldsymbol{\theta}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\boldsymbol{\theta}.$$

This makes

$$\boldsymbol{\mu}^{\mathrm{T}}\mathbf{H}\mathbf{H}^{\mathrm{T}}\boldsymbol{\mu} = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{F}\mathbf{R}_{\mathbf{F}}^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{D}\boldsymbol{\theta} \leqslant c_{\mathbf{F}}\boldsymbol{\theta}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\boldsymbol{\theta}.$$

A similar argument can be applied to $\boldsymbol{\mu}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}\boldsymbol{\mu}$:

$$\boldsymbol{\mu}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{H}\boldsymbol{\mu} \leqslant c_{\mathbf{D}}\boldsymbol{\psi}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}\boldsymbol{\psi}.$$

The term $\mathrm{tr}(\mathbf{H}^2)$ in equation (33) can be bounded by using Schur's inequality (Abadir and Magnus (2005), exercise 12.6, page 325):

$$\mathrm{tr}(\mathbf{H}^2) \leqslant \mathrm{tr}(\mathbf{H}^{\mathrm{T}}\mathbf{H}) = \mathrm{tr}(\mathbf{H}\mathbf{H}^{\mathrm{T}}).$$

In turn, the Cauchy–Schwartz inequality for traces can be used to make

$$\begin{aligned}
\mathrm{tr}(\mathbf{H}\mathbf{H}^{\mathrm{T}}) &= \mathrm{tr}(\mathbf{A}\mathbf{D}\mathbf{R}_{\mathbf{D}}^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{A}\mathbf{F}\mathbf{R}_{\mathbf{F}}^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{A}) \\
&\leqslant [\mathrm{tr}\{(\mathbf{R}_{\mathbf{D}}^{-1}\mathbf{S}_{\mathbf{D}})^2\}\,\mathrm{tr}\{(\mathbf{R}_{\mathbf{F}}^{-1}\mathbf{S}_{\mathbf{F}})^2\}]^{1/2} \\
&\leqslant (Nc_{\mathbf{D}}^2 Jc_{\mathbf{F}}^2)^{1/2}.
\end{aligned}$$

In conjunction with the previous components of $\mathrm{var}(\tilde{\sigma}_{\theta\psi}|\mathbf{Z})$, it follows that

$$\mathrm{var}(\tilde{\sigma}_{\theta\psi}|\mathbf{Z}) \leqslant \frac{1}{N^*}\left[\sigma_\varepsilon^2\,\{2(c_{\mathbf{D}}c_{\mathbf{F}}\sigma_{\theta N}^2\sigma_{\psi N}^2)^{1/2} + c_{\mathbf{F}}\sigma_{\theta N}^2 + c_{\mathbf{D}}\sigma_{\psi N}^2\} + 2\sigma_\varepsilon^4\left(\frac{NJ}{N^{*2}}c_{\mathbf{D}}^2c_{\mathbf{F}}^2\right)^{1/2}\right]$$

which will go to 0, almost surely, under assumptions 1–7. As a result, $\mathrm{var}(\tilde{\sigma}_{\theta\psi}) \to 0$ as N^* and J go to ∞.

A.6. Consistency of $\hat{\sigma}_\varepsilon^2$

The natural estimator of σ_ε^2 in equation (4) is

$$\hat{\sigma}_\varepsilon^2 = \frac{\mathbf{y}^\mathrm{T}\mathbf{M}\mathbf{y}}{\mathrm{tr}(\mathbf{M})} = \frac{\varepsilon^\mathrm{T}\mathbf{M}\varepsilon}{\mathrm{tr}(\mathbf{M})},$$

where $\mathbf{M} = \mathbf{I} - \mathbf{V}(\mathbf{V}^\mathrm{T}\mathbf{V})^{-1}\mathbf{V}^\mathrm{T}$, $\mathbf{V} \equiv (\mathbf{Z}, \mathbf{D}, \mathbf{F})$. It has expected value σ_ε^2 and

$$\mathrm{var}(\hat{\sigma}_\varepsilon^2) = \frac{1}{\mathrm{tr}(\mathbf{M})^2} 2\sigma_\varepsilon^4 \,\mathrm{tr}(\mathbf{M}^2) = \frac{2\sigma_\varepsilon^4}{\mathrm{tr}(\mathbf{M})}.$$

Then

$$\mathrm{tr}(\mathbf{M}) = N^* - K - N - J = N^*\left(1 - \frac{K}{N^*} - \frac{N}{N^*} - \frac{J}{N^*}\right)$$

where K is the number of columns in $\mathbf{Z}$ and $\mathrm{tr}(\mathbf{M}) \to \infty$, since $1 - K/N^* - N/N^* - J/N^*$ converges to a limit by using assumption 5.

Appendix B: Further analysis of limited mobility bias

An analysis of limited mobility bias is best done in the random-effects data generation process because movement between firms is remodelled every replication. In other words, ε, $\boldsymbol{\theta}$, $\boldsymbol{\psi}$, $\mathbf{D}$ and $\mathbf{F}$ vary from replication to replication. We define a baseline experiment with parameter values $J = 10000$, $T = 10$, $\mu_N = 10$, $p = 0.1$ and $\sigma_\varepsilon^2 = 1$. The number of workers per firm, N_j (which are drawn randomly from a log-normal distribution with mean μ_N) varies across replications. Each replication involves a completely new set of worker movements from firm to firm, and so the total number of workers who change firm each period varies by replication. The total number of observations, $T\Sigma_{j=1}^J N_j$, varies across replications for the same reason, even though the number of firms remains fixed. The population number of observations is $TJ\mu_N = 1000000$, in keeping with the size of many linked employer–employee data sets. The crucial parameter is the correlation between $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$, which again is chosen to be positive ($\rho_{\theta\psi} = 0.195$). This is the same for all experiments.

For each of the $R = 100$ data sets that make up the baseline experiment, we estimate equation (3) by using FEiLSDVj and compute the four biases for each replication. Because we do not have formulae to correct these estimators for bias, we compare 'true' and estimated outcomes replication by replication. For the covariance term this is

$$\tilde{\sigma}_{\theta\psi,r} - \sigma_{\theta\psi N,r}^2, \qquad r = 1, \ldots, 100.$$

There are similar expressions for $\tilde{\sigma}_\theta^2$, $\tilde{\sigma}_\psi^2$ and $\tilde{\rho}_{\theta\psi}$. In Fig. 1, there are 100 diamonds that represent the baseline experiment, associated with the cluster labelled '$p = 0.1$'.

The aim of the exercise is twofold. The first is to see whether our limited mobility bias formula that was derived in Section 3.5, namely equation (27), holds up in more complicated situations. The second is to quantify the extent of the bias as a function of the characteristics of a given data set by varying the simulations in single dimensions away from the baseline experiment. In other words, we vary one of the parameters J, T, μ_N, p and σ_ε^2, but keep the others fixed. In this appendix, we deal mainly with the first aim (by varying p) and we simply summarize our findings when varying J, T, μ_N and σ_ε^2.

The easiest way to illustrate the basic relationship between the bias in the covariance term and the number of movers M is to vary the probability of a match p. Simulations for three departures, for $p = 0.05$, $p = 0.15$ and $p = 0.20$, are plotted in Fig. 1, together with the baseline experiment $p = 0.10$. The main conclusion to emerge is that the relationship between the bias and M/J is very similar to that predicted by equation (27). The reason why a completely deterministic relationship is not being plotted is because the firm size distribution varies between replications, as does $\hat{\sigma}_\varepsilon^2$ and the number of firms J. Note that J is roughly constant, and so here it does not matter whether M or M/J is plotted on the x-axis. The same basic shape emerges when we plot the bias in the correlation, not the covariance, against M/J for the same four experiments (which are not reported).

When we simulate four further experiments for $\mu_N = 5, 10, 15, 20$, with $p = 0.10$, it turns out that the effect of changing μ_N is the same as changing p because it changes the average number of movers per firm. The same happens when we vary the number of time periods: $T = 5, 10, 15$. The longer the panel, each firm has on average more movers, and so the biases in the covariance become smaller. In contrast, varying the number of firms has no effect on the bias of the estimated correlation. This is because every new firm requires a new estimated parameter ψ_j, and so there is no improvement in variability across $\hat{\psi}_j$s. Recall that equation (27) does not apply in this situation. Because the number of movers doubles with twice as many firms, the bias appears to be homogeneous of degree 0 in M/J. Finally, as σ_ε^2 increases, the sampling variability of $\hat{\boldsymbol{\psi}}$ increases, which decreases the estimated correlation of $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$, and so the absolute value of the bias increases. Recall that all bias expressions are linear in σ_ε^2.

To conclude, altering any parameter which alters the number of movers in the data has an effect on the bias. Hence the bias in the correlation is decreasing in T, μ_N and p, and is unaffected by J. It is also increasing in σ_ε^2.

References

Abadir, K. and Magnus, J. (2005) *Matrix Algebra*. Cambridge: Cambridge University Press.

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

Abowd, J. and Kramarz, F. (1999) The analysis of labor markets using matched employer-employee data. In *Handbook of Labor Economics*, vol. 3B (eds O. Ashenfelter and D. Card), ch. 40, pp. 2567–2627. Amsterdam: Elsevier.

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

Alda, H., Bender, S. and Gartner, H. (2005) The linked employer-employee dataset of the IAB (LIAB). *Discussion Paper 06/2005*. Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg.

Andrews, M., Schank, T. and Upward, R. (2005) Practical estimation methods for linked employer-employee data. *Discussion Paper*. Lehrstuhle für Arbeitsmarkt- und Regionalpolitik, Friedrich-Alexander-Universität, Erlangen-Nürnberg.

Andrews, M., Schank, T. and Upward, R. (2006) Practical fixed effects estimation methods for the three-way error components model. *Stata J.*, **6**, 461–481.

Atakan, A. (2006) Assortative matching with explicit search costs. *Econometrica*, **3**, 667–680.

Barth, E. and 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. and Klose, C. (2000) The IAB employment subsample 1975-1995: opportunities for analysis provided by the anonymised sample. *Z. Wirts. Sozialwiss.*, **120**, 649–662.

Goux, D. and Maurin, E. (1999) Persistence of interindustry wage differentials: a reexamination using matched worker-firm panel data. *J. Lab. Econ.*, **17**, 492–533.

Gruetter, M. and Lalive, R. (2004) The importance of firms in wage determination. *Discussion Paper 1367*. Institute for the Study of Labor, Bonn.

Haltiwanger, J., Lane, J., Spletzer, J., Theeuwes, J. and Troske, K. (1999) *The Creation and Analysis of Employer-employee Matched Data*. Amsterdam: North-Holland.

Karr, A. (1992) *Probability*. Berlin: Springer.

Kölling, A. (2000) The IAB establishment panel. *Z. Wirts. Sozialwiss.*, **120**, 291–300.

Krueger, A. and Summers, L. (1988) Efficiency wages and the inter-industry wage structure. *Econometrica*, **56**, 259–293.

Magnus, J. and Neudecker, H. (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: Wiley.