The Stata Journal (2006)
6, Number 4, pp. 461–481

Practical fixed-effects estimation methods for the three-way error-components model

Martyn Andrews
School of Economic Studies
University of Manchester
Manchester, UK
martyn.andrews@man.ac.uk

Thorsten Schank
Department of Economics
Friedrich-Alexander-Universität Erlangen–Nürnberg
Nürnberg, Germany
thorsten.schank@wiso.uni-erlangen.de

Richard Upward
School of Economics
University of Nottingham
Nottingham, UK
richard.upward@nottingham.ac.uk

Abstract. Methods for fixed-effects estimation of the three-way error-components model are not yet standard. Where possible, we make the fixed-effects methods originally developed by Abowd, Kramarz, and Margolis (*Econometrica* 67: 251–333) for linked worker–firm data more accessible. We also show how these methods can be implemented in Stata. There is a caveat, however. If the researcher wants to recover estimates of the error components themselves, and the number of units at the higher level of aggregation is large, memory or matrix constraints may make using Stata to estimate the components themselves infeasible.

Keywords: st0112, grouping, linked employer–employee panel data, fixed effects

1 Introduction

Panel data with three or more dimensions of variation are increasingly available to researchers in various fields. One might have data on workers observed over time and on the firms in which they work. Or one might have children in schools or patients in hospitals. These data are also commonly described as being *multilevel* or *hierarchical*. In the examples given here, the lower-level units (workers, children, patients) are not merely nested in the higher level; they may also move between the higher-level units. Workers can change their employer, for example.

Researchers often use the linear *error-components model* for panel data. If the fixed (over time) error components are assumed to be uncorrelated with observed explanatory variables, then a random-effects estimator may be used. These models may be fitted in Stata in several ways, including the standard `xtreg` command for the two-way model and the new `xtmixed` command for models with a more complicated hierarchical structure. However, one may not wish to impose the assumption that the error components are uncorrelated with the observed explanatory variables, in which case one needs fixed-effects methods to estimate the parameters of interest.

In the linear model, one can model fixed-error components by using dummies—for example, individuals, firms, and time. In a two-way model with individual and time dummies, algebraic solutions are available for estimating all parameters of interest, including those associated with both sets of dummies (see Baltagi 2005, sec. 3.2). However, for data structures such as those considered here, there is no algebraic transformation that both sweeps away all the fixed error components in one go and allows them to be recovered later.

Although there is a growing literature in economics—much of it based on the work of Abowd, Kramarz, and Margolis (AKM)—analyzing data with three dimensions of variation (such as linked employer–employee panel data) is not yet routine. Because of various econometric obstacles, routine techniques and packages cannot be used. AKM's papers suggest that these issues are highly technical. Therefore, this paper's goal is to make these fixed-effects methods more accessible where possible and then to show how they can be implemented in Stata.

To illustrate these methods, we focus on a dataset of workers who are observed annually, together with the firms they work for. In section 2, we set out the generic model that best represents the econometrics of fixed-effects models using such data. In section 3, we describe the various methods that can be used to fit this generic model. In section 4, we present some illustrative results using an example dataset from the Institut für Arbeitsmarkt- und Berufsforschung (IAB) in Germany.

2 A generic model

Consider the following linear three-way error-components model:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{w}_{j(i,t)t}\boldsymbol{\gamma} + \mathbf{u}_i\boldsymbol{\eta} + \mathbf{q}_{j(i,t)}\boldsymbol{\rho} + \alpha_i + \phi_{j(i,t)} + \mu_t + \epsilon_{it} \tag{1}$$

Workers are indexed $i = 1, \ldots, N$. They are observed once per period $t = 1, \ldots, T_i$ in firm $j = 1, \ldots, J$. Workers can change firms over time, and the function $j(i,t)$ maps worker i to firm j at time t. y_{it} is the dependent variable, $\mathbf{x}_{it}$ and $\mathbf{u}_i$ are vectors of observable i-level covariates, and $\mathbf{w}_{j(i,t)t}$ and $\mathbf{q}_{j(i,t)}$ are vectors of observable j-level covariates. Both workers and firms are assumed to enter and exit the panel; i.e., we have an unbalanced panel with T_i observations per worker. There are $N^* = \sum_{i=1}^{N} T_i$ observations (worker periods) in total.

Both α_i and $\mathbf{u}_i$ are variables that are time invariant for workers, and similarly $\phi_{j(i,t)}$ and $\mathbf{q}_{j(i,t)}$ are fixed over time for firms. $\mathbf{x}_{it}$, on the other hand, varies across i and t, and $\mathbf{w}_{j(i,t)t}$ varies across j and t. Because the data are recorded at the (i,t) level, firm-level covariates also vary at that level, and referring to such variables as $\mathbf{w}_{jt}$ may therefore be less cumbersome.

Equation (1) therefore contains all four possible types of covariate that a researcher might have about workers and firms. There are K observed covariates in total.

The error components (or unobserved heterogeneities) comprise α_i for the worker and $\phi_{j(i,t)}$ for the firm. The third component, μ_t, represents the unobserved time effect.

The error components may be correlated with each other and with any of the observable covariates.

Throughout, we assume strict exogeneity; namely,

$$E[\epsilon_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},\mathbf{w}_{j(i,t)1},\ldots,\mathbf{w}_{j(i,t)T},\mathbf{u}_i,\mathbf{q}_{j(i,t)},\alpha_i,\phi_{j(i,t)},\mu_t]=0$$

This equation implies, among other things, that workers' mobility decisions are independent of ϵ_{it}. Although possibly implausible in the current context, this assumption is standard in the literature.

Assuming that the heterogeneity terms α_i and $\phi_{j(i,t)}$ are correlated with the observables is customary. Random-effects methods are therefore inconsistent, and so one needs fixed-effects methods to estimate the parameters of interest. Thus $[\boldsymbol{\eta},\boldsymbol{\rho}]$, the parameter vector associated with the time-invariant variables, is not identified. Rather than dropping $[\mathbf{u}_i,\mathbf{q}_{j(i,t)}]$, it is customary to define

$$\theta_i \equiv \alpha_i + \mathbf{u}_i\boldsymbol{\eta} \tag{2}$$

and

$$\psi_j \equiv \phi_j + \mathbf{q}_j\boldsymbol{\rho} \tag{3}$$

giving

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{w}_{j(i,t)t}\boldsymbol{\gamma} + \theta_i + \psi_{j(i,t)} + \epsilon_{it} \tag{4}$$

In the next section, we describe how to estimate the parameters of (4) by using various fixed-effects methods. We assume throughout that the unobserved time component μ_t is to be treated as fixed and estimated directly by using time dummies.[1] We have therefore dropped μ_t from (4); we have subsumed these time dummies into one of the vectors of observable covariates. Thus we are essentially analyzing a two-way model henceforth.

3 Econometric methods

3.1 Spell fixed effects

If one is not interested in the estimates of θ_i and $\psi_{j(i,t)}$ themselves, or in estimating the parameters on the time-invariant variables $\mathbf{u}_i$ and $\mathbf{q}_{j(i,t)}$, one can easily obtain consistent estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ from (4) by taking differences or by time-demeaning within each unique worker–firm combination (or spell). For each spell of a worker within a firm, neither θ_i nor $\psi_{j(i,t)}$ varies. When we define $\lambda_s \equiv \theta_i + \psi_{j(i,t)}$ as spell-level heterogeneity, which is swept out by subtracting averages at the spell level, both θ_i and $\psi_{j(i,t)}$ have disappeared:

$$y_{it} - \overline{y}_s = (\mathbf{x}_{it} - \overline{\mathbf{x}}_s)\boldsymbol{\beta} + (\mathbf{w}_{j(i,t)t} - \overline{\mathbf{w}}_s)\boldsymbol{\gamma} + (\epsilon_{it} - \overline{\epsilon}_s) \tag{5}$$

1. This assumption will always be practical so long as the time dimension of the panel is relatively short, which it usually is with these kinds of data.

The effects of $\mathbf{u}_i$ and $\mathbf{q}_{j(i,t)}$ are not identified, because $\mathbf{u}_i - \overline{\mathbf{u}}_s = \mathbf{0}$ and $\mathbf{q}_{j(i,t)} - \overline{\mathbf{q}}_s = \mathbf{0}$. Any variable x_{it} or w_{jt} that is constant *within a spell* will also not be identified. One observation per spell is used up in identifying each spell fixed effect.

This is essentially the first method that AKM discuss (Abowd, Kramarz, and Margolis 1999, sec. 3.3), except that they use differences rather than mean deviations. We call this method spell fixed effects, or FE(s). Because one cannot separate the worker and firm heterogeneities, AKM do not pursue this method further. Given estimates $\widehat{\lambda}_s$, one cannot recover $\widehat{\theta}_i$ and $\widehat{\psi}_j$.

Unless researchers are analyzing the heterogeneity after estimation, this spell fixed-effects method is a practical and simple solution that presents no computational difficulty. Stata's `xtreg, fe` command readily implements this method.

3.2 Least-squares dummy variables

The spell fixed-effects method outlined above is not useful if one wants to recover estimates of θ_i and $\psi_{j(i,t)}$, specifically if one wants to analyze these terms themselves or if one wants to recover estimates of $\boldsymbol{\rho}$ and $\boldsymbol{\eta}$ by using (2) and (3), respectively. An alternative is to use a least-squares dummy variable (LSDV) estimator when estimating (4).

However, using dummy variables to directly estimate (4) when the dataset is large will not usually be feasible, since this is a model with approximately $K + N + J$ parameters. In the two-way model, this problem is circumvented by using the within transformation that sweeps out the i-level heterogeneity. But because of the lack of patterning between workers and firms,[2] there is no algebraic transformation of the observables that both sweeps away all heterogeneity terms in one go and allows them to be recovered later. To circumvent this problem, AKM note that explicitly including dummy variables for the firm heterogeneity, but sweeping out the worker heterogeneity algebraically, gives the same solution as the LSDV estimator.[3]

More precisely, the researcher must generate a dummy variable for each firm,

$$F_{it}^{j} = 1\{j(i,t) = j\}, \quad j = 1, \ldots, j$$

where $1\{\ \}$ is the dummy variable indicator function and the function $j(i,t) = j$ maps worker i at time t to firm j. Now substitute

$$\psi_{j(i,t)} = \sum_{j=1}^{J} \psi_j F_{it}^{j}$$

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

3. In linear models, there is no distinction between removing the heterogeneity algebraically and adding two full sets of dummy variables, for workers and firms, and so the terminology "LSDV" applies to both.

into (4). The θ_i are removed by time-demeaning (or differencing) over i:

$$y_{it} - \overline{y}_i = (\mathbf{x}_{it} - \overline{\mathbf{x}}_i)\boldsymbol{\beta} + (\mathbf{w}_{j(i,t)t} - \overline{\mathbf{w}}_i)\boldsymbol{\gamma} + \sum_{j=1}^{J} \psi_j (F_{it}^j - \overline{F}_i^j) + \epsilon_{it}$$

Therefore, J demeaned (or differenced) firm dummies actually need creating.[4] To distinguish this estimator from LSDV above, we label this estimator FEiLSDVj. They are identical estimators but differ in how they are computed.

Because of the strict exogeneity assumption, these estimators of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and ψ_j are consistent and unbiased, conditional on the observed covariates. One is assuming that there are many individuals and relatively fewer firms. However, the LSDV estimators of θ_i are inconsistent, although unbiased. Adding one more individual does not mean that there is more information, because one also has an extra parameter to estimate.[5] In short, the firm effects, $\psi_{j(i,t)}$, can be estimated precisely, provided that firms have enough workers who join or leave. But worker effects, θ_i, cannot be estimated precisely with relatively few periods per worker.

Firm dummies are no different from any multicategory dummy, so long as workers can move from one category to another over time (e.g., region dummies but not ethnicity dummies). $F_{it}^j - \overline{F}_i^j$ will be zero for all J dummies for any worker i who does not change firm. Furthermore, if we have a sample of firms (rather than the population) $F_{it}^j - \overline{F}_i^j$ will be nonzero only for workers who change from one firm within the sample to another firm in the sample. Often only a tiny proportion of workers have any nonzero terms. Identification of $\psi_{j(i,t)}$ is driven by the total number of such movers in each firm j. Some small firms may have no movers, in which case $\psi_{j(i,t)}$ is not identified. Other small firms may have only a few movers, in which case estimates of $\psi_{j(i,t)}$ will be imprecise. Estimating $\psi_{j(i,t)}$ for small firms may not be sensible, and instead one should group small firms together (as AKM and others do.)

This estimator has two potential computational problems. The first is the number of firms, J, because the software needs to invert a matrix of dimension $(K + J) \times (K + J)$. For many applications, the number of firms is small enough that FEiLSDVj is computationally feasible. However, some datasets have tens of thousands of firms, or even hundreds of thousands. The second problem is that one must create and store J mean deviations for N^* observations, meaning that the data matrix is $N^* \times (K + J)$. This matrix may be prohibitively large for software packages that store all data in memory, such as Stata.

Some improvement in the storage efficiency of the J mean-deviated firm dummies can be achieved in Stata by using the lowest common multiple of all values of T_i if the panel is short enough. For example, if the data span a maximum of 5 years then T_i can

4. Differencing is ignored hereafter. Implementing the covariance transformation is easier for various reasons. Normally, the decision whether to fit the model in first differences or use the covariance transform depends on which method gives more efficient estimates. Both estimators are consistent. See Wooldridge (2002, sec. 10.6.3).

5. See Wooldridge (2002, chap. 10) for assumptions and properties of panel-data models.

be any value from 1 to 5. Multiplying $F_{it}^j - \overline{F}_i^j$ by the lowest common multiple (here 60) yields a set of integers that can be stored in Stata as single bytes rather than 4- or 8-byte fractions.[6] To implement this method, the researcher would need to create the mean deviations manually and use ordinary least squares (OLS) on the transformed data, rather than relying on `xtreg`.

The memory requirements of the data matrix for the FEiLSDVj estimator are then approximately $(N^*J)+4\{N^*(K+1)\}$ bytes. We require N^*J bytes for the mean-deviated firm dummies and $4\{N^*(K+1)\}$ bytes for the remaining K explanatory variables and the dependent variable, assuming that each is stored as 4 bytes.

3.3 A classical minimum distance method

The FEiLSDVj method requires inverting a potentially very large $(K+J)\times(K+J)$ cross-product matrix, as well as enough memory to store J mean-deviated firm dummies across N^* observations. To circumvent the second issue, we propose the following method, since only movers between firms identify firm effects.

We separate the model into observations for movers and nonmovers, subscripted by "1" and "2", respectively. There are N_1^* mover observations and N_2^* nonmover observations. We then write (4) in matrix notation, where each model is fitted separately:[7]

$$\begin{aligned} \widetilde{\mathbf{y}}_1 &= \widetilde{\mathbf{X}}_1\boldsymbol{\beta}_1 + \widetilde{\mathbf{F}}_1\boldsymbol{\psi}_1 + \boldsymbol{\epsilon}_1 \\ \widetilde{\mathbf{y}}_2 &= \widetilde{\mathbf{X}}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}_2 \end{aligned} \tag{6}$$

$\widetilde{\mathbf{y}}_1$, $\widetilde{\mathbf{y}}_2$, $\widetilde{\mathbf{X}}_1$, $\widetilde{\mathbf{X}}_2$, and $\widetilde{\mathbf{F}}_1$ have all been mean-deviated and defined; namely, $\widetilde{\mathbf{y}}_1 = \mathbf{M_D}\mathbf{y}_1$, etc., where $\mathbf{M_D} \equiv \mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$. Denote the variances of the two error terms as $\sigma_{\epsilon 1}^2$ and $\sigma_{\epsilon 2}^2$. We now drop all columns of $\widetilde{\mathbf{F}}_1$ that are the zero vector, that is, the J_2 firms that have no turnover. By definition, $\widetilde{\mathbf{F}}_2 \equiv \mathbf{0}$.

Because there are often few movers, eliminating $\widetilde{\mathbf{F}}_2 \equiv \mathbf{0}$ from the model means that we sidestep the memory constraints by fitting the model for movers and nonmovers separately. The classical minimum distance (CMD) estimator forms a restricted estimator for $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ from $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\psi}_1$.[8]

In general, denote $\boldsymbol{\pi}$ as the $S \times 1$ unrestricted parameter vector and denote $\boldsymbol{\delta}$ as the $P \times 1$ restricted parameter vector. The constraint is $\boldsymbol{\pi} = \mathbf{h}(\boldsymbol{\delta})$. In CMD estimation, one estimates $\boldsymbol{\pi}$ and then finds a $\boldsymbol{\delta}$ such that the distance between $\widehat{\boldsymbol{\pi}}$ and $\mathbf{h}(\boldsymbol{\delta})$ is minimized. An efficient CMD estimator uses any consistent estimator of asymptotic covariance matrix $\mathbf{V}$ to act as weighting matrix for the distance between $\widehat{\boldsymbol{\pi}}$ and $\mathbf{h}(\boldsymbol{\delta})$, denoted $\widehat{\mathbf{V}}$. In other words, efficient CMD solves

$$\min_{\boldsymbol{\delta}} \{\widehat{\boldsymbol{\pi}} - \mathbf{h}(\boldsymbol{\delta})\}' \, \widehat{\mathbf{V}}^{-1} \{\widehat{\boldsymbol{\pi}} - \mathbf{h}(\boldsymbol{\delta})\}$$

6. Storing the mean-deviated firm dummies as integers also appears to improve the accuracy of the matrix inversion procedure.

7. We dispense with the distinction between $\mathbf{x}_{it}$ variables and $\mathbf{w}_{j(i,t)t}$ variables.

8. See Wooldridge (2002, sec. 14.6) for more details.

whose solution is

$$\widehat{\boldsymbol{\delta}} = (\mathbf{H}'\widehat{\mathbf{V}}^{-1}\mathbf{H})^{-1}\mathbf{H}'\widehat{\mathbf{V}}^{-1}\widehat{\boldsymbol{\pi}}$$

when the mapping from $\boldsymbol{\delta}$ to $\boldsymbol{\pi}$ is linear: $\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\delta}$. Also the appropriate estimator of $\widehat{\mathrm{Avar}}(\widehat{\boldsymbol{\delta}})$ with which to conduct inference is

$$\widehat{\mathrm{Avar}}(\widehat{\boldsymbol{\delta}}) = \left\{\mathbf{H}'\widehat{\mathrm{Avar}}(\widehat{\boldsymbol{\pi}})^{-1}\mathbf{H}\right\}^{-1} = (\mathbf{H}'\widehat{\mathbf{V}}^{-1}\mathbf{H})^{-1}$$

A test of the validity of the restrictions is given by Wooldridge (2002, eqn. 14.76):

$$\left\{\widehat{\boldsymbol{\pi}} - \mathbf{h}(\widehat{\boldsymbol{\delta}})\right\}' \widehat{\mathbf{V}}^{-1} \left\{\widehat{\boldsymbol{\pi}} - \mathbf{h}(\widehat{\boldsymbol{\delta}})\right\} \sim \chi^2(S-P)$$

For the model at hand, the constraint $\boldsymbol{\pi} = \mathbf{H}\boldsymbol{\delta}$ is written as

$$\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\psi}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} = \begin{bmatrix} \mathbf{I}_K & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_J \\ \mathbf{I}_K & \mathbf{0} \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\psi} \end{pmatrix}$$

where $\boldsymbol{\pi}$ is $(2K+J)\times 1$, $\boldsymbol{\delta}$ is $(K+J)\times 1$, and $\mathbf{H}$ is $(2K+J)\times(K+J)$. The appropriate asymptotic covariance matrix is

$$\widehat{\mathbf{V}} = \begin{bmatrix} \widehat{\mathbf{V}}_1 & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{V}}_2 \end{bmatrix} = \begin{Bmatrix} \widehat{\sigma}_1^{-2} \begin{pmatrix} \widetilde{\mathbf{X}}_1'\widetilde{\mathbf{X}}_1 & \widetilde{\mathbf{X}}_1'\widetilde{\mathbf{F}}_1 \\ \widetilde{\mathbf{F}}_1'\widetilde{\mathbf{X}}_1 & \widetilde{\mathbf{F}}_1'\widetilde{\mathbf{F}}_1 \end{pmatrix} & \mathbf{0} \\ \mathbf{0} & \widehat{\sigma}_2^{-2}(\widetilde{\mathbf{X}}_2'\widetilde{\mathbf{X}}_2) \end{Bmatrix}$$

From the general expressions immediately above, it follows that the restricted estimator $\widehat{\boldsymbol{\delta}} = (\mathbf{H}'\widehat{\mathbf{V}}^{-1}\mathbf{H})^{-1}\mathbf{H}'\widehat{\mathbf{V}}^{-1}\widehat{\boldsymbol{\pi}}$ is given by

$$\widehat{\boldsymbol{\delta}} = \begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\psi}} \end{pmatrix} = \left\{\widehat{\mathbf{V}}_1^{-1} + \begin{pmatrix} \widehat{\mathbf{V}}_2^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right\}^{-1} \left\{\widehat{\mathbf{V}}_1^{-1}\begin{pmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\psi}}_1 \end{pmatrix} + \widehat{\mathbf{V}}_2^{-1}\begin{pmatrix} \widehat{\boldsymbol{\beta}}_2 \\ \mathbf{0} \end{pmatrix}\right\} \tag{7}$$

and that

$$\widehat{\mathrm{Avar}}(\widehat{\boldsymbol{\delta}}) = (\mathbf{H}'\widehat{\mathbf{V}}^{-1}\mathbf{H})^{-1} = \left\{\widehat{\mathbf{V}}_1^{-1} + \begin{pmatrix} \widehat{\mathbf{V}}_2^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right\}^{-1} \tag{8}$$

a $(K+J)\times(K+J)$ matrix. These expressions use standard (unrobust) covariance matrices. A robust version of this covariance matrix replaces $\widehat{\mathbf{V}}_1$ and $\widehat{\mathbf{V}}_2$ in (8) by robust equivalents. Following this approach in (7) is wrong, however, because if the true constraint could be imposed on the model, one would not end up with the LSDV estimator.

A standard criticism is that movers and nonmovers are different groups of individuals and so one should model them separately. Before imposing H_0: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, one should test these restrictions, although doing so is rarely done in practice. Under H_0,

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}_1 - \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\psi}}_1 - \widehat{\boldsymbol{\psi}} \end{pmatrix}' \widehat{\mathbf{V}}_1^{-1} \begin{pmatrix} \widehat{\boldsymbol{\beta}}_1 - \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\psi}}_1 - \widehat{\boldsymbol{\psi}} \end{pmatrix} + (\widehat{\boldsymbol{\beta}}_2 - \widehat{\boldsymbol{\beta}})'\widehat{\mathbf{V}}_2^{-1}(\widehat{\boldsymbol{\beta}}_2 - \widehat{\boldsymbol{\beta}}) \sim \chi^2(K) \tag{9}$$

The *only* price paid with this approach is that one cannot constrain $\sigma^2_{\epsilon 1} = \sigma^2_{\epsilon 2}$. The only difference between this and the LSDV estimator is because $\widehat{\mathbf{V}}_1^{-1}$ and $\widehat{\mathbf{V}}_2^{-1}$ come from separate regressions.

3.4 Postestimation analysis of the error components

Once estimates of $(\boldsymbol{\beta}, \boldsymbol{\psi})$ have been made (using either the FEiLSDVj or CMD method), the researcher can recover estimates of the error components themselves. First, compute

$$\widehat{\psi}_{j(i,t)} = \sum_{j=1}^{J} \widehat{\psi}_j F^j_{it} \tag{10}$$

and then

$$\widehat{\theta}_i = \overline{y}_i - \overline{\widehat{\psi}}_i - \overline{\mathbf{x}}_i \widehat{\boldsymbol{\beta}} - \overline{\mathbf{w}}_i \widehat{\boldsymbol{\gamma}} \tag{11}$$

where $\overline{\widehat{\psi}}_i$ averages $\widehat{\psi}_{j(i,t)}$ over t for each i.

It is not possible to identify one firm effect per *group*, where a group is defined by the movement of workers between firms. A group contains all the workers who have ever worked for any of the firms in that group, as well as all the firms at which any of the workers were employed. A second (unconnected) group is defined only if no firm in the first group has ever employed any workers in the second and no firm in the second group has ever employed any workers in the first. It is not possible to identify one firm per group because within each group the mean-deviated firm dummies sum to zero. Some normalization is therefore required between groups. In section 4, we show how to identify groups and how to perform this normalization.

Identifying the effects of time-invariant variables

AKM suggest that one can recover estimates of $\widehat{\alpha}_i$ and $\widehat{\phi}_j$ by estimating (2) and (3) as follows. Thus one can analyze distributions of $\widehat{\alpha}_i$ and $\widehat{\phi}_j$, specifically to see whether they are correlated. First, run the auxiliary regressions,

$$\widehat{\theta}_i = \text{constant} + \mathbf{u}_i \boldsymbol{\eta} + \text{error} \tag{12}$$

$$\widehat{\psi}_j = \text{constant} + \mathbf{q}_j \boldsymbol{\rho} + \text{error} \tag{13}$$

which give consistent estimates of $\boldsymbol{\eta}$, $\boldsymbol{\rho}$ (Abowd, Kramarz, and Margolis 1999, sec. 3.4.4). Because α_i is dropped from (2), the identifying assumption is that $\text{Cov}(\mathbf{u}_i, \alpha_i) = 0$ or else there is omitted-variable bias. Similarly, $\text{Cov}(\mathbf{q}_{j(i,t)}, \phi_{j(i,t)}) = 0$ is assumed in (3). One needs only N observations to estimate (2) and J observations to estimate (3). AKM estimate these equations by generalized least squares, because of the aggregation to the firm level. Because there are other potential causes of heteroskedasticity, one could use OLS and adjust the covariance matrix for clustering at the firm level. Second, the researcher computes

$$\widehat{\alpha}_i = \widehat{\theta}_i - \mathbf{u}_i \widehat{\boldsymbol{\eta}}$$
$$\widehat{\phi}_j = \widehat{\psi}_j - \mathbf{q}_j \widehat{\boldsymbol{\rho}}$$

θ and ψ can be defined at three levels of aggregation:

i, t	θ_i replicated T_i times	$\psi_{j(i,t)}$
i	θ_i	$\overline{\psi}_i = \sum_{t=1}^{T_i} \psi_{j(i,t)}/T_i$
j	$\overline{\theta}_j = \sum_{(it)\in j} \theta_i/N_j^*$	$\psi_{j(i,t)}$

(N_j^* is the total number of worker-years observed in firm j.) AKM show that statistics based on aggregating $\widehat{\theta}_i$ and $\widehat{\alpha}_i$ to the level of the firm are consistent as T_i goes to infinity (see also Chamberlain 1984).

4 An illustrative example

To illustrate these methods, we use data from a linked worker–firm dataset made available by the IAB. The firm data comprise a panel of 4,376 establishments (or "plants") from the former West Germany observed over 1993–1997. The worker data comprise a panel of 1,930,260 workers who are employed in these plants. A common establishment identifier is available in both datasets, allowing them to be linked.[9] After we eliminate observations with missing or incomplete information, the resulting linked dataset has 5,145,098 worker-year observations (the i, t level). For each row in the data, the identity, j, of the plant is recorded.

The first row of table 1 reports the total sample size in terms of the total number of rows in the data (N^*), the number of workers (N), the number of plants (J), and the number of unique worker–firm combinations, or "spells" (S). The total number of spells is only slightly greater than the total number of workers. This is a consequence of having a sample of plants: a worker will be observed with more than one spell only if he moves from one plant in the sample to another, which is actually unlikely.

9. Kölling (2000) provides more information on the IAB establishment panel; Bender, Haas, and Klose (2000) has details on the worker data; and Alda, Bender, and Gartner (2001) has details on the linked data.

Table 1: Sample sizes

Sample	N^*	N	J	S
Whole sample	5,145,098	1,930,260	4,376	1,953,774
Workers who move to other IAB plants	72,253	23,393	1,821	46,907
Workers who don't move to other IAB plants	5,072,845	1,906,867	4,376	1,906,867
Workers in plants with movement to other IAB plants	4,883,331	1,816,368	1,821	1,839,882

Identifying unobserved plant effects is driven only by those workers who change plants. Thus an important subsample comprises those workers who have two or more spells ($S_i > 2$) in IAB plants ("IAB movers"). There are only 23,393 of these movers, and they work in 1,821 plants. However, these 1,821 plants employ more than 1.8 million workers because they tend to be larger than plants who employ no IAB movers.

To illustrate our methods, we estimate log-wage equations using a small set of covariates of each type ($\mathbf{x}$, $\mathbf{w}$, $\mathbf{q}$, and $\mathbf{u}$). The dependent variable y_{it} is the log daily wage in Pfennigen. Table 2 gives the sample means for the relevant variables.

Table 2: Sample means

Description	Variable type	Variable name	Mean	SD
Log daily wage in Pfennigen	y_{it}	`lw`	9.763	0.278
Female	u_i	`female`	0.213	0.409
Married	x_{it}	`married`	0.624	0.484
Age	x_{it}	`age`	39.643	10.539
Age^2/100	x_{it}	`age_2`	16.827	8.628
Single-plant enterprise	q_j	`single`	0.269	0.443
Log plant employment	w_{jt}	`lN`	7.702	1.454
(Log plant employment)^2	w_{jt}	`lN_2`	61.429	22.283

A simple OLS estimate of (1) provides a useful benchmark. This estimate treats α_i and $\phi_{j(i,t)}$ as part of the error term, whereas μ_t is estimated using a dummy for each year. Here we correct the standard errors for possible correlation within i, since the errors in (4) are probably not independent for the same individual. However, we might also consider clustering at the j level, since errors may also be correlated across individuals within the same firm.

```
. regress lw female married age age_2 single lN lN_2 year2-year5, cluster(i)
Linear regression                                      Number of obs = 5145098
                                                       F(11,1930259) =71255.43
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2441
Number of clusters (i) = 1930260                       Root MSE      =   .2417

------------------------------------------------------------------------------
             |               Robust
          lw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.1773087    .000522  -339.70   0.000    -.1783317   -.1762856
     married |   .0126403    .000415    30.46   0.000      .011827    .0134536
         age |   .0351288   .0001435   244.72   0.000     .0348474    .0354101
       age_2 |  -.0356353   .0001781  -200.11   0.000    -.0359844   -.0352863
      single |  -.0317087   .0004474   -70.87   0.000    -.0325856   -.0308317
          lN |   .0793973   .0009911    80.11   0.000     .0774549    .0813397
        lN_2 |  -.0028624   .0000618   -46.28   0.000    -.0029836   -.0027412
       year2 |   .0295777   .0001956   151.21   0.000     .0291944    .0299611
       year3 |   .0723557   .0002432   297.52   0.000     .0718791    .0728324
       year4 |   .0983264   .0002655   370.33   0.000      .097806    .0988468
       year5 |   .1078701   .0002919   369.55   0.000      .107298    .1084422
       _cons |   8.514487   .0046803  1819.21   0.000     8.505314    8.523661
------------------------------------------------------------------------------
```

The strict exogeneity assumption for the OLS estimator is different from that discussed above for the LSDV estimator. The implied error term is $\theta_i + \psi_{j(i,t)} + \epsilon_{it}$—see (4)—and each component is assumed to be contemporaneously uncorrelated with the observed covariates. We now want to investigate whether any of these parameter estimates are likely to be biased because of potential correlation between the unobserved error components and the variables of interest.

Fitting two-way fixed-effects models by using the `xtreg` command is simple. The within-i transformation eliminates the unobserved worker-level component of the error term, θ_i, and assumes that $\phi_{j(i,t)}$ is uncorrelated with the covariates. The standard errors here are robust to cross-sectional heteroskedasticity and within-panel i correlation (see StataCorp [2005, 293] and Wooldridge [2002, eqn. 10.59]). For reference, denote this estimator as FE(i):

(*Continued on next page*)

```
. xtreg lw female married age age_2 single lN lN_2 year2-year5, fe i(i)
> cluster(i)
Fixed-effects (within) regression               Number of obs      =   5145098
Group variable (i): i                           Number of groups   =   1930260

R-sq:  within  = 0.3029                         Obs per group: min =         1
       between = 0.1088                                        avg =       2.7
       overall = 0.0964                                        max =         5

                                                F(9,5145089)       =  97217.88
corr(u_i, Xb)  = -0.6637                        Prob > F           =    0.0000

                               (Std. Err. adjusted for 1930260 clusters in i)
```

lw	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
female	(dropped)					
married	.0061755	.0002999	20.59	0.000	.0055877	.0067633
age	.0582696	.000171	340.77	0.000	.0579345	.0586048
age_2	-.0353585	.0001952	-181.15	0.000	-.0357411	-.034976
single	-.0049728	.0020442	-2.43	0.015	-.0089793	-.0009662
lN	-.0016697	.0020872	-0.80	0.424	-.0057605	.0024211
lN_2	.0010928	.0001366	8.00	0.000	.0008251	.0013604
(output omitted)						
sigma_u	.35916844					
sigma_e	.06832495					
rho	.96507601	(fraction of variance due to u_i)				

We cannot estimate a coefficient on `female` because it does not vary within i. There are some significant changes in some of the parameter estimates. For example, the effect of age has gone up from 0.035 to 0.058, whereas the effect of age^2 remains the same. Thus the estimated quadratic wage–age profile is steeper, with a much older turning point. This outcome implies a negative correlation between θ_i and age—older workers have lower (unobserved) earning power. This conclusion is reflected in the large correlation (-0.6637) between θ_i and the estimated effect of all the covariates.

The within-j transformation eliminates the unobserved plant-level component, $\psi_{j(i,t)}$, and assumes that α_i is uncorrelated with the covariates. Here `xtreg` does not permit us to cluster on i because the panels are not nested within clusters (individuals may move between firms). The standard errors here are therefore robust to cross-sectional heteroskedasticity and within-panel j correlation. For reference, denote this estimator as FE(j):

```
. xtreg lw female married age age_2 single lN lN_2 year2-year5, fe i(j)
> cluster(j)
Fixed-effects (within) regression               Number of obs      =   5145098
Group variable (i): j                           Number of groups   =      4376
R-sq:  within  = 0.2151                         Obs per group: min =         1
       between = 0.0000                                        avg =    1175.8
       overall = 0.1810                                        max =    102106
                                                F(10,5145088)      =   1068.04
corr(u_i, Xb)  = 0.0100                         Prob > F           =    0.0000
                                   (Std. Err. adjusted for 4376 clusters in j)
------------------------------------------------------------------------------
             |               Robust
          lw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.1779654   .0042996   -41.39   0.000    -.1863947    -.169536
     married |   .0247614   .0012382    20.00   0.000     .0223339    .0271889
         age |   .0327271   .0007152    45.76   0.000      .031325    .0341292
       age_2 |  -.0336805   .0008039   -41.90   0.000    -.0352565   -.0321046
      single |  (dropped)
          lN |  -.0694038   .0491958    -1.41   0.158    -.1658524    .0270449
        lN_2 |   .0042823   .0038799     1.10   0.270    -.0033243    .0118889
  (output omitted)
-------------+----------------------------------------------------------------
     sigma_u |  .35985478
     sigma_e |  .20663177
         rho |  .75204046   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

The coefficient estimates on `female` and `age` are both similar to those from the OLS regression, suggesting that the correlation between individual-level variables and $\phi_{j(i,t)}$ is not important. However, the wage effect of plant size is now considerably different from those estimated by either the OLS or FE(i) model, and there is a doubling of the marriage premium.

We now want to eliminate *both* the unobserved worker- and plant-level error components. The simplest way to do this is to estimate (5) by using FE(s):

(*Continued on next page*)

```
. egen s=group(i j)
. xtreg lw female married age age_2 single lN lN_2 year2-year5, fe i(s) cluster(i)
Fixed-effects (within) regression               Number of obs      =   5145098
Group variable (i): s                           Number of groups   =   1953774

R-sq:  within  = 0.3029                         Obs per group: min =         1
       between = 0.1119                                        avg =       2.6
       overall = 0.0994                                        max =         5

                                                F(8,5145090)       = 109076.69
corr(u_i, Xb)  = -0.6669                        Prob > F           =    0.0000

                          (Std. Err. adjusted for 1930260 clusters in i)

------------------------------------------------------------------------------
             |               Robust
          lw |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  (dropped)
     married |   .0057338   .0002979    19.25   0.000     .0051499    .0063178
         age |   .0582328   .0001709   340.75   0.000     .0578979    .0585678
       age_2 |  -.0350417   .0001944  -180.26   0.000    -.0354227   -.0346607
      single |  (dropped)
          lN |  -.0102442   .0021745    -4.71   0.000     -.014506   -.0059823
        lN_2 |   .0021531   .0001401    15.37   0.000     .0018785    .0024276
  (output omitted)
-------------+----------------------------------------------------------------
     sigma_u |  .36006301
     sigma_e |    .067745
         rho |  .96581076   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

By defining a spell, `s`, in this way, we treat as one spell all unique combinations of i and j. Therefore, a worker who has two periods of employment with employer A, separated by a period with employer B, is treated as having just two spells in total.

If the correct model is given by (4), if the strict exogeneity assumption holds, and if the error components are correlated with the observed data, then these estimates should be preferred to either of the standard fixed-effects or OLS estimates. Here estimates from FE(s) are generally close to those from FE(i), implying that ϕ_j is uncorrelated with the observed covariates. The correlation of 0.0100 reported in the FE(j) regression confirms this implication. However, we are now unable to estimate either of the coefficients on `female` or `single`.

We now want to estimate (4), but we also want to then recover estimates of α_i and $\phi_{j(i,t)}$. If we had enough memory, we could use the LSDV methods outlined in section 3.2. In our example, we have $N^* = 5{,}145{,}098$, $J = 1{,}821$, and $K = 11$, meaning that we require about 10 GB of memory to proceed. At press time, this amount of memory is not available to us (or to many researchers) and so we must use the CMD method described in section 3.3.

The dummy variable `mover` identifies workers who change plant during the sample period. The variable `plantin` counts the number of worker-years in each plant of workers who move.

```
. sort i j
. by i: gen byte mover = j[1]!=j[_N]
. egen plantin = total(mover), by(j)
. save cmd, replace
```

Only workers with `mover=1` contribute to estimates of $\psi_{j(i,t)}$, and one cannot estimate $\psi_{j(i,t)}$ for plants with `plantin=0`. To estimate (6), we use `xtreg` only for movers and include a full set of firm dummies:

```
. keep if mover==1
. tab j, gen(F_)
. local J1 = r(r)
. xtreg lw married age age_2 single lN lN_2 year2-year4 F_*, fe i(i)
```

We then save the coefficient estimates $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\psi}}$ and the variance–covariance matrix $\widehat{\mathbf{V}}_1$, removing the constant from both:

```
. matrix B1 = e(b)'
. matrix B1 = B1["x".."F_`J1'",1]
. matrix V1 = e(V)
. matrix V1 = V1["married".."F_`J1'","married".."F_`J1'"]
```

To calculate a robust version of $\widehat{\text{Avar}}(\widehat{\boldsymbol{\delta}})$ (8) we also need a robust equivalent of `V1`. For example,

```
xtreg lw married age age_2 single lN lN_2 year2-year4 F_*, fe i(i) cluster(i)
matrix V1r = e(V)
matrix V1r = V1r["married".."F_`J1'","married".."F_`J1'"]
```

The process is then repeated for the nonmovers. We do not issue a `clear` command on its own. Instead `use, clear` loads the data without destroying any of the relevant matrices in memory.

```
. use cmd if mover==0, clear
. xtreg y married age age_2 lN lN_2 year2-year4, fe i(i)
. matrix BETA2 = e(b)'
. matrix BETA2 = BETA2["married".."year4",1]
. local K = rowsof(BETA2)
. matrix V2 = e(V)
. matrix V2 = V2["married".."year4","married".."year4"]
```

Again if we want to compute a robust version of $\widehat{\text{Avar}}(\widehat{\boldsymbol{\delta}})$, we also need the following:

```
. xtreg y married age age_2 lN lN_2 year2-year4, fe i(i) cluster(i)
. matrix V2r = e(V)
. matrix V2r = V2r["married".."year4","married".."year4"]
```

Now we can compute the restricted estimator $\widehat{\boldsymbol{\delta}}$, given by (7). To do this computation, we need to construct the vector

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}_2 \\ \mathbf{0} \end{pmatrix}$$

and the matrix

$$\begin{pmatrix} \widehat{\mathbf{V}}_2^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$$

which is achieved by adding blocks of zeros to $\widehat{\boldsymbol{\beta}}_2$ and $\widehat{\mathbf{V}}_2$:

```
. matrix B2 = BETA2\J(`J1',1,0)
. matrix V2inv = J(`J1'+`K',`J1'+`K',0)
. matrix V2inv[1,1] = invsym(V2)
```

Equation (7) is then computed

```
. matrix DELTA = invsym(invsym(V1)+V2inv)*((invsym(V1)*B1)+(V2inv*B2))
```

and the robust variance–covariance matrix of $\widehat{\boldsymbol{\delta}}$ is given by (8):

```
. matrix VARDELTAr = invsym(invsym(V1r)+V2inv)
```

We can label the resulting matrices by using the variable names of $(\widehat{\boldsymbol{\beta}}_1, \widehat{\mathbf{V}}_1)$, and we can then display the results in the usual format.

```
. local rownames: rownames B1
. matrix rownames DELTA = `rownames'
. matrix rownames VARDELTAr = `rownames'
. matrix colnames VARDELTAr = `rownames'
. matrix DELTA = DELTA'
. ereturn post DELTA VARDELTAr
. ereturn display
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
married	.0057685	.0002993	19.28	0.000	.0051819	.006355
age	.0582999	.0001714	340.11	0.000	.0579639	.0586358
age_2	-.0351287	.0001949	-180.23	0.000	-.0355107	-.0347467
lN	-.0103505	.0022001	-4.70	0.000	-.0146627	-.0060383
lN_2	.0021577	.0001414	15.26	0.000	.0018806	.0024348
year2	-.0060567	.000087	-69.60	0.000	-.0062272	-.0058861
year3	.0145369	.0000935	155.55	0.000	.0143537	.01472
year4	.0073687	.0000883	83.50	0.000	.0071957	.0075417
F_1	.1060651	.0791987	1.34	0.180	-.0491615	.2612917
F_2	.1250478	.0923152	1.35	0.176	-.0558868	.3059823
(output omitted)						
F_1821	.4486626	.0328059	13.68	0.000	.3843643	.512961

These results can be compared with those of the FE(s) model reported on page 474, since these are the two models that allow for correlation between worker *and* firm fixed effects and covariates. The sample size is the same in both models. Parameter estimates from the CMD model appear close, well within the 95% confidence intervals from the FE(s) model.

The χ^2 statistic to test whether the restriction imposed by pooling movers and nonmovers is given by (9).

```
. matrix DELTA = e(b)'
. matrix VARDELTA = e(V)
. matrix BETA = DELTA["married".."year4",1]
. matrix PSI = DELTA["F_1".."F_$J1",1]
```

```
. matrix CHI2r = ((B1-DELTA)'*syminv(V1r)*(B1-DELTA) +
> (BETA2-BETA)'*V2inv[`K',`K']*(BETA2-BETA))
. display as text "Chi^2 statistic: " el(CHI2r,1,1)
Chi^2 statistic: 344.05358
. display as text "P-value: " chi2tail(`K',el(CHI2r,1,1))
P-value: 1.682e-69
```

Our test statistic strongly rejects the pooling hypothesis H_0: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$, namely, that the models for movers and nonmovers are the same. Therefore, fitting this model by LSDV, even if one could, would be wrong.

Postestimation analysis of error components

The CMD method allows us to recover the estimates of θ_i and $\psi_{j(i,t)}$ so that they can be analyzed and possibly used in auxiliary regressions such as (12) and (13).

We have a vector, $\widehat{\boldsymbol{\psi}}$, that contains the firm-level error component for each firm that has a mover. Using (10), we can map this vector back to the data in the following way:

```
. use cmd, clear
. egen j1 = group(j) if plantin>0
. generate psi=.
. forvalues j=1(1)`J1' {
  2. quietly replace psi = PSI[`j',1] if j1==`j'
  3. }
. assert psi==. if plantin==0
```

The variable `psi` now contains the appropriate value of $\widehat{\boldsymbol{\psi}}$. We need a new variable, `j1`, that contains the index only for those firms with movement. We then `assert` that plants with no movers do not have an estimated value of $\psi_{j(i,t)}$.

Another complication is that estimates of $\widehat{\psi}_j$ cannot be directly compared across groups, as defined on page 468: which $\psi_{j(i,t)}$ is set equal to zero for normalization in each group is arbitrary. The same issue applies to the resulting $\widehat{\theta}_i$. Abowd, Creecy, and Kramarz (2002) therefore suggest normalizing estimates of $\psi_{j(i,t)}$ so that they have the same mean (zero) across groups. To do this, we must first define the groups. We have written an ado-file that creates a new variable to record which group each firm is in. The syntax of `grouping` is simple:

`grouping` *newvar*, `ivar(`*varname*`)` `jvar(`*varname*`)`

(*Continued on next page*)

In our data, we have 33 groups, as shown below.

```
. grouping g, ivar(i) jvar(j)
New variable g contains grouping indicator
Group 1: 4857672 person-years allocated to groups
Group 2: 4860228 person-years allocated to groups
Group 3: 4862179 person-years allocated to groups
Group 4: 4863445 person-years allocated to groups
Group 5: 4863770 person-years allocated to groups
Group 6: 4865494 person-years allocated to groups
  (output omitted)
Group 30: 4881458 person-years allocated to groups
Group 31: 4882145 person-years allocated to groups
Group 32: 4882738 person-years allocated to groups
Group 33: 4883331 person-years allocated to groups
```

But almost all rows in the data belong to group 1. All those workers in plants with movement to other plants are allocated to a group. To normalize the estimates of $\psi_{j(i,t)}$ across groups, type

```
. egen psigbar = mean(psi), by(g)
. replace psi = psi-psigbar
```

To recover estimates of θ_i, we can use (11). The easiest way to implement this task in Stata is to use the `matrix score` command:

```
. matrix x = DELTA["married".."year4",1]'
. matrix score xb = x
. gen theta_it = lw - xb - psi
. egen theta = mean(theta_it), by(i)
```

Finally, we can estimate the auxiliary regressions and see whether the components are themselves correlated. In the i-level regression, we allow for within-j clustering because errors may be correlated across individuals within the same firm.

```
. regress theta female if t==1, cluster(j)
Regression with robust standard errors                 Number of obs = 1816368
                                                       F(  1,  1819) =  116.02
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0071
Number of clusters (j) = 1820                          Root MSE      =  .35695
```

theta	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
female	-.0724758	.0067285	-10.77	0.000	-.0856722	-.0592794
_cons	7.986986	.0057305	1393.77	0.000	7.975747	7.998225

```
. by j, sort: gen n=_n
```

```
. regress psi single if n==1, robust
Regression with robust standard errors                 Number of obs =    1821
                                                       F(  1,  1819) =    8.06
                                                       Prob > F      =  0.0046
                                                       R-squared     =  0.0048
                                                       Root MSE      =  .17472

------------------------------------------------------------------------------
             |               Robust
         psi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      single |   -.024473   .0086226    -2.84   0.005    -.0413842   -.0075618
       _cons |   .0105364   .0046327     2.27   0.023     .0014505    .0196223
------------------------------------------------------------------------------
```

The first regression sample comprises those 1,816,368 workers who work in the 1,821 plants for which we can estimate $\psi_{j(i,t)}$ (see table 1). The second regression sample comprises one observation for each of these 1,821 plants.

The coefficients on `female` and `single` are both smaller than those estimated from the original OLS regression, suggesting that these original estimates were biased. However, these auxiliary regressions impose the usual identifying assumptions that the unobserved component of the error is uncorrelated with the observed component, so $\text{Cov}(\mathbf{u}_i, \alpha_i) = 0$ and $\text{Cov}(\mathbf{q}_{j(i,t)}, \phi_{j(i,t)}) = 0$.

5 Conclusion

We have shown how, using standard Stata code, one can fit fixed-effects three-way error-components models.

Researchers who are interested in estimating unobserved i- and j-level heterogeneity, and who have many j-level units, must use the direct least-squares algorithm of Abowd, Creecy, and Kramarz. In this paper, we explained how the researcher can make the feasible number of plants as large as possible without having to resort to the direct least-squares algorithm. Our CMD method is virtually identical to the correct FEiLSDVj method and differs only because the error variances are different in the mover and nonmover regressions.[10]

The estimates of $\psi_{j(i,t)}$ rely entirely on workers who change plants, as in any fixed-effects model. If one has a sample of plants, as here, there are few movers (we have 1.9 million workers but only 23,000 movers). The estimates of $\psi_{j(i,t)}$ therefore need interpreting with caution.

If researchers are not interested in estimating the worker and firm heterogeneities themselves but merely want to control for them, using spell-level fixed effects is straightforward.

10. In fact, these are estimated as 0.0851^2 and 0.068^2, respectively.

Acknowledgments

We thank the IAB for kindly supplying the data used in this paper, particularly Lutz Bellmann, Stephan Bender, and Holger Alda. We also thank the Institute of Social and Economic Research at Essex, as well as the economics departments at Aberdeen, Kent, Manchester, and Warwick.

We thank those who offered comments to our presentations at the Symposium of Multisource Databases, Universität Erlangen–Nürnberg, July 2004, and the 10th Annual Stata Users Group conference, London, 2004.

Financial support from the British Academy under grant SG-35691 is also gratefully acknowledged.

This paper's views are solely those of the authors, not the IAB. We performed all calculations with Stata/SE 9.

6 References

Abowd, J., R. Creecy, and F. Kramarz. 2002. Computing person and firm effects using linked longitudinal employer–employee data. Technical Report 2002-06, U.S. Census Bureau. http://lehd.dsd.census.gov/led/library/techpapers/tp-2002-06.pdf.

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

Alda, H., S. Bender, and H. Gartner. 2001. The linked employer–employee dataset of the IAB (LIAB). IAB Discussion Paper 06/2005. http://doku.iab.de/discussionpapers/2005/dp0605.pdf.

Baltagi, B. H. 2005. *Econometric Analysis of Panel Data*. 3rd ed. New York: Wiley.

Bender, S., A. Haas, and C. Klose. 2000. IAB employment subsample 1975–1995: Opportunities for analysis provided by the anonymised subsample. IZA Discussion Paper No. 117. ftp://ftp.iza.org/dps/dp117.pdf.

Chamberlain, G. 1984. Panel data. In *Handbook of Econometrics*, ed. Z. Griliches and M. Intrilligator, vol. 2, 1247–1318. Amsterdam: Elsevier.

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

StataCorp. 2005. *Stata 9 Longitudinal/Panel Data Reference Manual*. College Station, TX: Stata Press.

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

About the authors

Martyn Andrews is Senior Lecturer in Econometrics in the School of Economic Studies, University of Manchester, Manchester, UK. His research interests are in applied labor economics. His recent work includes papers on linked employer–employee data, employer search and matching, and the economics of education.

Thorsten Schank is a researcher at Friedrich-Alexander-Universität Erlangen–Nürnberg, Nürnberg, Germany. His research interests are in applied labor economics.

Richard Upward is an associate professor at the School of Economics, University of Nottingham, Nottingham, UK. His research interests are primarily in applied labor economics and applied microeconometrics. His recent work includes papers on linked employer–employee data, employer search and matching, labor mobility, and unemployment.