Using Stata for a memory saving fixed effects estimation of the three-way error component model[*]

Thomas Cornelißen[†]

October 2006

Abstract

The paper proposes a memory saving decomposition of the design matrix to facilitate fixed effects estimation of the three-way error component model with high numbers of observations and groups.

A common way to estimate such a model is to include two of the effects as dummy variables and to sweep out the other effect by the fixed effects transformation. If the number of groups is high, the design matrix that includes the dummy variables can be prohibitively large for computer packages that need to store the whole data set in memory. The decomposition of the design matrix proposed here shows a way of how to create the cross-product matrices for the least squares normal equations without explicitly creating the dummy variables for the group effects. As the cross-product matrices are of much lower dimension than the design matrix, this procedure reduces the computer memory required considerably. For example, a model computation shows that in a linked employer-employee data set with 20 million observations and 10 thousand firms, the memory requirement drops from 800 gigabytes to 1 gigabyte. The method is implemented in Stata by making use of the new Mata environment available in Stata 9.0. Besides implementing the memory-saving estimation method, the program also takes care of identification issues (grouping algorithm) and provides useful summary statistics. The paper presents the Stata program and comments its output.

Keywords: Linked employer-employee data, fixed effects, three-way error component model

JEL classification: C23, C63

1 Introduction

The three-way error component model has received a revived interest in labour economics due to the rising availability of linked employer-employee data which allow to analyse many problems from a different angle than before, for example by controlling for unobserved heterogeneity of workers and firms (Abowd and Kramarz 1999, Hamermesh 1999). The model applies not only to linked employer-employee data, but also to other types of matched data sets, such as for example data concerning pupils and schools.

If the researcher is concerned that unobserved heterogeneity is correlated with the observed characteristics, the model has to be estimated by fixed effects methods. In recent years, such methods have been applied to the estimation of wage equations in linked employer-employee data sets for different countries (see for example Goux and Maurin 1999, Abowd, Kramarz and Margolis 1999, Abowd, Creecy and Kramarz 2002, Gruetter and Lalive 2004, Alda 2006). Due to the size of the linked employer-employee data sets used in these studies, the authors usually encounter computer restrictions and resort to approximative methods, use time-consuming iterative solutions or apply the exact method to only a reduced number of groups.

This paper deals with one particular restriction that researchers implementing fixed effects methods of the three-way error component model may encounter: limitations of computer memory for storing a large number of dummy variables in order to capture one of the effects. The paper presents a memory-saving way to estimate the three-way fixed effects model and presents a Stata implementation of the method. The Stata implementation is a ready-to-use ado file which also takes care of the identification problem and provides useful summary statistics alongside with the estimation.

The paper proceeds as follows. Section 2 presents the model, section 3 points out the computer restrictions and shows the potential reduction in required memory size that can be gained by implementing the proposed method. Section 4 describes the decomposition of the design matrix on which the memory saving estimation procedure is based and section 5 summarizes the steps of the estimation. Section 6 presents the implementation of the method in a Stata ado-file and comments the output of the program. Section 7 concludes. Throughout the paper I refer to linked employer-employee data sets calling the two effects to be estimated person and firm effects. I will refer to stayers as those individuals that are observed in only one firm and to movers as those who are observed in several firms. Of course, the case applies also for other types of matched data sets.

2 A linear fixed effects model with three-way error component

Consider the model:

$$\tilde{y} = \tilde{X}\beta + \tilde{D}\theta + \tilde{F}\psi + \tilde{\epsilon}, \tag{1}$$

where $\tilde{X}(N^* \times K)$ is the design matrix of time varying characteristics; $\tilde{F}(N^* \times J)$ is the design matrix for the firm effects; and $\tilde{D}(N^* \times N)$ is the design matrix for the person effects. N^* is the number of person-years in the dataset, J is the number of firms, N is the number of persons and K is the number of time varying regressors. The ˜ reflects that (1) is the untransformed model.

Only two of the error components are written out explicitly. The third effect is the time-dimension which is necessary in order to identify the two other effects. For ease of notation it is subsumed in $\tilde{X}$ together with the other time-varying regressors.

A common way to estimate such a model is to include one of the effects (here the firm effect) as dummy variables, and to sweep out the other effect (here the person effect) by the within transformation or fixed effects transformation[1]. This transformation consists in subtracting the group mean (here the person mean) for all observations. The D matrix becomes the null matrix, the person effects are eliminated from the model. Abowd, Kramarz and Margolis (1999) note that this procedure is algebraically equivalent to the full dummy variable model. Andrews et al. (2006) call this procedure the „FEiLSDVj" method in order to emphasize that the model combines the classical fixed effects (FE) model and the least squares dummy variable model (LSDV), as one effect is eliminated by the fixed effects transformation and the other included as dummy variables.

Write the transformed model as:

$$y = X\beta + F\psi + \epsilon, \tag{2}$$

where ϵ is an error term satisfying the assumptions of the classical linear regression model.

---

[1]The third effect (time effect) can also be incorporated as dummy variables. Because the time dimension is usually limited, the number of time effects is usually small and therefore does not pose a special problem.

The system of normal equations is:

$$A \begin{pmatrix} \beta \\ \psi \end{pmatrix} = B \tag{3}$$

with

$$A = \begin{pmatrix} X'X & X'F \\ F'X & F'F \end{pmatrix} \text{ and} \tag{4}$$

$$B = \begin{pmatrix} X'y \\ F'y \end{pmatrix} \tag{5}$$

3   Computer restrictions

Computational problems of estimating the parameter vector from these normal equations arise

1. when (X, F) is too large to fit the memory, as some software packages, such as Stata, require the design matrix to be stored in memory, or

2. when A ist too large to be inverted, or

3. when the number of observations is so large that numerical problems arise when forming the sums that are elements of the cross product matrices A and B.

The main purpose of this paper is to propose a solution to point 1. However, the program set up in Stata to implement the solution also takes points 2 and 3 into account by using the Mata environment. Provided that there is enough computer memory, Mata can handle matrices of a dimension of up to 2 billion × 2 billion compared to only 11,000 × 11,000 in the Stata environment (Stata SE). Mata also provides computer routines with high numerical precision that can help to take into acocunt point 3. In addition, cross-checking the results obtained with those obtained in similar but smaller data sets is another way to test whether the size of the data set poses problems of numerical precision[2].

Other ways to handle the estimation problem are the approximate procedures as well as two-step and iterative solutions solutions to the exact problem presented

---

[2]Helpful comments about that topic can be found under the thread „data set larger than RAM" on the Statalist discussion board http://www.stata.com/statalist/archive/ .

in Abowd, Kramarz and Margolis (1999), Abowd, Creecy and Kramarz (2002), Andrews et al. (2006a) and Grütter (2006)[3]. If the sole aim is to control for unobserved heterogeneity and not to compute the person and firm effects explicitly, the „spell fixed effects" method proposed in Andrews et al. (2006a) is a good alternative to the „FEiLSDVj" method.

Table 1 illustrates which memory requirements can come about in practical applications. The table reports two samples of the German linked employer-employee data set (LIAB) provided by the Institute for Employment Research (IAB). The first column summarizes the estimation sample used in Alda (2006) for the estimation of person and firm fixed effects in a wage equation. This is based on the longitudinal version 2 of the LIAB data. The second column of the table shows a sample that could be generated from the LIAB cross-sectional version 1.

The computation of the memory requirements is built on the assumption that all firm effects are identifyable and that the storage of each element of the design matrix requires 4 bytes. In practice, however, not all firm effects are identifyable and the memory requirement will be less. The table therefore presents an upper bound of memory requirement[4].

When applying the „FEiLSDVj" method described above by direct creation of the time-demeaned firm dummy variables, the design matrix (X, F) must be stored. The size of this matrix is reported in the before last row of Table 1 labeled „Sum X and F". The memory requirement to store the design matrix is roughly 17 gigabytes in the sample of the first column and roughly 800 gigabytes in the sample of the second column. This is far more than the memory and, in the second example, even the hard disk space available to most researchers.

It therefore seems that the estimation of person and firm effects using the „FEiLSDVj" method with several millions of observations and several thousands of firms would be impossible with restricted memory ressources. However, note that while the design matrix (X, F) is of dimension $(N^* \times (K + J))$, the cross-product matrices A and B given in (4) and (5) are of dimension $((K + J) \times (K + J))$ and

---

[3]The classical minimum distance estimator proposed by Andrews et al. (2006a) delivers the same coefficient estimates as the „FEiLSDVj" method, but it delivers different standard errors, because it is based on separate estimations for movers and stayers, and the error term variance of both estimations is not constrained to be equal.

[4]Andrews et al. (2006a) suggest to multiply the time-demeaned firm dummies by the least common multiple of the number of observations per person. This transforms the time-demeaned firm dummies into integers and lowers the storage requirements per element to 2 bytes. With this procedure one can half the storage requirements for the F matrix. However, in many cases this reduction will not be enough to meet the memory limitations the researcher is confronted with.

Table 1: Memory requirements for two different sample configurations of German LEE data

	Sample of Alda 2006	LIAB cross–sectional version
Data set	LIAB longitudinal v.2	LIAB cross-sectional v.1
Observations (N*)	2,200,000	20,000,000[1)]
Persons (N)	670,000	-
Firms (J)	1,900	10,000[2)]
No. of regressors (K)[3)]	50	50
Size of X in (in MB)[4)]	440	4,000
Size of F in (in MB)[5)]	16,720	800,000
Sum X and F (in MB)	17,160	804,000
Size of $(X, F)'(X, F)$ (in MB)[6)]	15	404

Notes: All values in the table are approximate. [1)]Computed as the sum of yearly observations given in Table 2 in Alda 2005. [2)]According to Alda et al. 2006, up to 2004 around 10,000 different establishments have taken part in the IAB establishment panel. [3)]Hypothetical value for illustration purposes. [4)]Computed as N* x K x 4 bytes. [5)]Computed as N* x J x 4 bytes. [6)]Computed as (K+J) x (K+J) x 4 bytes

$((K + J) \times 1)$ only. They require much less storage space. The last row of Table 1 reports the memory requirement for $A = (X, F)'(X, F)$, which is only 15 megabytes for the example in the first column and 404 megabytes for the example of the second column. Therefore, once the cross-product matrices have been computed, far less than 1 gigabyte of memory is sufficient to store those matrices in both of the above examples[5].

How can A and B be computed, if the underlying design matrix (X, F) does not fit into the memory? A solution for that problem lies in the fact that each element of A and B is a cross product sum of no more than two regressors. This implies that for computing one element of A or B, only two regressors need to be stored in memory. While the X-part of the design matrix is provided as a dataset, the F-part of the cross-product matrix can be created during the estimation process without actually generating the F-part of the design matrix, i.e. the dummy variables. The information needed for that purpose is condensed in the group identifiers. The following decomposition is based on the fact that the F matrix is a sparse matrix, i.e. large

---

[5]The cross product matrix $B = (X, F)'y$ is negligibly small compared to the matrix A.

parts of it are null sub-matrices which deliver no contribution to A or B. Therefore, in the process of the formation of A and B, only certain parts of the F matrix need to be created and time and memory can be saved.

4 The decomposition

Let the persons in the dataset be indexed by i ($i = 1 \ldots N$) and the time periods for each individual be indexed by t ($t = 1 \ldots T_i$). T_i is the number of time periods that individual i is observed. The total number of observations is then $N^* = \sum_i T_i$. The vector y and the design matrices X and F in (2) have row dimension N^* and rows are indexed by the index it. The columns of X are indexed k ($k = 1, \ldots K$) and the columns of F are indexed j ($j = 1 \ldots J$).

The decomposition starts from the idea that the A and B matrices can be decomposed by observations or subsets of observations (this idea is also developed in Ritchie 1995). For example, A (B) can be represented as a sum of matrices A_i (B_i) for each individual:

$$A = \sum_i A_i = \sum_i \begin{pmatrix} X_i'X_i & X_i'F_i \\ F_i'X_i & F_i'F_i \end{pmatrix} \text{ and} \tag{6}$$

$$B = \sum_i B_i = \sum_i \begin{pmatrix} X_i'y_i \\ F_i'y_i \end{pmatrix}, \tag{7}$$

where X_i is ($T_i \times K$), F_i is ($T_i \times J$) and y_i is ($T_i \times 1$). The matrices involve only those observations that are associated with individual i.

For the current purpose it makes sense to do the individual-wise decomposition only for those parts of the matrices, where the F matrix is involved, i.e.:

$$A = \begin{pmatrix} X'X & 0 \\ 0 & 0 \end{pmatrix} + \sum_i \begin{pmatrix} 0 & X_i'F_i \\ F_i'X_i & F_i'F_i \end{pmatrix} \text{ and} \tag{8}$$

$$B = \begin{pmatrix} X'y \\ 0 \end{pmatrix} + \sum_i \begin{pmatrix} 0 \\ F_i'y_i \end{pmatrix}. \tag{9}$$

The decomposition continues with the idea that the F matrix has a different structure for stayers and for movers. In this context movers are defined as workers who change employer at least once during the whole observation period and stayers are those workers who never change employer.

Recall that the model is a transformed model. Group means by person have been substracted („time-demeaning" /"within-transformation"). As stayers never change firms, the time-demeaned firm dummies are all zero. The F matrix for stayers is the

null matrix.

Therefore we get:

$$A = \begin{pmatrix} X'X & 0 \\ 0 & 0 \end{pmatrix} + \sum_{i \in \text{ Movers}} \begin{pmatrix} 0 & X_i'F_i \\ F_i'X_i & F_i'F_i \end{pmatrix} \text{ and} \tag{10}$$

$$B = \begin{pmatrix} X'y \\ 0 \end{pmatrix} + \sum_{i \in \text{ Movers}} \begin{pmatrix} 0 \\ F_i'y_i \end{pmatrix}. \tag{11}$$

(10) and (11) are important simplifications of (8) and (9). As the F matrix is the null matrix in the sub sample of stayers, the cross product sub matrices $X'F$, $F'F$ and $F'y$ only need to be computed for movers, which are usually a comparatively small fraction of the sample.[6] As these matrices can be computed individual by individual, the F matrix does not need to exist completely at any point of time. For example, it suffices to create the matrix F_i for one individual and to compute $X_i'F_i$, $F_i'F_i$ and $F_i'y_i$. F_i is of dimension $(T_i \times J)$ and therefore should fit the memory. However, by analysing the structure of F_i more precisely, the matrix can be reduced further and more memory space can be saved.

Look at F_{i^*} for a worker i^* who is observed at $T_{i^*} = 3$ different points in time and changes the firm once. The non time-demeaned matrix $\tilde{F}_{i^*}$ is:

$$\tilde{F}_{i^*} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \ldots & 0 \\ 1 & 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 1 & 0 & \ldots & 0 \end{pmatrix} \tag{12}$$

Worker i^* is employed during two time periods in firm 1 and during the third time period in firm 4. He is never employed in any other firm, which means that to the right the individual F matrix is filled up with zeros.

The corresponding time-demeaned design matrix of the firm effects for individual i^* is:

$$F_{i^*} = \begin{pmatrix} 1/3 & 0 & 0 & -1/3 & 0 & \ldots & 0 \\ 1/3 & 0 & 0 & -1/3 & 0 & \ldots & 0 \\ -2/3 & 0 & 0 & 2/3 & 0 & \ldots & 0 \end{pmatrix} \tag{13}$$

What is important is that for each worker, only very few columns of F_{i^*} will be different from null vectors, because a given worker is employed in very few firms relative to the total set of firms. Consequently, many elements of the cross product matrices $X_i'F_i$, $F_i'F_i$ and $F_i'y_i$ are equal to zero. In the appendix (section 7), $(X_{i^*}'F_{i^*})$,

---

[6]The matrix $F'X$ is the transpose of $X'F$ an therefore in what follows it is not discussed seperately.

$(F'_{i*}F_{i*})$ and $(F'_{i*}y_{i*})$ are computed for the above example. In $(F'_{i*}F_{i*})$ the only non-zero elements are those where both row and column indices refer to a firm where worker i was employed at some moment of time. In $(X'_{i*}F_{i*})$ only the columns that are indexed with reference to a firm where worker i was employed are non-zero. In $(F'_{i*}y_{i*})$ only the rows that are indexed with reference to a firm where worker i was employed are non-zero.

A typical worker is usually employed in very few firms and thus contributes to only very few elements of the cross product matrices. F_i is a sparse matrix, and so are $(X'_{i*}F_{i*})$, $(F'_{i*}F_{i*})$ and $(F'_{i*}y_{i*})$. One can write F_i more compactly by leaving out the zero columns. Call this reduced matrix F_i^S. This is a $T_i \times s$ matrix, where s is the number of firms in which individual i was employed. In the above example, F_i^S would be a (3×2) matrix which reads:

$$
F_{i*}^S = \begin{pmatrix} 1/3 & -1/3 \\ 1/3 & -1/3 \\ -2/3 & 2/3 \end{pmatrix} \tag{14}
$$

Instead of computing $(X'_i F_i)$, $(F'_i F_i)$ and $(F'_i y_i)$ one can compute $(X'_i F_i^S)$, $(F_i^{S'} F_i^S)$ and $(F_i^{S'} y_i)$, which saves memory and time. However, one needs the information to which firm the columns of F_i^S refer, because once the cross products are computed, the results need to be added to the correct elements of the A and the B matrix, which is not a problem because this information is stored in the group identifiers. The next section summarises the algorithm for the fixed effects estimation of the linear three-way error component model in a memory saving way.

5 The algortihm to compute the least squares solution

The memory-saving way to compute the matrices A and B of the normal equations uses the information in which firm a given worker is employed. This allows to compute only those elements of A and B that the worker contributes to. The zero elements of the sparse matrices involved are dropped from the computations.

The steps are the following:

1. Create null matrices A of dimension $((K + J) \times (K + J))$ and B of dimension $((K + J) \times 1)$.

2. Compute $X'X$ and $X'y$ on the combined sample of movers and stayers. Fill

in these cross products at the appropriate sub matrices of A and B as shown in (8) and (9).

3. For each mover $i(i \in \text{Mover})$ create the time-demeaned matrix F_i but leave out columns that are zero, call this reduced matrix F_i^S. This is a $T_i \times s$ matrix, where s is the number of firms in which individual i was employed. Now,

   a. form $F_i^{S\prime} F_i^S$ and update the A matrix by adding the resulting cross-products to the appropriate elements of A,

   b. form $X_i^\prime F_i^S$ as well as its transpose $(X_i^\prime F_i^S)^\prime = F_i^{S\prime} X_i$ and update the A matrix by adding the resulting cross-products to the appropriate elements of A,

   c. form $F_i^{S\prime} y$ and update the B matrix by adding the resulting cross-products to the appropriate elements of of A.

4. Once A and B are completed, solve for the coefficient vector (β, ψ).

6   Implementation in Stata

This section uses a small simulated LEE data set in order to illustrate the Stata implementation of the estimation method. The memory-saving estimation of the „fixed effects least squares dummy variables regression" has been implemented in the stata ado-file *felsdvreg* which is available from the author upon request. This section shows that *felsdvreg* leads to exactly the same results that are obtained when the „FEiLSDVj" method is implemented in the usual way by creating the dummy variables in the data set. The section also comments and explains the output generated by *felsdvreg*.

The data set used for the illustration has 100 observations. It comprises 20 workers, for which the dummy variables $p1 \ldots p20$ have been created, and 15 firms, for which the dummy variables $f1 \ldots f20$ have been created. The dependent variable is called y, the two independent time-varying regressors are called $x1$ and $x2$.

A simple pooled linear regression of y on $x1$ and $x2$ gives the following result:

```
. reg y x1 x2
      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  2,    97) =    5.26
       Model | 1121.26227      2  560.631137           Prob > F      =  0.0068
    Residual | 10341.1204     97  106.609489           R-squared     =  0.0978
-------------+------------------------------           Adj R-squared =  0.0792
       Total | 11462.3827     99  115.781643           Root MSE      =  10.325
```

```
--------------------------------------------------------------------------
        y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------+---------------------------------------------------------------
       x1 |    1.11365   .3580307     3.11   0.002     .4030576    1.824241
       x2 |  -.3307562    .348666    -0.95   0.345    -1.022762    .3612493
    _cons |  -2.158526   1.046275    -2.06   0.042    -4.235091   -.0819599
--------------------------------------------------------------------------
```

The omission of person and firm fixed-effects is likely to bias the coefficient estimates if the omitted heterogeneity is correlated with the included regressors.

A common way to estimate a model with person and firm fixed effects is to include the firm effects as dummies and to eliminate the person affects by the within transformation („FEiLSDVj" method). This can be implemented as follows:

```
. xtreg y x1 x2  f1-f15, fe i(i)

Fixed-effects (within) regression          Number of obs      =       100
Group variable (i): i                      Number of groups   =        20

R-sq:  within  = 0.6518                     Obs per group: min =         1
       between = 0.0000                                    avg =       5.0
       overall = 0.2932                                    max =         9

                                           F(11,69)           =     11.74
corr(u_i, Xb)  = -0.2330                    Prob > F           =    0.0000


--------------------------------------------------------------------------
        y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------+---------------------------------------------------------------
       x1 |   1.029258   .2151235     4.78   0.000     .6000987    1.458418
       x2 |   -.709482   .2094198    -3.39   0.001    -1.127263   -.2917009
       f1 |  (dropped)
       f2 |  (dropped)
       f3 |   -13.2617   3.258081    -4.07   0.000    -19.76139   -6.762004
       f4 |  (dropped)
       f5 |   .6932951   3.139439     0.22   0.826    -5.569713    6.956304
       f6 |  -5.433107   3.908214    -1.39   0.169    -13.22978    2.363566
       f7 |    3.12687   3.882176     0.81   0.423    -4.617858     10.8716
       f8 |  (dropped)
       f9 |   6.016403   4.967271     1.21   0.230    -3.893031    15.92584
      f10 |  -16.76837   3.245567    -5.17   0.000     -23.2431   -10.29364
      f11 |  (dropped)
      f12 |  -6.752864   3.941654    -1.71   0.091    -14.61625    1.110519
      f13 |  -2.526721   3.844219    -0.66   0.513    -10.19573    5.142287
      f14 |  -12.72366   3.047553    -4.18   0.000    -18.80336   -6.643961
      f15 |  (dropped)
    _cons |    2.87504   1.113735     2.58   0.012     .6531994     5.09688
----------+---------------------------------------------------------------
  sigma_u |   9.170253
  sigma_e |  5.4861156
      rho |  .73642876   (fraction of variance due to u_i)
--------------------------------------------------------------------------
F test that all u_i=0:     F(19, 69) =      4.01            Prob > F = 0.0000
```

Stata drops six of the fifteen firm effects, which apparently are not identified. The coefficient estimates on $x1$ and $x2$ differ from those in the pooled regression, reflecting the effect of controlling for unobserved heterogeneity.

Some of the non-identified firm effects belong to firms without movers, because for a firm without movers the firm effect is never identified. The firms with movers can be divided into groups within which there is worker mobility, but between which there is no mobility. Within each such group, one firm effect is not identified, i.e. one firm effect has to be taken as the reference, and all other firm effects are expressed as differences from the reference firm.

In the present sample data set, two firms have no movers (firm 1 and firm 2), and the remaining firms are divided into 4 groups. Therefore, six firm effects are not identified. However, it is arbitrary which firm effect in each group is used as the reference.

The table shows to which group the firms with movers belong[7]:

Group	Firms
1	3,4,5
2	6,7,8,9
3	10,11,12
4	13,14,15

When looking at the firm dummies that have been dropped from the above estimation, we see that in the *xtreg* command firm 4 is the reference in group 1, firm 8 in group 2, firm 11 in group 3 and firm 15 in group 4.

We can repeat the estimation and decide by ourselves which firm effect to drop in each group. When dropping the first firm in each group (i.e. the firm with the lowest firm ID), the results read:

```
. xtreg y x1 x2  f4-f5 f7-f9 f11-f12 f14-f15, fe i(i)

Fixed-effects (within) regression          Number of obs     =        100
Group variable (i): i                      Number of groups  =         20

R-sq:  within  = 0.6518                     Obs per group: min =          1
       between = 0.0015                                    avg =        5.0
       overall = 0.0913                                    max =          9

                                            F(11,69)          =      11.74
corr(u_i, Xb)  = -0.5330                    Prob > F          =     0.0000
```

---

[7]An algorithm to determine the groups is derived in Abowd et al. (2002). The algorithm is included in the program *felsdvreg* drawing on the Stata code by Andrews at al. 2006a

```
------------------------------------------------------------------------------
         y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------+------------------------------------------------------------------
        x1 |   1.029258   .2151235     4.78   0.000     .6000987    1.458418
        x2 |  -.709482    .2094198    -3.39   0.001    -1.127263   -.2917009
        f4 |    13.2617   3.258081     4.07   0.000     6.762004    19.76139
        f5 |   13.95499   2.818964     4.95   0.000     8.331314    19.57867
        f7 |   8.559977   3.882525     2.20   0.031     .8145504     16.3054
        f8 |   5.433107   3.908214     1.39   0.169    -2.363566    13.22978
        f9 |   11.44951   4.792492     2.39   0.020     1.888749    21.01027
       f11 |   16.76837   3.245567     5.17   0.000     10.29364     23.2431
       f12 |   10.01551   3.407205     2.94   0.004     3.218319     16.8127
       f14 |  -10.19694   3.074528    -3.32   0.001    -16.33046   -4.063427
       f15 |   2.526721   3.844219     0.66   0.513    -5.142287    10.19573
     _cons |  -6.044057    1.03021    -5.87   0.000     -8.09927   -3.988844
-----------+------------------------------------------------------------------
   sigma_u |  10.169633
   sigma_e |  5.4861156
       rho |  .77458273   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(19, 69) =      8.64             Prob > F = 0.0000
```

The relative effect between firms in the same group are exactly the same (e.g. the difference between f6 and f7 is 8.559977 in both *xtreg* estimations), but between groups they are not the same, because relative firm effects between groups are affected by the decision which firm serves as the reference in each group. In fact, relative firm effects between groups are not identified and therefore firm effects should not be compared between groups.

The person effects can be recovered by:

```
. predict peffxt, u
. tab i, sum(peffxt)

           |        Summary of u[i]
         i |        Mean   Std. Dev.       Freq.
-----------+------------------------------------
         1 |  -9.3451653           0           1
         2 |  -3.7514441           0           2
         3 |   12.987279           0           2
         4 |  -4.6659427           0           4
         5 |  -3.8792346           0           1
         6 |   1.1379685           0           6
         7 |  -.44613671           0           6
         8 |   .45241559           0           4
         9 |  -16.23423           0           1
        10 |  -12.18615           0           9
        11 |  -.40414947           0           9
        12 |  -3.9539671           0           6
        13 |  -11.948536           0           6
        14 |  -4.2723632           0           5
        15 |   1.7324725           0           2
```

13

```
        16 |   -11.586735              0              4
        17 |   -13.570383              0              7
        18 |    21.664909              0              8
        19 |    14.427181              0              9
        20 |    11.061296              0              8
-----------+---------------------------------
```

In this small data set, the same results can be obtained by including all person and firm effects in a pooled regression:


```
. reg y x1 x2 p1- p20 f4-f5 f7-f9 f11-f12 f14-f15, noc

      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F( 31,    69) =   10.46
       Model | 9754.77005      31  314.670002          Prob > F      =  0.0000
    Residual | 2076.72508      69  30.0974649          R-squared     =  0.8245
-------------+------------------------------           Adj R-squared =  0.7456
       Total | 11831.4951     100  118.314951          Root MSE      =  5.4861


------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   1.029258   .2151235     4.78   0.000     .6000987    1.458418
          x2 |   -.709482   .2094198    -3.39   0.001    -1.127263   -.2917009
          p1 |  -15.38922     5.5063    -2.79   0.007    -26.37399   -4.404454
          p2 |  -9.795501   3.908692    -2.51   0.015    -17.59313   -1.997874
          p3 |   6.943222   3.907763     1.78   0.080    -.8525509      14.739
          p4 |     -10.71   2.743058    -3.90   0.000    -16.18225   -5.237749
          p5 |  -9.923292   5.590565    -1.78   0.080    -21.07616    1.229581
          p6 |  -4.906089   2.592838    -1.89   0.063    -10.07866     .266481
          p7 |  -6.490194   2.584415    -2.51   0.014    -11.64596   -1.334426
          p8 |  -5.591642   4.256923    -1.31   0.193    -14.08397    2.900687
          p9 |  -22.27829   6.193958    -3.60   0.001    -34.63489   -9.921679
         p10 |  -18.23021   2.742264    -6.65   0.000    -23.70087   -12.75954
         p11 |  -6.448207   3.084513    -2.09   0.040    -12.60164   -.2947709
         p12 |  -9.998024    2.75384    -3.63   0.001    -15.49178   -4.504264
         p13 |  -17.99259   2.770329    -6.49   0.000    -23.51925   -12.46594
         p14 |  -10.31642   3.662598    -2.82   0.006     -17.6231   -3.009737
         p15 |  -4.311585   4.374143    -0.99   0.328    -13.03776    4.414591
         p16 |  -17.63079   3.830932    -4.60   0.000    -25.27329   -9.988293
         p17 |  -19.61444   2.297134    -8.54   0.000     -24.1971   -15.03178
         p18 |   15.62085   3.046805     5.13   0.000     9.542643    21.69906
         p19 |   8.383124   3.515348     2.38   0.020     1.370196    15.39605
         p20 |   5.017239   2.268042     2.21   0.030     .4926188    9.541858
          f4 |    13.2617   3.258081     4.07   0.000     6.762004    19.76139
          f5 |   13.95499   2.818964     4.95   0.000     8.331314    19.57867
          f7 |   8.559977   3.882525     2.20   0.031     .8145504     16.3054
          f8 |   5.433107   3.908214     1.39   0.169    -2.363566    13.22978
          f9 |   11.44951   4.792492     2.39   0.020     1.888749    21.01027
         f11 |   16.76837   3.245567     5.17   0.000     10.29364     23.2431
         f12 |   10.01551   3.407205     2.94   0.004     3.218319     16.8127
         f14 |  -10.19694   3.074528    -3.32   0.001    -16.33046   -4.063427
         f15 |   2.526721   3.844219     0.66   0.513    -5.142287    10.19573
------------------------------------------------------------------------------
```

The coefficient estimates on $x1$ and $x2$ as well as the corresponding standard errors are exactly the same as in the preceding two *xtreg* estimations. Person and firm effects can be read directly from the coefficient estimates. The firm effects are exactly the same than in the preceding *xtreg* estimation (p.12). The person effects only differ by a constant of -6.044057 which is the constant estimated in the *xtreg* model. *xtreg* is programmed such that the mean person effect is subtracted from the person effects and displayed as the regression constant. In other words, *xtreg* normalizes the sum of all person effects to zero.

As described in the previous section, the explicit creation of all firm dummies combined with the use of *xtreg* (let alone the creation of all person and firm dummies with the use of *reg*) can require more computer memory than is available. In the case when there is a large number of firms it can therefore be necessary to apply a memory-saving way to the solution of the „FEiLSDVj" estimator. I have programmed the algorithm presented in the preceding section as a Stata ado-file called *felsdvreg*. This routine can be applied to the present data set as follows:

```
felsdvreg y x1 x2, i(i) j(j) f(feffhat) p(peffhat) m(mover) g(group)
 xb(xb) r(res) mnum(mnum) pobs(pobs)
```

The options are the following: The option **i()** is used to pass the variable name of the person ID, the option **j()** does the same for the firm ID. The options **p()** and **f()** define names of new variables to be created in order to store the person and firm effects after estimation. So do the options **xb()** and **res()** to store the linear combinations $x'\hat{\beta}$ and the residual $\hat{\epsilon}$. The remaining options define names of new variables that store a dummy variable that indicates a person who is a mover (**m()**), a group variable that counts and identifies the number of groups of firms that are connected through mobility (**g()**), a variable that contains the number of movers per firm (**mnum()**) and a variable that stores the number of observations per persons (**pobs()**).

In the following, the output is commented in several steps:

```
. felsdvreg y x1 x2, i(i) j(j) f(feffhat) p(peffhat) m(mover) g(group)
 xb(xb) r(res) mnum(mnum) pobs(pobs)
Selecting sample and dropping superfluous observations 17:21:30
Preserve dataset 17:21:30
Fit restricted model 17:21:30
Generate smooth firm and person ids 17:21:30
Sort 17:21:30
Determine stayers and movers 17:21:30

41  Unique worker firm combinations.
```

```
Number of firms workers are employed in.
   __000006 |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |          7       35.00       35.00
          2 |          7       35.00       70.00
          3 |          4       20.00       90.00
          4 |          2       10.00      100.00
------------+-----------------------------------
      Total |         20      100.00


Number of movers out of all persons:
      mover |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |          7       35.00       35.00
          1 |         13       65.00      100.00
------------+-----------------------------------
      Total |         20      100.00


Number of observations per person:

       pobs |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |          3       15.00       15.00
          2 |          3       15.00       30.00
          4 |          3       15.00       45.00
          5 |          1        5.00       50.00
          6 |          4       20.00       70.00
          7 |          1        5.00       75.00
          8 |          2       10.00       85.00
          9 |          3       15.00      100.00
------------+-----------------------------------
      Total |         20      100.00


Distribution of number of movers per firm:

   __00000F |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |          2       13.33       13.33
       1- 5 |          7       46.67       60.00
       6-10 |          5       33.33       93.33
      11-20 |          1        6.67      100.00
------------+-----------------------------------
      Total |         15      100.00
```

Until here, the program has analysed the structure of the data set. The first table
tells us in how many firms the workers are employed. The seven workers employed
in only one firm are stayers. Out of the remaining 13 workers, 7 are observed in two
firms, 4 in three firms and 2 in four firms. The second table is a summary of the
first and gives the total number of stayers and movers. The third table indicates
the numbers of observations per person. For example, 3 workers are observed only

at one point in time, 3 workers are observed 9 times, etc.

The fourth table shows the distribution of the number of movers per firm. The purpose of this table is to get an impression of the quality of the estimation of the firm effects. The estimation of the firm effects is better the more movers there are and one might think of the firm effects that are identified by few movers as effects that are poorly measured. In this example data set two firms have no movers and all 15 firms have less than 20 movers.

The 15 firms can be divided into groups within which there is worker mobility, but between which there is no mobility. As noted above, within each such group, one firm effect is not identified, i.e. one firm effect has to be taken as the reference, and all other firm effects are expressed as differences from the reference. The *felsdvreg* program goes on by defining the groups[8]:

```
Group firms without movers under artificial firm IDs 17:21:31
Determine number of groups 17:21:31

New variable group contains grouping indicator
100 person-years (14 firms) to be allocated.
Group 0 contains firms without movers.
Group 0: 10 person-years allocated to groups. + 10(1) Left: 90(13)
Group 1: 36 person-years allocated to groups. + 26(3)   64 left.
Group 2: 51 person-years allocated to groups. + 15(4)   49 left.
Group 3: 75 person-years allocated to groups. + 24(3)   25 left.
Group 4: 100 person-years allocated to groups. + 25(3)    0 left.
```

	Person-years	Persons	Mover	Firms
group |	N(__000000)	N(__000009)	sum(__00000D)	N(__000008)
0 |	10	5	0	2
1 |	26	5	3	3
2 |	15	2	2	4
3 |	24	5	5	3
4 |	25	3	3	3
|				
Total |	100	20	13	15

```
Note: Group 0 in the table regroups firms without movers. No firm effect
in group 0 is identified.
15-2-4 = 9 firm effects are identified.
Computed as: number of firms - number of firms without movers - number
of groups (excl. group 0)
```

The two firms without movers are gathered in group 0. The remaining firms of the sample are divided into 4 groups. The table shows the number of person-years,

---

[8]The grouping algorithm incorporated in *felsdvreg* draws on the Stata implementation by Andrews et al. (2006a)

persons, movers and firms in each of the groups. As indicated, only 9 of the 15 firm effects are identified because 2 firms have no movers and their firm effects cannot be identified, and 4 more firms effects are not identified because they serve as reference in their groups.

The remaining estimation output reads as follows:

```
Time-demean variables 17:21:31
Start Mata environment 17:21:31
Compute Total sum of squares 17:21:31
Create moment matrices 17:21:31
Filling in elements for movers 17:21:31
Take out unidentified firm effects 17:21:31
Dimension A
  11
Solving for beta 17:21:31
Restore dataset 17:21:31
Generate smooth firm and person IDs again 17:21:31
Predicting x'b and assigning firm effects 17:21:31
Computing residuals and person effects 17:21:31
Solving for standard errors 17:21:31
N=100
-------------------------------------------------------------------------------
            |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------+------------------------------------------------------------------
        x1  |   1.029258   .2151235     4.78   0.000     .6000987    1.458418
        x2  |  -.7094819   .2094198    -3.39   0.001    -1.127263   -.2917009
-------------------------------------------------------------------------------
F-test that all person and firm effects are equal to zero: F(28,69)=9.81
Prob > F = 8.711e-15

If the covariances are positive, the following can be interpreted as
shares in explaining the variance of y:

Cov(y, xb) / Var(y):                          .10029458
Cov(y, peffhat) / Var(y):                     .56511312
Cov(y, feffhat) / Var(y):                     .15341486
Cov(y, res) / Var(y):                         .18117743

Job Done! 17:21:31
```

Note that the coefficient estimates on $x1$ and $x2$ and their standard errors are exactly the same as in the estimations shown above, where the dummy variables where included 'manually'.

The variance decomposition at the end of the output gives an indication of how strongly the four components (i) observed time-varying characteristics, (ii) person effects, (iii) firm effects and (iv) the residual contribute to explaining the variance of the dependent variable. The shares sum to 1, however, the covariances indicated can become negative and then it becomes difficult to interpret the numbers as shares.

We now look at the person and firm effects:

```
. tab j, sum(feffhat)

            |       Summary of feffhat
        j |        Mean   Std. Dev.        Freq.
-----------+-----------------------------------
        1 |           0           0            3
        2 |           0           0            7
        3 |           0           0            8
        4 |   13.261698           0            9
        5 |   13.954993           0            9
        6 |           0           0            5
        7 |   8.5599766           0            4
        8 |   5.4331064           0            4
        9 |    11.44951           0            2
       10 |           0           0           10
       11 |   16.768372           0            9
       12 |   10.015507           0            5
       13 |           0           0            7
       14 |  -10.196942           0           13
       15 |   2.5267205           0            5
-----------+-----------------------------------
    Total |    4.048878   8.2788071          100


. tab i, sum(peffhat)

            |       Summary of peffhat
        i |        Mean   Std. Dev.        Freq.
-----------+-----------------------------------
        1 |  -15.389222           0            1
        2 |  -9.7955017           0            2
        3 |    6.943222           0            2
        4 |      -10.71           0            4
        5 |  -9.9232922           0            1
        6 |  -4.9060888           0            6
        7 |  -6.4901938           0            6
        8 |  -5.5916419           0            4
        9 |  -22.278286           0            1
       10 |  -18.230206           0            9
       11 |  -6.4482064           0            9
       12 |   -9.998024           0            6
       13 |  -17.992594           0            6
       14 |  -10.316421           0            5
       15 |  -4.3115845           0            2
       16 |  -17.630793           0            4
       17 |  -19.614441           0            7
       18 |   15.620852           0            8
       19 |   8.3831244           0            9
       20 |   5.0172386           0            8
-----------+-----------------------------------
    Total |  -6.0440572   10.846544          100
```

The firm effect of the firm without movers and of the reference firm in each group are set to zero. The firm effects are exactly the same as in the second *xtreg* estimation (p.12) and the pooled estimation with the person and firm dummies (p.14).

The person effects are the same as in the pooled regression with the person and firm dummies (p.14), and they differ from the effects of the second *xtreg* regression (p.12) only by the constant -6.044057 of the *xtreg* model, because *felsdvreg* does not by default normalize the sum of the person effects to zero[9].

After the estimation, the researcher may be interested in correlating the person and firm effects with each other or with other regressors. However, it should be kept in mind that what is actually identified are relative person and firm effects within each group, and that person and firm effects of different groups should not be compared. This can be illustrated by computing the correlation of person and firm effects over all groups with different normalisations.

The first command correlates the person and firm effects over all groups, the second command correlates only the effects of group 1:

```
. corr  feffhat peffhat
(obs=100)

            |  feffhat  peffhat
------------+------------------
    feffhat |   1.0000
    peffhat |  -0.5645   1.0000

. corr  feffhat peffhat if group==1
(obs=26)

            |  feffhat  peffhat
------------+------------------
    feffhat |   1.0000
    peffhat |  -0.2006   1.0000
```

Now we normalise the firm and person effects so that they sum to zero within each group by subtracting the average group firm and the average group person effect. We construct a new variable *gmean* which for each group is equal to the sum of the mean firm and the mean person effect of the group. After this normalisation, the person and firm effects are deviations from the group means that are stored in *gmean*. After that we compute the correlation over all groups and the correlation

---

[9]If the option *cons* is chosen in *felsdvreg*, it does normalize the sum of the person effetcs to zero and displays a regression constant.

using only the effects of group 1:

```
. by group: egen pmean=mean(peffhat)
. by group: egen fmean=mean(feffhat)
. gen peffnorm=peffhat-pmean
. gen feffnorm=feffhat-fmean
. gen gmean=pmean+fmean

. tab group, sum(gmean)

                 |        Summary of gmean
       group |        Mean    Std. Dev.        Freq.
------------+------------------------------------
           0 |  -7.3857074           0           10
           1 |  -1.2363138           0           26
           2 |  -2.6100435           0           15
           3 |  -7.2913418           0           24
           4 |   4.8250484           0           25
------------+------------------------------------
       Total |  -1.9951788    4.6572344          100


. corr  feffnorm peffnorm
(obs=100)
             | feffnorm peffnorm
-------------+------------------
    feffnorm |   1.0000
    peffnorm |   0.0227    1.0000


. corr  feffnorm peffnorm if group==1
(obs=26)
             | feffnorm peffnorm
-------------+------------------
    feffnorm |   1.0000
    peffnorm |  -0.2006    1.0000
```

The normalisation has changed the result from the correlation over all groups[10]. It is now 0.0227 whereas before it was -0.5645. The result of the correlation within the group of -0.2006 is still the same. One could argue that the normalisation of person and firm effects to an equal group mean makes comparison across groups more appropriate and therefore the correlation over all groups after normalisation is appropriate whereas the one before normalisation was not. However, it seems difficult to argue that a deviation of +1 from a group mean of -7.29 of group 3

---

[10]This normalization is not exactly implemented in *felsdvreg*, But the program has two options for normalization: The option *normalize* normalizes the firm effects to mean zero within each group and adds the mean firm effects that are subtracted in each group to the person effects. The option *cons* normalizes the person effects to sum to zero over all observations and displays the overall mean person effect as the regression constant. Both options can be combined.

means the same as a deviation of +1 from the group mean of 4.82 in group 4. The normalisation does not change the fact that relative firm effects within groups are identified but relative firm effects between groups are not identified. It is therefore preferable to correlate only effects of the same group.

Andrews et al. (2006b) show that the correlation between worker and firm effects is biased and that the bias is greater the lower the observed worker mobility between firms. After estimation one may therefore want to select firm and person effects that fulfill certain minimum requirements with respect to the minimum number of movers per firm or the minimum number of observations per person. This is possible with the variables defined in the options **mnum()** and **pobs()** and returned by *felsdvreg*.

7 Conclusion

The paper has proposed a memory saving decomposition of the design matrix to facilitate fixed effects estimation of the three-way error component model. This is applicable, for example, to linked employer-employee data sets but it is also applicable to other matched data that allow to estimate a three-way error component model, such as pupil and school effects for example.
A common way to estimate such a model is to take into account one of the effects by including dummy variables, and to sweep out the other effect by the within transformation (fixed effects transformation). If the number of groups is high, creating and storing the dummy variables can require much computer memory space. The paper has presented two model samples for the estimation of person and firm effects in German linked employer-employee data, where the storage requirements would be 17 and 800 gigabytes respectively. To date, this requirement is prohibitively high for many researchers. The decomposition of the design matrix presented in this paper reduces the storage requirements in the model samples to below 1 gigabyte. A Stata ado file for the memory-saving computation of the fixed effects model has been described. Besides implementing the memory-saving estimation method, the program also implements a grouping algorithm to determine the identified effects and provides useful summary statistics. The program is available from the author upon request.

References

Abowd J. and Kramarz F. (1999): „The Analysis of Labor Markets Using Matched Employer-Employee Data" in O. Ashenfelter and D. Card (eds.) Handbook of Labor Economics, Volume 3B, Chapter 40 (Amsterdam: North Holland, 1999), pp. 2629-2710.

Abowd J., Kramarz F. and Margolis D. (1999): „High Wage Workers and High Wage Firms", *Econometrica*, 67(2), pp. 251-334.

Abowd J., Creecy R. and Kramarz F. (2002) „Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data", Technical Paper No. TP-2002-06, U.S. Census Bureau.

Alda H., Dundler A., Müller D. and Spengler A. (2006): „Aufbereitung eines Paneldatensatzes aus den Querschnittsdaten des IAB-Betriebspanels", FDZ Datenreport Nr. 02/2006, IAB, Nürnberg.

Alda, H. (2005): Datenbeschreibung der Version 1 des LIAB-Längsschnittmodells, FDZ Datenreport Nr. 3/2005, IAB, Nürnberg.

Alda, H. (2006): „Beobachtbare und unbeobachtbare Betriebs- und Personeneffekte auf die Entlohnung", Beiträge zur Arbeitsmarkt- und Berufsforschung Nr. 298, IAB, Nürnberg.

Andrews M., Schank T., Upward R. (2006a): „Practical fixed effects estimation methods for the three-way error components model", erscheint demnächst in: The Stata Journal

Andrews M., Gill L., Schank T., Upward R. (2006b): „High wage workers and low wage firms: negative assortative matching or statistical artefact?", Discussion Paper No. 42, Lehrstuhl für Arbeitsmarkt- und Regionalpolitik, Friedrich-Alexander-Universität Erlangen-Nürnberg.

Goux, D. and Maurin E. (1999): „Persistence of Interindustry Wage Differentials: A Reexamination Using Matched Worker-Firm Panel Data", Journal of Labor Economics, 17(3), pp. 492-533.

Gruetter M. and Lalive R. (2004) „The Importance of Firms in Wage Determination", IEW Working Paper No. 207, Institute for Empirical Research in Economics, University of Zurich.

Grütter M. (2006) „The anatomy of the wage structure„ , Verlag Dr. Kovac, Hamburg.

Hamermesh, D. „LEEping into the future of labor economics: the research potential of linking employer and employee data", Labour Economics, 1999, vol. 6, issue 1, pages 25-41.

Ritchie, F. (1995): „Efficient Access to Large Datasets for Linear Regression Models", Working Paper 95/12, Department of Economics, University of Stirling.

Appendix

In the above example, $X'_{i*}F_{i*}$, $F'_{i*}F_{i*}$ and $F'_{i*}y_{i*}$ are:

$$F'_{i*}F_{i*} = \begin{pmatrix} 1/3 & 1/3 & -2/3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1/3 & -1/3 & 2/3 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1/3 & 0 & 0 & -1/3 & 0 & \ldots & 0 \\ 1/3 & 0 & 0 & -1/3 & 0 & \ldots & 0 \\ -2/3 & 0 & 0 & 2/3 & 0 & \ldots & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \phi_{i11} & 0 & 0 & \phi_{i14} & 0 & \ldots & 0 \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 \\ \phi_{i14} & 0 & 0 & \phi_{44} & 0 & \ldots & 0 \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 \end{pmatrix}, \tag{15}$$

$$\phi_{i11} = (\frac{1}{3})^2 + (\frac{1}{3})^2 + (\frac{-2}{3})^2$$

$$\phi_{i14} = (\frac{1}{3})(\frac{-1}{3}) + (\frac{1}{3})(\frac{-1}{3}) + (\frac{-2}{3})(\frac{2}{3})$$

$$\phi_{i44} = (\frac{-1}{3})^2 + (\frac{-1}{3})^2 + (\frac{2}{3})^2$$

$$X'_{i*}F_{i*} = \begin{pmatrix} x_{i11} & x_{i21} & x_{i31} \\ x_{i12} & x_{i22} & x_{i32} \\ \vdots & \vdots & \vdots \\ x_{i1K} & x_{i2K} & x_{i3K} \end{pmatrix} \begin{pmatrix} 1/3 & 0 & 0 & -1/3 & 0 & \ldots & 0 \\ 1/3 & 0 & 0 & -1/3 & 0 & \ldots & 0 \\ -2/3 & 0 & 0 & 2/3 & 0 & \ldots & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \xi_{i11} & 0 & 0 & \xi_{i14} & 0 & \ldots & 0 \\ \xi_{i21} & 0 & 0 & \xi_{i24} & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ \xi_{iK1} & 0 & 0 & \xi_{iK4} & 0 & \ldots & 0 \end{pmatrix}, \tag{16}$$

$$\xi_{ij1} = (\frac{1}{3})x_{i11} + (\frac{1}{3})x_{i21} + (\frac{-2}{3})x_{i31}$$

$$\xi_{ij4} = (\frac{-1}{3})x_{i11} + (\frac{-1}{3})x_{i21} + (\frac{2}{3})x_{i31}$$

$$
F'_{i*} y_{i*} = \begin{pmatrix} 1/3 & 1/3 & -2/3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1/3 & -1/3 & 2/3 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{pmatrix} = \begin{pmatrix} v_{i1} \\ 0 \\ 0 \\ v_{i4} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{17}
$$

$$
v_{i1} = (\frac{1}{3})y_{i1} + (\frac{1}{3})y_{i2} + (\frac{-2}{3})y_{i3}
$$

$$
v_{i4} = (\frac{-1}{3})y_{i1} + (\frac{-1}{3}yx_{i2} + (\frac{2}{3})y_{i3}
$$