A Review of Stata Routines for Fixed Effects Estimation in Normal Linear Models

Daniel F. McCaffrey
The RAND Corporation
Pittsburgh, PA
danielm@rand.org

J. R. Lockwood
The RAND Corporation
Pittsburgh, PA
lockwood@rand.org

Kata Mihaly
The RAND Corporation
Washington, DC
kmihaly@rand.org

Tim R. Sass
Florida State University
Tallahassee, FL
tsass@fsu.edu

Abstract. Availability of large multi-level longitudinal databases in various fields including labor economics (with workers and firms observed over time) and education research (with students, teachers and schools observed over time) has increased the application of models with one or multiple levels of fixed effects (e.g., teacher and student effects). There has been a corresponding rapid development of Stata routines designed for estimating these types of models. The routines parameterize the fixed effects portions of models differently. In cases where estimates of the fixed effects parameters are of interest, it is critical to understand precisely what parameters are being estimated by different routines. This article catalogs the estimates of reported fixed effects provided by different routines for several canonical cases of both one-level and two-level fixed effects models. We also discuss issues regarding computational efficiency and standard error estimation.

Keywords: st0001, longitudinal data, linked employer-employee data, fixed effects estimators, software review

1 Introduction

In labor economics, health policy, and education research there is interest in estimating the effects of individual units, e.g., firms, hospitals, doctors, schools, or teachers from databases with measures on the units and individual persons (e.g., workers, patients, or students) attached to these units. Commonly the unit effects are estimated with fixed effects. Administrative databases with data from very large numbers of persons linked to hundreds or even thousands of units (e.g., Abowd et al. (2006); Harris and Sass (2006)) are increasingly available for such analyses. Models with huge numbers of fixed effects for firms pose computational challenges especially in cases in which the models also include fixed effects for persons. The Stata community has been active in developing routines for efficient estimation of such models including areg and xtreg and user-generated routines such as felsdvreg (Cornelissen (2008)), a2reg (Ouazad (2008)), gpreg (Guimaraes and Portugal (2009); Schmieder (2009)), and felsdvregdm (Mihaly et al. (2010)).
Models with fixed effects for units are overparameterized because the means for the individual firms cannot be estimated separately from the mean of the individual persons. In many applications where fixed effects models are used, the primary goal is the estimation of the effects of time-varying covariates, with the fixed effects for persons and often units being nuisance parameters. In such cases, estimates of the effects of time-varying covariates are invariant to different approaches to handling the overparameterization of the fixed effects. However, in many recent applications, such as studies of hospital or teacher quality (e.g., Bazzoli et al. (2008), Goldhaber et al. (2010)) there is interest in obtaining estimates of the fixed effect parameters and often standard error estimates as well. Stata routines have taken three different approaches to solving the indeterminacy due to overparameterization of unit means:

1. Estimation of unit means that conflate the unit means with the person means
2. Estimation of contrasts between each of the unit means and the mean of a “hold-out” unit
3. Estimation of contrasts between each of the unit means and the average of the unit means.

Each of these alternative parameterizations has advantages and limitations. All three lead to the same rank ordering of units by estimated unit fixed effects. However, they do not provide estimates of the same quantities and they are not all equally appropriate for all uses. For instance, estimates of the unit means are not estimates of causal effects. A large value for a unit mean does not imply a particularly effective unit because all units or the average person may have a large value of the outcome of interest. Also, analysts are increasingly interested in using post hoc “shrinkage” estimators (e.g., Jacob and Lefgren (2008)) to reduce the error variance in estimates of units with small numbers of persons (e.g., teachers with very small classes). However, shrinking estimated unit means yields estimates that cannot be compared across units because the overall mean is differentially weighted in each shrunken unit mean estimate. Contrasts between each unit mean and a holdout can be interpreted as causal effects but have the limitation of being sensitive to the arbitrary choice of the holdout unit. Moreover, in this case, the variability among the estimates yields extremely biased estimates of the variability of the true unit means.

Additional problems with indeterminacy arise when analysts want to control for unit-level variables (for cross-sectional unit data) or time-invariant unit-level variables (for longitudinal data unit data) when estimating unit effects. For example, in education, the units might be teachers effects by year and the analysts might want to control for overall year means. There is no way to separate the effects of the unit-level variables in cross-sectional data or time-invariant unit-level variables in longitudinal data from differences in unit-level effects. For instance, an analyst would not be able to determine if higher student achievement in a given year was the result of all teachers performing better in that year or the test being easier in that year; i.e., there is no way to identify both year means and all the teacher-by-year effects.
Two conventions to estimation in the face of this indeterminacy exist: 1) the unit-level variables in models for cross-sectional data or time-invariant unit-level variables in models longitudinal data are removed from the models, conflating the effects of these factors with the individual unit effects; or 2) individual unit effects are estimated only among units with the same value of the unit-level or time-invariant unit-level variables (e.g., among firms of the same size or among teachers teaching during the same school year). The common practice has been the first approach which estimates unit effects as the combined effects of the unit-level variables and units themselves (e.g. \texttt{ggreg}). However, \texttt{felsdvregdm} takes the alternative approach providing estimates of unit effects only among those units with common values on the unit-level variables. Both solutions can be useful and not all analysts will want to the use the conflated effects. Hence, analysts need to understand what has been estimated for accurate interpretation of differences among units.

With longitudinal data, the inclusion of fixed effects for persons in addition to units further complicates these issues. The Stata routines used for estimating models with a single level of fixed effects (i.e., just unit effects) must be specified differently when the model includes both unit and person effects. The change in specification can change the parameterization of the unit effects. For instance, \texttt{areg} applied to a one-level fixed effects model estimates the parameters of a model with a different parameterization of the units than the model it estimates when the model includes two levels of fixed effects. Moreover, specialized routines developed specifically for two-level fixed effect models also use different parameterizations of the unit effects that do not necessarily match those of other routines.

We have found that the subtle differences in the parameterization of the unit fixed effects across different routines is not well understood by the user community. To help clarify these differences we make some direct comparisons of estimates for the available Stata routines under scenarios designed to span the possible indeterminacies in modeling with fixed effects and highlight the differences in model parameterizations. Because some of the routines for estimating effects from models with two levels of fixed effects differ from those available for estimation of models with just one level of fixed effects and the routines available for both behave differently in these two settings, we repeat our comparisons with these two alternative modeling conditions.

In the remainder of this paper, we first consider estimates of unit effects when there are no person fixed effects and then turn to models with both unit and person fixed effects. In each case, we first explicitly specify the model and then compare estimates under three different scenarios: a simple model with only unit fixed effects (or unit and person fixed effects), a model with fixed effects and person level covariates, and a model with fixed effects and unit level covariates.
Fixed Effects Estimation in Stata

2 One Level of Fixed Effects

2.1 One-Level Fixed Effects Model

The basic model with a single level of fixed effects assumes that the outcome for a “person” \( i \) with \( K_p \) person-level predictors \( x_i \) linked to “unit” \( j \) with \( K_U \) unit-level predictors \( u_j \) is given by

\[
y_i = \mu + u_j(i) \gamma + x_i \beta + \psi_j(i) + \epsilon_i
\]

where \( \epsilon_i \) is a mean zero error term and there is a separate mean \( \psi_j \) for each unit, with the index \( j(i) \) indicating the unit \( j \) to which person \( i \) is linked. The model for the outcomes from a typical sample of data from \( N \) persons is given by:

\[
Y = 1 \mu + U \gamma + X \beta + F \psi + \epsilon
\]

where \( 1 \) is a conforming vectors of ones, \( F \) is an \((N \times J)\) incidence matrix consisting of only zeros and ones with a single one in each row (i.e. each observation is linked to exactly one of the \( J \) units) so that \( F1 = 1 \), and \( U \) and \( X \) are \( N \times K_U \) and \( N \times K_P \) matrices containing covariates. The model can also be fit to panel data with multiple observations over time on individual persons; see Mihaly et al. (2010) for an example of such a model. In the next section we explicitly consider such a model with the addition of person-level fixed effects. Without the inclusion of person-level fixed effects, the time component has no specific implication for our results and we ignore it in the presentation in this section.

Least squares is the standard approach for estimating the model parameters. Least squares estimates are solutions to the normal equations Searle (1971):

\[
\begin{bmatrix}
1' (1 \mu + U \gamma + X \beta + F \psi) \\
U' (1 \mu + U \gamma + X \beta + F \psi) \\
X' (1 \mu + U \gamma + X \beta + F \psi) \\
F' (1 \mu + U \gamma + X \beta + F \psi)
\end{bmatrix}
= 
\begin{bmatrix}
1' Y \\
U' Y \\
X' Y \\
F' Y
\end{bmatrix}
\]

For models with one level of fixed effects, solutions to the normal equations for the fixed effects are simple closed form expressions of the unit and overall sample averages or adjusted averages (the raw unit or sample average less the coefficient-weighted sum of the corresponding covariate average). Consequently, the estimates from the various routines are functions of the unit and sample means or adjusted means.

Because \( F1 = 1 \), if \( \tilde{\mu}, \tilde{\gamma}, \tilde{\beta} \) and \( \tilde{\psi} \) are a solution to Equation 3 then so is \( \tilde{\mu} + 1c \), and \( \tilde{\psi} - 1c \) for any constant \( c \). In other words, the design matrix \([1, X, F]\) is less than full column rank and solutions to the normal equations are not unique. Moreover, additional degrees of freedom are lost if \( U \) is included.

As noted above, to accommodate the indeterminacy of the model, the model is typically reparameterized to a model that yields normal equations with a unique solution. For instance, \( u_j \gamma \) is combined with \( \psi_j \), or \( \psi_j \) is replaced with the contrast \( \psi_j - \psi_j \) and
\( \mu \) is replaced with \( \mu + \psi_J \). The resulting solutions to the normal equations for the various reparameterized models correspond to alternative solutions to normal equations. The alternative model parameterizations for the fixed effects and unit-level variables produce the same overall model fit and residuals because the predicted value for each observation is invariant to the alternative parameterizations of fixed effects. Inferences about the elements of \( \beta \) are likewise invariant. However, the alternative parameterizations yield different estimands for the fixed effects and corresponding different estimates and standard errors.

### 2.2 Stata Routines for One-Level Fixed Effects Model

There are a number of routines available in Stata to estimate Model 2 and to produce estimates of the unit effects, including `reg`, `areg`, `xtreg`, `fese` (Nichols (2008)) and `felsdvregdm` (Mihaly et al. (2010)) \(^1\). These routines use different approaches to the overparameterization of the model and select different solutions to the normal equations. Tables 1 to 3 characterize the estimates that these routines provide for the overall constant, the unit effects and the standard errors of the estimated unit effects \(^2\). The three tables refer to three canonical instances of the model: Table 1 considers the simplest case where there are no covariates, only an intercept; Table 2 considers a more complex case that includes only \(X\), person-level variables, along with the intercept; and Table 3 considers a case with person level variables and includes a classifying unit-level variable \(u_j\) (e.g., large versus small firms or grade 4 versus grade 5 teachers).

None of the routines is designed for estimating models with continuous unit-level effects. We consider discrete unit-level variable. The units are partitioned into subsamples called “reference collections” defined by values of the unit-level variables: each unit belongs to exactly one reference collection and units within a reference collection have common values of the unit level-variables. For instance, in a sample of firms the unit-level variable may be an indicator for large firms, so that firms are partitioned into two reference collections one with large firms and one with small firms. Similarly, if experience is the unit-level variable for a sample of teachers then teacher could be partitioned into two reference collections: novice and experienced teachers.

Even in a very simple case with only unit fixed effects and an intercept (Table 1), the various estimation routines estimate different quantities and produce different estimates. When the model is expanded to include only person-level variables in addition to unit effects and the intercept (Table 2), the routines typically provide analogous results to those from the simpler model without person-level variables. However, rather than using raw unit or sample means, estimates now use means adjusted by the covariate means scaled by the estimated regression coefficients. When the model includes unit-level variables (Table 3), the differences among the estimation methods become more

---

1. Note that `felsdvregdm` requires longitudinal data while the others do not.
2. Estimands reported in the tables are based on running the routines on simulated data in Stata/MP 11 on 64-bit Linux machines and using versions of `a2reg`, `gpreg`, `fese`, `felsdvreg` and `felsdvregdm` downloaded from Statistical Software Components (SSC) or from the Stata Journal archive on September 14, 2010.
pronounced. In the remainder of this section we compare and contrast the behaviors of the routines under these different scenarios.

**reg**

With **reg**, the unit effects must be explicitly included in the model statement. With even a modest numbers of units, creating unit indicators and typing their names in the routine call is tedious and users typically avoid this by using the **xi** command to generate the indicators and add them to the model. This is the approach we consider for our comparisons. As shown in Table 1 in models without unit-level variables ($u_j$) and with the default behavior of **xi**, **reg** uses the traditional reparameterization of the model replacing $\psi_j$ with $\psi_j^* = \psi_j - \psi_1$, where $\psi_1$ is the unit with the first label in alphabetical order. Which unit gets held out can be manipulated with **xi**. It uses the analogous covariate-adjusted parameterization for the Table 2 model. However, for the Table 3 model, the estimates produced by **reg** do not follow any simple or obvious pattern. The estimates produced involve complicated contrasts of unit means within and between the levels of the unit variables and the exact estimates can depend on the specification of the model including how the variables are ordered or included in the procedure call.

**areg and xtreg**

**areg** and **xtreg** behave identically in all of the cases we consider. Both were designed for efficient computation in models with large numbers of fixed effects at one level under the assumption that fixed effects are included as nuisance parameters to control for differences among units that could bias the estimates of interest, the $\beta$ coefficients. Computational efficiency is achieved by absorbing the unit fixed effects via the “within” transformation in which the corresponding unit-level means are subtracted from each element of $Y$ and each element of every column of $X$, and the adjusted outcomes are regressed on the adjusted covariates indicators.

This method does not directly produce estimates of the unit effects. Rather, these estimates are recovered post-estimation using the **predict** command with either the d or xbd options in **areg**, and analogously, with either the u or xbu options in **xtreg**. Standard errors are not provided in any case. These options correspond to different parameterizations of the unit effects. For the Table 1 model, the d or u options provide estimates that equal deviations from the grand mean, and the xbd and xbu options return simple unit averages. These procedures also provide the grand mean as an estimate of the intercept. The combination of the grand mean and individual unit means as estimates for the vector of model parameters is not a solution to the normal equations for least squares estimation. These estimates cannot be combined to predict values or produce residuals. For the Table 2 model, the d or u options provide estimates equal to covariate-adjusted deviations from the grand mean, while the xbd and xbu options still provide simple unit averages. For the Table 3 model, both routines ignore the inclusion of the unit level variables when estimating the unit effects, and therefore
provide estimates identical to those from the Table 1 model.

**fese**

The user-developed command `fese` was created specifically to estimate fixed effects and their standard errors building on the `areg` procedure. Unlike most Stata procedures, `fese` does not estimate the intercept by default. To estimate the intercept, the user must explicitly include a constant variable equal to one in the procedure call. `fese` reparameterizes the model by replacing $\psi$ with $\psi^* = \psi + \mu + u_j\gamma$.

`fese` produces the same point estimates for the unit effects and intercepts as `areg` with `predict` and the `xb` option and `xtreg` with `predict` and the `xbu` option, for the Table 1 and 3 models. Unlike `areg` and `xtreg` it also produces standard errors. However, the standard error estimates incorrectly adjust for a degree of freedom given to the intercept and one for every unit mean even though the model is over parameterized, and the model degrees of freedom equal the number of units, not the number of units plus one. Consequently, the standard errors are biased upward, although by a trivial amount when there are many units. Like `areg` and `xtreg`, the `fese` estimates are identical for the Table 1 and Table 3 models because it ignores unit-level variables when estimating unit effects. The standard errors again incorrectly include all the covariates and the units in the calculation of model degrees of freedom, even though there are only $J$ independent model degrees of freedom, which again results in an upward bias in the estimated standard errors. For the Table 2 model it estimates unit means adjusted for the unit mean of the covariate.

**felsdvregdm**

The user-developed `felsdvregdm` was designed for estimating two-level models but will work when the model does not include person effects provided there are repeated measures on persons. Mihaly and colleagues developed `felsdvregdm` to contrast each unit to the average unit and to provide standard errors for these estimated contrasts. In simple cases with no unit-level variables, it reparameterizes the model by replacing $\psi$ with $\psi^* = \psi - \bar{\psi}$, where $\bar{\psi}$ is the average of all the unit-specific means. When there are unit-level variables, `felsdvregdm` estimates unit effects as the difference between the unit mean and the average of the unit means for all units in the unit’s reference collection. Let $\bar{\psi}_g$ equal the average of the unit-specific means for units in reference collection $g = 1, \ldots, G$, then `felsdvregdm` reparameterizes the model with $\psi^* = \psi - \bar{\psi}_g$ for each unit is reference collection $g$. These are reflected in the estimators reported in Tables 1 to 3. As shown in the tables, `felsdvregdm` also provides standard errors for all estimated parameters.
Table 1: Description of estimates from various Stata procedures for the constant and unit means and standard errors of unit means for a one-level model with only unit means and an intercept. The mean outcome for unit $j = 1, \ldots, J$ is $\bar{y}_j$; the average of the unit means is $\bar{y}$; and the mean of the individual values $\tilde{y}$.

<table>
<thead>
<tr>
<th>Stata Procedure</th>
<th>Unit Effect</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>reg: with xi command</td>
<td>$\bar{y}_j - \bar{y}_1$</td>
<td>OLS Std. Error for Contrast</td>
</tr>
<tr>
<td>areg: with predict d</td>
<td>$\bar{y}_j - \bar{y}$</td>
<td>Not</td>
</tr>
<tr>
<td>areg: with predict xbd</td>
<td>$\bar{y}_j$</td>
<td>Provided</td>
</tr>
<tr>
<td>xtreg: with predict u</td>
<td>$\bar{y}_j - \bar{y}$</td>
<td>Not</td>
</tr>
<tr>
<td>xtreg: with predict xbu</td>
<td>$\bar{y}_j$</td>
<td>Provided</td>
</tr>
<tr>
<td>fese (must explicitly include constant)</td>
<td>$\bar{y}_j$</td>
<td>$\sqrt{(N - J)/(N - J - 1)}$</td>
</tr>
<tr>
<td>felsdvregdm</td>
<td>$\bar{y}_j - \bar{y}$</td>
<td>OLS Std. Error for Effect</td>
</tr>
</tbody>
</table>
Table 2: Description of estimates from various Stata procedures for the constant and unit means and standard errors of unit means for a one-level model with unit means, an intercept, and person-level predictors. The mean outcome for unit $j = 1, \ldots, J$ is $\bar{y}_j$; the average of the unit means is $\bar{y}$; and the mean of the individual values $\tilde{y}$. The unit means for the vector of person-level predictors is $\bar{x}_j$, the average of the unit means is $\bar{x}$, and the mean of the individual values is $\tilde{x}$.

<table>
<thead>
<tr>
<th>Stata Procedure</th>
<th>Unit Effect</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>reg</strong> with xi command</td>
<td>$\hat{\beta} - \bar{x}_j' \hat{\beta} - (\bar{y}_1 - \bar{x}_1' \hat{\beta})$</td>
<td>OLS Std. Error for Contrast</td>
</tr>
<tr>
<td><strong>areg</strong> with predict d</td>
<td>$\hat{\beta} - \bar{x}_j' \hat{\beta} - (\bar{y} - \bar{x}' \hat{\beta})$</td>
<td>Not</td>
</tr>
<tr>
<td><strong>areg</strong> with predict xbd</td>
<td>$\hat{\beta} - \bar{x}_j' \hat{\beta}$</td>
<td>Provided</td>
</tr>
<tr>
<td><strong>xtreg</strong> with predict u</td>
<td>$\hat{\beta} - \bar{x}_j' \hat{\beta} - (\bar{y} - \bar{x}' \hat{\beta})$</td>
<td>Not</td>
</tr>
<tr>
<td><strong>xtreg</strong> with predict xbu</td>
<td>$\hat{\beta} - \bar{x}_j' \hat{\beta}$</td>
<td>Provided</td>
</tr>
<tr>
<td><strong>fese</strong> (must explicitly include constant)</td>
<td>$\tilde{y}_j - \bar{x}_j' \hat{\beta}$</td>
<td>$\sqrt{(N - J)/(N - J - 1)}$</td>
</tr>
<tr>
<td><strong>felsdvregdm</strong></td>
<td>$\tilde{y}_j - \bar{x}_j' \hat{\beta}$</td>
<td>OLS Std. Error for Effect</td>
</tr>
</tbody>
</table>
Table 3: Description of estimates from various Stata procedures for the constant and unit means and standard errors of unit means for a one-level model with unit means and reference collection means. The mean outcome for unit $j_g = 1, \ldots, J_g$ of reference collection $g = 1, \ldots, G$ is $\bar{y}_{gj}$; the average of the unit means for reference collection $g$ is $\bar{y}_g$; and the mean of the individual values $\bar{y}$. For all procedures unit means are estimated for all reference collections.

<table>
<thead>
<tr>
<th>Stata Procedure</th>
<th>Coefficient for Unit Effect</th>
<th>Unit Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>reg</strong>: with xi command</td>
<td>No set pattern arbitrary units and group means dropped</td>
<td>OLS Std. Error for Estimates</td>
</tr>
<tr>
<td>areg: with predict d</td>
<td>$\tilde{y}$. Estimated $j = 1, \ldots, J$</td>
<td>Not Provided</td>
</tr>
<tr>
<td>areg: with predict xbd</td>
<td>$\tilde{y}$. Estimated $j = 1, \ldots, J$</td>
<td>Not Provided</td>
</tr>
<tr>
<td>xtreg: with predict u</td>
<td>$\tilde{y}$. Estimated $j = 1, \ldots, J$</td>
<td>Not Provided</td>
</tr>
<tr>
<td>xtreg: with predict xbu</td>
<td>$\tilde{y}$. Estimated $j = 1, \ldots, J$</td>
<td>Not Provided</td>
</tr>
<tr>
<td><strong>fese</strong> (must explicitly include constant)</td>
<td>$\tilde{y}$. Estimated $j = 1, \ldots, J$</td>
<td>$\sqrt{(N - J)/(N - J - G)}$</td>
</tr>
<tr>
<td><strong>felsdvregdm</strong></td>
<td>Not Provided $g = 1, \ldots, G$</td>
<td>OLS Std. Error for Effect</td>
</tr>
</tbody>
</table>
3 Two Levels of Fixed Effects

3.1 Two-Level Fixed Effects Model

When there are repeated measures on the persons (e.g., workers, patients or students), person-level fixed effects can be included in the model allowing for more flexibility in modeling potential differences among the persons associated with the different units. Models with both person and unit level fixed effects expand the basic Model 1 by accounting for potential differences among the persons associated with the different units. The existence of multiple groups will change the specific form of the estimates but the general patterns will not change.

The model for the outcomes from a typical sample of data from $N^\ast$ observations from $N$ persons is given by:

$$Y = 1\mu + U\gamma + V\eta + X\beta + Z\delta + F\psi + D\theta + \epsilon$$  \hspace{1cm} (5)

where $1$, $U$, $X$ and $F$ are defined as they were with Model 1, and $V$ and $Z$ are similarly defined matrices of the time-varying unit and person variables. $\theta$ is the vector of person fixed effects $D$ is a $(N^\ast \times N)$ matrix of indicator variables for persons and $\theta$ is a vector of person-level fixed effects. The normal equations for this model are:

$$
\begin{pmatrix}
1' (1\mu + U\gamma + V\eta + X\beta + Z\delta + F\psi + D\theta) \\
U' (1\mu + U\gamma + V\eta + X\beta + Z\delta + F\psi + D\theta) \\
V' (1\mu + U\gamma + V\eta + X\beta + Z\delta + F\psi + D\theta) \\
X' (1\mu + U\gamma + V\eta + X\beta + Z\delta + F\psi + D\theta) \\
V' (1\mu + U\gamma + V\eta + X\beta + Z\delta + F\psi + D\theta) \\
F' (1\mu + U\gamma + V\eta + X\beta + Z\delta + F\psi + D\theta) \\
D' (1\mu + U\gamma + V\eta + X\beta + Z\delta + F\psi + D\theta)
\end{pmatrix} = 
\begin{pmatrix}
1'Y \\
UY \\
VY \\
XY \\
YZ \\
FY \\
DY
\end{pmatrix}
$$  \hspace{1cm} (6)

3. We assume that there is no stratification or grouping of variables (Mihaly et al., 2010, Cornelissen, 2008). The existence of multiple groups will change the specific form of the estimates but the general patterns will not change.
We let $\tilde{\mu}$, $\tilde{\gamma}$, $\tilde{\eta}$, $\tilde{\beta}$, $\tilde{\delta}$, $\tilde{\psi}$, and $\tilde{\theta}$ be an arbitrary solution to the normal equations (Equation 6). Because the sum of the columns of $F$ and the sum of the columns of $D$ equal 1, we can add a constant to $\tilde{\mu}$ and subtract it from every unit effect or every person effect to generate an alternative solution to the normal equations. Similarly, we can add a constant to every unit effect and subtract it from every person effect and again generate an alternative solution to the normal equations.

For the simple case with just an intercept and fixed effects for units and persons, one solution to the normal equations can be obtained by setting the person effects equal to the person means, using the within transformation to remove person-level means from $Y$ and the unit indicators (i.e., the columns of $F$) and regressing the adjusted outcomes on the adjusted unit indicators. The adjusted unit indicators are linearly dependent. The regression estimates can be obtained by dropping one of the adjusted unit indicators and regressing the outcomes on those that remain. This solution via the within transformation is equivalent to regressing the outcomes on the person indicators and obtaining the residuals and regressing the unit indicators on the person indicators and obtaining the residuals and then regressing the outcome residuals on the unit indicator residuals using a G-inverse. Other methods are available for solving the normal equations (e.g., a2reg or gpreg), but this simple method yields an intuitive solution that is commonly used.

For models with $V$, $X$, and $Z$ but no time-invariant unit-level covariates $U$, an example of an arbitrary solution to the normal equations can be found by regressing the outcomes on the person effect indicators and the time-varying variables and obtaining the residuals, regressing the unit indicators on the person effect indicators and the time-varying variables, and then regressing the residuals on the residuals using a G-inverse (e.g., dropping one of the adjusted unit indicators). For models with time-invariant unit-level covariates, arbitrary solutions to the normal equations can be found by again first regressing outcomes, unit indicators and time-invariant unit-level covariates on all other variables and then regressing the residuals of the outcomes on the residuals for the other variables using any G-inverse to find a solutions. For instance, dropping one unit indicator from every reference collection would provide a solution.

### 3.2 Stata Routines for Two-Level Fixed Effects Model

We consider six routines for estimating unit effects in the two-level fixed model: areg, xtreg, and the user developed felsdevreg, feldsvregdm, a2reg, and gpreg. As discussed above, areg and xtreg were developed to improve computational efficiency for models with a large number of fixed at one level but they can be slow for models with large number of fixed effects at two levels such as many teachers and many students. felsdevreg, feldsvregdm, a2reg, and gpreg were all developed to overcome the computational limitations of areg and xtreg but take different approaches to achieve com-

---

4. reg could again be used to estimate the parameters of models with two levels of fixed effects but both the unit and person-level fixed effects would need to be entered explicitly resulting in computationally inefficient estimation and precluding use except in cases with small numbers of units and persons. Hence, we do not discuss the use of reg for parameter estimation with two-level models.
putational efficiency.

As with the one-level fixed effects model, we again consider three scenarios: a simple model with just an intercept and unit and person fixed effects; a model with an intercept, unit and person fixed effects, and time-varying unit and person variables \((v_{jt} \text{ and } z_{it})\); and a model with an intercept, unit and person fixed effects, and time invariant unit variables \((u_j)\). The solutions to the normal equations are not simple closed form expressions of unit averages or adjusted unit averages; however, for most cases the estimates are simple expressions of the arbitrary solution to the normal equations and we present results in terms of those solutions for individual units and the averages across units.

For models that include time-varying variables, the estimates follow the same patterns as estimates for models with only fixed effects although the arbitrary solutions to the normal equations now adjust for the time-varying variable and the person fixed effects. Hence, we do not include a table for this model.

**areg and xtreg**

Using `areg` or `xtreg` to estimate unit fixed effects in the two-level fixed effects model is analogous to estimating the unit effects with `reg` in the one-level fixed effects model. The person level effects are absorbed by the within transformation but the unit effects, which are of interest, are explicitly included in the model, typically with the `xi` option. Consequently, the parameterization of the unit effects from `areg` and `xtreg` in the two-level model is the same as the parameterization of the unit effects from `reg` in the one-level model.

As shown in Table 4 in a simple model without time-varying or time-invariant covariates, the unit effects are estimated as contrasts between solutions to the normal equations for the unit and the unit with the smallest ID. The estimands are the contrasts between the \(\psi\) for each unit and the holdout. The procedures provide standard errors for the estimated contrasts.

As shown in Table 5 like `reg` in the one-level fixed effects model, `areg` and `xtreg` rely on Stata defaults for estimation with collinear predictors. As a result, in models with unit effects and time-invariant unit-level predictors, both routines provide estimates that drop an arbitrary set of columns resulting in unit level effects that may include contrasts of units with the same or different values of the time-invariant unit-level predictors. The exact set of columns dropped can be sensitive to the data set and reordering of unit identification numbers. The estimates will in general be very difficult to interpret as any specific function of the \(\psi\) parameters.

**felsdvreg**

For the model with just an intercept and fixed effects but no covariates, `felsdvreg` also provides the same estimates of the unit effects as `areg` and `xtreg` – the contrasts between each unit and the unit with the lowest identifier as estimates of the corresponding
contrasts $\psi^* = \psi - \psi_1$. It also provides standard error estimates consistent with this parameterization. To fit this model `felsdvreg` requires that a variable equal to 1 for every observation be specified as a predictor for the model, but it is omitted from the estimated model because it is collinear with the person effects. However, `felsdvreg` does provide the overall sample mean as an estimate of $\mu$. The procedure does not add one degree of freedom to the model degrees of freedom for the intercept because it is already in the column space spanned by the person effects. Formally the estimates are a solution to the normal equations because the estimate of the intercept can be subtracted from the person effects.

`felsdvreg` also uses the same model parameterization and provides the same estimates and standard errors as `areg` and `xtreg` for models with only time-varying unit-level or person-level covariates. For all models without time-invariant unit-level predictors, the difference between `felsdvreg` and `areg` or `xtreg` is the computational efficiency of `felsdvreg` and its ability to model much larger data sets than the other two routines.

For models that include time-invariant unit level predictors ($U$ in Model 5), `felsdvreg` deviates from `areg` and `xtreg` by using a consistent and well-defined model parameterization that follows the convention of combining the time-invariant unit covariates with the unit means and estimates unit effects as contrasts of these combined quantities. In particular, `felsdvreg` reparameterizes the unit effects to $\psi^* = \psi + u' \gamma - \psi_1 - u_1' \gamma$. `felsdvreg` ignores the time-invariant unit-level covariates when they are included in the model. It will produce the same estimates whether or not users include time-invariant unit-level covariates in the model statement when calling the routine.

`felsdvregdm` uses the same computational algorithm as `felsdvreg` to estimate the parameters of models with two-levels of fixed effects, so it provides efficient estimation for large data sets. However, it uses different model parameterization than `felsdvreg` for models with and without time-invariant unit-level covariates and consequently provides different estimates of the unit effects.

`felsdvregdm` uses the same reparameterization for the two-level fixed effects model that is uses for the one-level fixed effects model. That is it replaces $\psi$ with $\psi^* = \psi - \bar{\psi}_r$, where $\bar{\psi}_r$ is the average of the unit means. For models without time-invariant unit-level covariates the means is over all units. For models with time-invariant unit-level covariates the means is over all units with the same value of the time-invariant unit level covariate. As shown in Tables 4 and 5, the estimates equal corresponding contrasts of arbitrary solutions of the normal equations. As with the one-level model, `felsdvregdm` provides standard errors for all estimates.

---

5. Whether or not the model includes time-invariant unit-level covariates, the unit effect estimates from `felsdvreg` are solutions to the same normal equation, but formulas for the estimates are different in Tables 4 and 5. This is because when the model includes time-invariant unit-level covariate they are part of the estimate even if they are not used in estimation. When the models does not include these variables they are not part of the estimates.
For models without time-invariant unit-level covariates, \texttt{a2reg} uses an analogous model parameterization to the one used by \texttt{areg}, \texttt{xtreg}, and \texttt{felsdvreg} except the holdout unit is the unit with the highest value of the ID variables. \texttt{a2reg} uses a very fast computational method that does not yield standard errors. It requires that a variable equal to one for every observation be explicitly included in the model to estimate a model with no covariates. However, the computational method does not involve matrix inversion and can find solutions even if there is collinearity among the predictors and the unit fixed effects and it does not recognize that this extra variable is collinear with the intercept which \texttt{a2reg} also includes in the model. Consequently, the routine estimates one extra model degree of freedom because it counts the redundant variables as separate variables in the model and F-tests produced by the routine use the wrong reference distribution.

\texttt{a2reg} also makes no special adjustments for the lack of identification created by modeling with time-invariant unit-level predictors and unit fixed effects. Because the computational algorithm can find solutions even if there is collinearity among the predictors and the unit fixed effects, the routine treats time-invariant unit-levels covariates like any other covariates. Solutions are found ignoring the redundancy between the covariates and the unit effect indicators. The resulting estimates will be solutions to the normal equations; however, the estimates can differ from the estimates produced by the other procedures by arbitrary values that do not relate to any sample statistics. Even in our relatively simple example cases, we could not map the estimates to any straightforward combination of the model parameters. The \texttt{a2reg} estimates of unit effects in this context do not appear to provide estimates of any parameters of interest.

\texttt{gpreg}

Like \texttt{a2reg}, \texttt{gpreg} also uses an iterative procedure to solve the normal equations and find least squares estimates. Unlike most of the other procedures, \texttt{gpreg} directly solves the normal equations \cite{McCaffrey2004} without any explicit reparameterization of the model. With some data sets that we examined, for models without time-invariant unit-level covariates, the estimates produced by \texttt{gpreg} equalled those produced by \texttt{felsdvregdm} and so the estimands would match the reparameterization used by that procedure. With other data sets, the estimates produced by \texttt{gpreg} differed from those produced by \texttt{felsdvregdm} by a small constant (denoted $C$ in Table \ref{table:results}). We could not map this constant onto any straightforward combination of unit-level quantities and hence we could not map the estimand of \texttt{gpreg} to a well-defined straightforward combination of the $\psi$.

For models with time-invariant unit-level covariates, \texttt{gpreg} explicitly folds these variables into the unit effect by definition. Consequently the routine produces the same estimates for models with and without these covariates. Again, the estimates do not have a simple interpretation as contrast of the $\psi$. 
Fixed Effects Estimation in Stata

Table 4: Description of estimates from various Stata procedures unit means and standard errors of unit means for a two-level model with only unit means, person means, and an intercept. The $\tilde{\mu} = 0$, $\tilde{\psi}_j$, $j = 1, \ldots, J$, and $\tilde{\theta}_k$, $k = 1, \ldots, K$ are an arbitrary solution to the normal equations for the OLS estimates of these parameters and $\bar{\psi}$ is the average of the $\tilde{\psi}_j$. For \texttt{gpreg} $C$ is an unspecified constant that is not a simple function of unit-level quantities.

<table>
<thead>
<tr>
<th>Stata Procedure</th>
<th>Unit Effect</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>\texttt{areg}</td>
<td>$\psi_j - \psi_1$ \texttt{\textbackslash j = 2, \ldots, J}$</td>
<td>OLS Std. Error for contrast</td>
</tr>
<tr>
<td>\texttt{xtreg}</td>
<td>$\bar{\psi}_j - \bar{\psi}_1$ \texttt{\textbackslash j = 2, \ldots, J}$</td>
<td>OLS Std. Error for contrast</td>
</tr>
<tr>
<td>\texttt{felsdreg}</td>
<td>$\bar{\psi}_j - \bar{\psi}_1$ \texttt{\textbackslash j = 2, \ldots, J}$</td>
<td>OLS Std. Error for contrast</td>
</tr>
<tr>
<td>\texttt{felsdregdm}</td>
<td>$\bar{\psi}_j - \bar{\psi}$ \texttt{\textbackslash j = 1, \ldots, J}$</td>
<td>OLS Std. Error for Effect</td>
</tr>
<tr>
<td>\texttt{a2reg}</td>
<td>$\bar{\psi}_j - \bar{\psi}_j$ \texttt{\textbackslash j = 1, \ldots, J - 1}$</td>
<td>Not Provided</td>
</tr>
<tr>
<td>\texttt{gpreg}</td>
<td>$\bar{\psi}_j - C$ \texttt{\textbackslash j = 1, \ldots, J}$</td>
<td>Not Provided</td>
</tr>
</tbody>
</table>
Table 5: Description of estimates from various Stata procedures for the constant and unit means and standard errors of unit means for a two-level model with unit means and reference collection means. The mean outcome for unit $j_g = 1,\ldots,J$ of reference collection $g = 1,\ldots,G$ is $\bar{y}_{jg}$; the average of the unit means for reference collection $g$ is $\bar{y}_g$; and the mean of the individual values $\bar{y}$. For all procedures unit means are estimated for all reference collections. For \texttt{gpreg} $C$ is an unspecified constant that is not simple function of unit-level quantities.

<table>
<thead>
<tr>
<th>Stata Procedure</th>
<th>Coefficient for Effect</th>
<th>Unit Std. Effect</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>\texttt{areg}</td>
<td>No set pattern arbitrary units and group means dropped</td>
<td>OLS Std. Errors for Estimate</td>
<td></td>
</tr>
<tr>
<td>\texttt{xtreg}</td>
<td>No set pattern arbitrary units and group means dropped</td>
<td>OLS Std. Errors for Estimate</td>
<td></td>
</tr>
<tr>
<td>\texttt{felsdvreg}</td>
<td>$\psi_j + u_j'\tilde{\gamma} - \psi_1 - u_1'\tilde{\gamma}$ provided $j = 2,\ldots,J$</td>
<td>OLS Std. Error for contrast</td>
<td></td>
</tr>
<tr>
<td>\texttt{felsdvregd}</td>
<td>$\bar{\psi}_g - \bar{\psi}_G.$ provided $g = 1,\ldots,G - 1$</td>
<td>OLS Std. Error for Effect</td>
<td></td>
</tr>
<tr>
<td>\texttt{a2reg}</td>
<td>Does not correspond to a combination of unit values</td>
<td>Not Provided</td>
<td></td>
</tr>
<tr>
<td>\texttt{gpreg}</td>
<td>Not provided $\bar{\psi}_j + u_j'\tilde{\gamma} - C$</td>
<td>Not provided $j = 1,\ldots,J$</td>
<td></td>
</tr>
</tbody>
</table>
4 Computational Efficiency

In addition to the requisite estimates and parameterization, the choice of which Stata routine to use will also depend on the size of the sample, speed considerations and available computing resources. The available Stata routines use various computational algorithms to reduce the computational time required to estimate the parameters of interest. The various routines, however, are not equally efficient at computing estimates and some routines can take very long to produce estimates and standard errors. In some instances some routines cannot compute estimate within the available computing resources.

Although timing comparisons are sensitive to the particular computer architecture used in the comparison study, they provide useful information about the relative efficiency of alternative routines and some guidance about utility of alternative procedures for conducting users’ desired analyses. The timing data reported here are based on running Stata MP version 10 with 4 processors under Windows XP Professional x64 on a machine with Quad-Core Intel Xeon CPUs (X5365) running at 3.00 GHz and 32 GB of RAM.

For these comparisons we estimated teacher effects by grade for grade 4 to 8 mathematics teachers teaching Palm Beach County, Dade County or the eight largest counties in Florida during the XX to XX school years. All are large school systems and the data sets include from 99,397 records from 39,888 students to 774,156 records from 304,006 students and 2,587 teachers to 14,831 teacher-by-grade units.

For the case of one-level models, areg and xtreg are equivalent in their speed and use of resources, since both difference out the single fixed effect. Use of reg with the xi command is inferior in terms of speed and memory usage since it requires generating explicit indicator variables for the fixed effects. The use of reg is further constrained by the 11,000 variable limit inherent in Stata. Estimation with the reg command does have the advantage of producing estimated standard errors for the fixed effects. However, if one wants to obtain standard errors for the fixed effects in a one-level model, then felsdvregdm is clearly the superior choice. fese is impractically slow for even moderately large samples. reg is slower than felsdvregdm even in samples of modest size and does not scale up like felsdvregdm.

Computational efficiency is a greater concern for estimating two-level models and the choice of estimation routine is less clear cut. There are tradeoffs in terms of the parameters which are estimated, parameterization of the fixed effects, computational speed and memory requirements. Like in the one-level case, use of the xi command to generate explicit indicator variables is practical for only relatively small samples. If one is interested in point estimates alone, then a2reg would appear to be the clear choice, as it is much faster than any of the alternatives, it can be used with very large data sets, and does not require vast amounts of memory. However, as noted above, one must be careful to check for collinearities ex-ante, as a2reg can produce spurious results in such cases. While a2reg can be bootstrapped to produce estimates of the standard errors, the computational advantages of a2reg are lost since computational time is directly
proportional to the number of bootstraps. Even a 100-repetition bootstrap would make \texttt{a2reg} much slower than other alternatives.

When choosing between \texttt{gpreg} and \texttt{felsdvreg}, for small and moderately large samples (< 750,000 observations), \texttt{felsdvreg} is faster. However, as the sample size grows, the speed advantage of \texttt{felsdvreg} dissipates. Being an iterative-based routine, the time to estimate models with \texttt{gpreg} grows in direct proportion with sample size whereas estimation time for \texttt{felsdvreg} grows as approximately the cube of the number of units due to the matrix inversion. \texttt{felsdvreg} is probably not practical for samples larger than 1.5 million observations should we test this out? whereas Guimaraes and Portugal (2009) report estimating a two-level model on a sample of over 26 million observations with \texttt{gpreg}.

\texttt{felsdvreg}, \texttt{felsdvregdm}, and \texttt{gpreg} are written in Mata to avoid the 11,000 variable limit in Stata. This means, however, that one must have sufficient memory to hold the data in Stata and allow room for Mata to operate. DFM: this was Tim’s comment and I think he would need to provide these details since they are beyond my stata knowledge. [Maybe more here about the different algorithms available in \texttt{gpreg} and the multiples of the original data set size that are required for Stata with each program ? ? ?]

Since \texttt{felsdvregdm} is a modification to \texttt{felsdvreg}, it exhibits the same relationship of execution time to sample size and the same limitations on scalability as does \texttt{felsdvreg}. In general, \texttt{felsdvregdm} is slower than \texttt{felsdvreg}, but the difference is not huge and tends to diminish with sample size.

5 Conclusion

The recent increase in large administrative data for the estimation of firm, teacher or school effects has been matched by an increase in the development of Stata routines for estimating fixed effects in models with one-level of fixed effects (only firms or teachers) or two levels of fixed effects (firms and workers or teachers and students). These routines differ along three important dimensions: the estimands they estimate (e.g., the reparameterization of the model to account for the lack of identification of effects); whether or not they provide standard error estimates; and computational time.

When models do not include unit-level variables or time-invariant unit-level variables for longitudinal data, the estimands for the various procedures tend to fall into one of three types: 1) they equal a contrast between a unit and a holdout unit means; 2) they equal a contrast between the unit mean and the average of all the unit means; or 3) they equal the unit mean or the unit mean less the grand mean. Contrasts between unit and holdout unit means may or may not be of interest depending on whether the holdout unit is of interest. The unit means cannot be interpreted as causal effects. Contrasts of the means are causal effects but the standard errors may not be easy to recover.

When the models include unit-level variables or time-invariant unit-level variables for longitudinal data, the estimands of several procedures do not correspond to any parameters of interest and may be dependent on features of the data. This is true for
Table 6: Comparison of computational time to estimate fixed effects and analytic standard errors from various Stata procedures for models with one or two levels of fixed effects to student achievement data from Palm Beach, Dade or the eight largest counties in Florida. Times are give in hr:min:sec format. N/F is used to indicate that a procedure exceeded the available resources and failed to yield estimates.

<table>
<thead>
<tr>
<th></th>
<th>Palm Beach</th>
<th>Dade</th>
<th>8 Largest Counties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Student-Year Obs.</td>
<td>99,397</td>
<td>250,698</td>
<td>774,156</td>
</tr>
<tr>
<td>Number of Students</td>
<td>39,888</td>
<td>100,077</td>
<td>304,006</td>
</tr>
<tr>
<td>Number of Teacher-by-Grade Units</td>
<td>2,587</td>
<td>5,239</td>
<td>14,831</td>
</tr>
</tbody>
</table>

One-Level Models:

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Palm Beach</th>
<th>Dade</th>
<th>8 Largest Counties</th>
</tr>
</thead>
<tbody>
<tr>
<td>reg</td>
<td>0:03:17</td>
<td>0:20:10</td>
<td>N/F</td>
</tr>
<tr>
<td>areg</td>
<td>0:00:09</td>
<td>0:00:24</td>
<td>0:01:18</td>
</tr>
<tr>
<td>xtreg</td>
<td>0:00:12</td>
<td>0:00:22</td>
<td>0:00:53</td>
</tr>
<tr>
<td>fese</td>
<td>9:49:01</td>
<td>N/F</td>
<td>N/F</td>
</tr>
<tr>
<td>felsdvregdm</td>
<td>0:01:35</td>
<td>0:12:53</td>
<td>3:51:48</td>
</tr>
</tbody>
</table>

Two-Level Models:

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Palm Beach</th>
<th>Dade</th>
<th>8 Largest Counties</th>
</tr>
</thead>
<tbody>
<tr>
<td>areg</td>
<td>1:00:16</td>
<td>9:25:22</td>
<td>N/F</td>
</tr>
<tr>
<td>xtreg</td>
<td>1:04:37</td>
<td>10:26:45</td>
<td>N/F</td>
</tr>
<tr>
<td>felsdvregdm</td>
<td>0:02:03</td>
<td>0:13:26</td>
<td>3:48:54</td>
</tr>
<tr>
<td>felsdvreg</td>
<td>0:01:19</td>
<td>0:08:09</td>
<td>2:59:02</td>
</tr>
<tr>
<td>gpreg</td>
<td>0:21:44</td>
<td>0:59:01</td>
<td>2:36:54</td>
</tr>
<tr>
<td>a2reg</td>
<td>0:00:51</td>
<td>0:01:58</td>
<td>0:04:11</td>
</tr>
</tbody>
</table>
**McCaffrey et al.**

`reg` with the one-level fixed effects model, and `areg`, `xtreg`, and `a2reg` with the two-level fixed effects model. These procedures should be avoided in these contexts since the interpretation of the estimates is difficult and could lead to erroneous inferences about the effects of the units. Other procedures fold these unit-level variables into the unit effects. In many circumstances this may be an appropriate choice. Analysts may be interested in the total differences without any concern about the source of those differences. In other cases, it may be more appropriate to consider differences among units that share common values of the unit-level variables. `felsdvregdm` offers this alternative parameterization of effects.

Computing standard errors of the fixed effects involves the inversion of a matrix or other computationally demanding calculations to obtain the diagonals of the inverse of a matrix. The matrix is roughly $J \times J$ where $J$ is the number of units. When $J$ is large estimating the standard errors is very computationally demanding. For this reason many of procedures do not provide standard error estimates for the unit fixed effects. In additional routines that provide the standard errors typically require more computational time than procedures that do not. As demonstrated in the table, several procedures that do and do not provide standard errors provide the same estimates of the unit effects. Analysts might want to use procedures that do not provide standard errors for exploratory work and rely on the more computationally demanding procedures only when they have finalized their models.

The computational time required by the various procedures differs tremendously. However, in most cases the additional time required to produce the estimates appears to be a cost required for estimating standard errors. Bootstrap iterative resampling methods provide an alternative to estimating standard errors that can avoid inversion of a very large matrix, but the savings in time from not inverting the matrix appears to be offset by the time required for the iterative calls to the procedure.

Computational requirements of the procedures grows faster than linearly with number of units, and for very large problems, such as estimating effects for all the teachers in a large state, procedures such as `felsdvreg` and `felsdvregdm` will exceed computational resources. In these cases, procedures such as `a2reg` might be the only approach to estimation but analysts will need to take care to remove unit-level (or unit-level time-invariant) variables prior to using the procedure.

The wide array of procedures for estimated fixed effects provides users with flexibility in choosing the estimand to be estimated and programs that scale differently to problems of different sizes. No one procedure covers the full space of providing all the various estimands with a procedure that scales to even the largest of problems. However, by carefully considering the estimand, the standard errors provided, and the computational efficiency of each procedure, analysts are likely to find a method of estimating the values of interest to them with one of the tools provided by Stata.
6 References


Nichols, A. 2008. FESE: Stata module to calculate standard errors for fixed effects. Statistical Software Components S456914, Boston College Department of Economics.

Ouazad, A. 2008. A2REG: Stata module to estimate models with two fixed effects. Statistical Software Components S456942, Boston College Department of Economics.

Schmieder, J. 2009. GPREG: Stata module to estimate regressions with two dimensional fixed effects. Statistical Software Components S457048, Boston College Department of Economics.