The Stata Journal (2012)
12, Number 3, pp. 406–432

A review of Stata commands 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
Georgia State University
Atlanta, GA
tsass@gsu.edu

Abstract. Availability of large multilevel longitudinal databases in various fields of research, including labor economics (with workers and firms observed over time) and education (with students, teachers, and schools observed over time), has increased the application of models with one level or multiple levels of fixed effects (for example, teacher and student effects). There has been a corresponding rapid development of Stata commands designed for fitting these types of models. The commands 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 commands. In this article, we catalog the estimates of reported fixed effects provided by different commands 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: st0267, longitudinal data, linked employer–employee data, fixed-effects estimators, regress, areg, a2reg, gpreg, reg2hdfe, xtreg, fese, felsdvregdm, software review

1 Introduction

In research pertaining to labor economics, health policy, and education, there is interest in estimating the effects of individual units (for example, firms, hospitals, doctors, schools, or teachers) from databases with measures on the units and individual persons (for example, workers, patients, or students) attached to them.[1] The feature common to these data is two types of entities connected to each other, where the connections may or may not result in nesting, and where the quantities of interest are the effects of one set of the entities, which we call "units". Commonly, the unit effects are estimated with fixed effects. Administrative databases with data from very large numbers of persons

1. Individuals within units do not need to be people; for instance, the effects of countries (units) and firms within countries may be of interest. Alternatively, units may be people, and "persons" may be repeated measurements on each person.

linked to hundreds or even thousands of units (for example, Abowd, Kramarz, and Roux [2006]; Harris and Sass [2011]) are increasingly available for such analyses. Models with huge numbers of fixed effects for units pose computational challenges, especially in cases in which the models also include fixed effects for persons. The Stata community has been active in developing commands for efficient estimation of such models, including `areg`, `xtreg`, and user-written commands such as `a2reg` (Ouazad 2008), `felsdvreg` (Cornelissen 2008), `felsdvregdm` (Mihaly et al. 2010), `fese` (Nichols 2008), `gpreg` (Schmieder 2009), and `reg2hdfe` (Guimarães and Portugal 2010).

Models with fixed effects for units are overparameterized because the means for the individual units 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 (for example, Bazzoli et al. [2008]; Goldhaber, Gross, and Player [2011]), there is interest in obtaining estimates of the fixed-effects parameters and often the standard-error estimates of the fixed-effects parameters. Stata commands 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 "holdout" 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 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.

In addition, analysts are increasingly interested in using post hoc "shrinkage" estimators (for example, Jacob and Lefgren [2008]) to reduce the error variance in estimates of units with small numbers of persons (for example, 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, while estimating unit effects, want to control for unit-level variables (for cross-sectional unit data) or for time-invariant unit-level variables (for longitudinal unit-level data). For example, in education, the units might be teacher effects by year, and the analyst might want to control for overall year means. One cannot separate the effects of the unit-level variables in cross-sectional data or time-invariant unit-level variables in longitudinal data from the differences in unit-level effects. For instance, an analyst could not determine if higher student achievement in a given year was due to all teachers performing better in that year or to the test being easier in that year; that is, there is no way to identify both year means and all the teacher-by-year effects. For linear models, the indeterminacy of the parameters is often referred to as a problem of collinearity because the unit-level or time-invariant variables are collinear with the indicator variables for the unit 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 for longitudinal data are removed from the models, conflating the effects of these factors with the individual unit effects; or 2) the individual unit effects are estimated only among units with the same value of the unit-level or time-invariant unit-level variables (for example, 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 (for example, `gpreg` or `reg2hdfe`). However, `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 use the conflated effects. Hence, analysts need to understand what has been estimated to accurately interpret the differences among units.

When analysts model longitudinal data with repeated measures on persons, the inclusion of fixed effects for persons, in addition to those for units, further complicates these issues. The Stata commands used for fitting models with a single level of fixed effects (that is, 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. Moreover, specialized commands developed specifically for "two-level" fixed-effects models, that is, models with fixed effects for units and persons, also use different parameterizations of the unit effects that do not necessarily match those of other commands.[2]

We have found that the subtle differences in the parameterization of the unit fixed effects across different commands have implications for how users can interpret and use the resulting estimates that are not described in the documentation for the commands.

2. At a single time point, persons are nested within units, so it is natural to refer to units and persons as levels in the hierarchical structure. Hence, we refer to models with only unit effects as one-level models and to models with unit and person effects as two-level models. In longitudinal data, persons can move across units from one measurement time to the next so that they are not nested within units. It may be more precise to refer to the models as one-way and two-way models because the data are not formally hierarchical; thus the units and persons are two factors with limited crossing. However, we refer to levels of effects because this language is intuitive, is somewhat common, and draws attention to the extra complexity of adding a second set of fixed effects to a model.

To help clarify these differences, we make some direct comparisons of estimates for the available Stata commands under scenarios designed to span the possible indeterminacies in modeling with fixed effects. We also highlight the differences in model parameterizations. Because some of the commands for estimating effects from models with two levels of fixed effects differ from those available for estimating effects from models with just one level of fixed effects and because the commands available for both behave differently in these two settings, we repeat our comparisons with these two alternative modeling conditions.

In the remainder of this article, 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.

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

$$y_i = \mu + u_{j(i)}^{'}\gamma + x_i^{'}\boldsymbol{\beta} + \psi_{j(i)} + \epsilon_i \tag{1}$$

where ϵ_i is a mean zero error term, and there is a separate mean ψ_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

$$\boldsymbol{Y} = \mathbf{1}\mu + \boldsymbol{U\gamma} + \boldsymbol{X\beta} + \boldsymbol{F\psi} + \boldsymbol{\epsilon} \tag{2}$$

where $\mathbf{1}$ is a conforming vector of ones, $\boldsymbol{F}$ is an $N \times J$ incidence matrix consisting of only 0s and 1s with a single 1 in each row (that is, each observation is linked to exactly one of the J units) so that $\boldsymbol{F}\mathbf{1} = \mathbf{1}$, and $\boldsymbol{U}$ and $\boldsymbol{X}$ are $N \times K_U$ and $N \times K_P$ matrices containing covariates. We focus only on cross-sectional models in this section because without the inclusion of person-level fixed effects, the time component has no specific implications for our results.

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

$$\begin{pmatrix} \mathbf{1}'\left(\mathbf{1}\mu + \boldsymbol{U\gamma} + \boldsymbol{X\beta} + \boldsymbol{F\psi}\right) \\ \boldsymbol{U}'\left(\mathbf{1}\mu + \boldsymbol{U\gamma} + \boldsymbol{X\beta} + \boldsymbol{F\psi}\right) \\ \boldsymbol{X}'\left(\mathbf{1}\mu + \boldsymbol{U\gamma} + \boldsymbol{X\beta} + \boldsymbol{F\psi}\right) \\ \boldsymbol{F}'\left(\mathbf{1}\mu + \boldsymbol{U\gamma} + \boldsymbol{X\beta} + \boldsymbol{F\psi}\right) \end{pmatrix} = \begin{pmatrix} \mathbf{1}'\boldsymbol{Y} \\ \boldsymbol{U}'\boldsymbol{Y} \\ \boldsymbol{X}'\boldsymbol{Y} \\ \boldsymbol{F}'\boldsymbol{Y} \end{pmatrix} \tag{3}$$

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 commands are functions of the unit and sample means or adjusted means.

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

As noted above, to accommodate this indeterminacy, the analyst reparameterizes the model to one that yields normal equations with a unique solution. For instance, $u_j^{'}\boldsymbol{\gamma}$ is combined with ψ_j, or ψ_j is replaced with the contrast $\psi_j - \psi_J$, and μ 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 (3). 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 $\boldsymbol{\beta}$ are likewise invariant. However, the alternative parameterizations yield different estimands for the fixed effects and corresponding different estimates and standard errors.

2.2 Stata commands for one-level fixed-effects model

Several commands are available in Stata to estimate (2) and produce estimates of the unit effects, including `regress`, `areg`, `xtreg`, `fese` (Nichols 2008), and `felsdvregdm` (Mihaly et al. 2010).[3] These commands use different approaches to the overparameterization of the model and select different solutions to the normal equations. Tables 1 to 3 (at the end of this section) characterize the estimates that these commands provide for the overall constant, the unit effects, and the standard errors of the estimated unit effects.[4] The three tables refer to three canonical instances of the model: table 1 considers the simplest case, which has no covariates, only an intercept μ, and in which coefficients $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are set to 0; table 2 considers a more complex case that adds only $\boldsymbol{X}$ person-level variables along with the intercept and unit effects, and in which $\boldsymbol{\gamma}$ remains equal to 0; and table 3 considers a case with person-level variables, which includes a classifying unit-level variable u_j (for example, large versus small firms or grade 4 versus grade 5 teachers), and all parameters of the model, μ, $\boldsymbol{\psi}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\beta}$ are estimated.

3. Note that `felsdvregdm` requires longitudinal data while the others do not.
4. Estimands reported in the tables are based on running the commands on simulated data in Stata/MP 12 on 64-bit Linux machines and using versions of `a2reg`, `fese`, `gpreg`, and `reg2hdfe` downloaded from Statistical Software Components on 29 March 2012. `felsdvreg` and `felsdvregdm` were downloaded from the *Stata Journal* archive using, respectively, `net install st0143_2.pkg` and `net install st0185.pkg` on 29 March 2012.

None of the commands are designed for fitting models with continuous unit-level effects. We consider discrete unit-level variables. 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 teachers could be partitioned into two reference collections: novice and experienced teachers.

Even in a simple case with only unit fixed effects and an intercept (table 1), the various estimation commands estimate different quantities and produce different estimates. When the model is expanded to include person-level variables in addition to unit effects and the intercept (table 2), the commands typically provide results analogous 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 pronounced. In the remainder of this section, we compare and contrast the behaviors of the commands under these different scenarios.

regress

With `regress`, the unit effects must be explicitly included in the model statement. With even a small number of units, creating unit indicators and typing their names in the command is tedious; users typically avoid this by using the `xi` command to generate the indicators and add them to the model.[5] 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`, `regress` uses the traditional reparameterization of the model, replacing ψ_j with $\psi_j^* = \psi_j - \psi_1$, where ψ_1 is the unit with the first label in alphabetical order. The unit that 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 provided by `regress` 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 the cases we consider. Both were designed for efficient computation in models with many fixed effects at one level under the as-

5. Editor's note: With the introduction of factor-variable notation in Stata 11, users no longer need to use the `xi` command prefix to generate the indicator variables. See [U] **11.4.3 Factor variables** for more information.

sumption that fixed effects are included as nuisance parameters to control for differences among units that could bias the estimates of interest, the β coefficients. Computational efficiency is achieved by absorbing the unit fixed effects via the "within" transformation or the Frisch–Waugh–Lovell decomposition (Lovell 2008), in which the corresponding unit-level means are subtracted from each element of $\boldsymbol{Y}$ and each element of every column of $\boldsymbol{X}$, and the adjusted outcomes are regressed on the adjusted covariates and indicators.

This method does not directly produce estimates of the unit effects. Rather, these estimates are recovered postestimation by using the `predict` command with either the `d` or the `xbd` option in `areg` and, analogously, with either the `u` or the `xbu` option 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` option provides estimates that equal deviations from the grand mean, and the `xbd` or `xbu` option returns 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` option provides estimates equal to covariate-adjusted deviations from the grand mean, while the `xbd` or `xbu` option still provides simple unit averages. For the table 3 model, both commands 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-written command `fese` was created specifically to estimate fixed effects and their standard errors by 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 ψ with $\psi^* = \psi + \mu + u_j^{'}\boldsymbol{\gamma}$.

For the models in tables 1 and 3, `fese` produces the same point estimates for the unit effects and intercepts as `areg` with `predict` and the `xbd` option and as `xtreg` with `predict` and the `xbu` option. Unlike `areg` and `xtreg`, `fese` 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 equals 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-written `felsdvregdm` was designed for fitting two-level models but will work when the model does not include person effects, provided there are repeated measures on persons. Mihaly et al. (2010) wrote `felsdvregdm` to contrast each unit with 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 ψ with $\psi^* = \psi - \overline{\psi}$, where $\overline{\psi}$ 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 $\overline{\psi}_g$ equal the average of the unit-specific means for units in reference collection $g = 1, \dots, G$, and then `felsdvregdm` reparameterizes the model with $\psi^* = \psi - \overline{\psi}_g$ for each unit in 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 $j = 1, \dots, J$ is $\overline{y}_j$; the average of the unit means is $\overline{y}_.$; and the mean of the individual values is $\widetilde{y}_.$.

Stata procedure	Constant	Unit effect	Std. error
`regress`: with `xi` command	$\overline{y}_1$	$\overline{y}_j - \overline{y}_1$, $j = 2, \dots, J$	OLS std. error for contrast
`areg`: with `predict d`	$\widetilde{y}_.$	$\overline{y}_j - \widetilde{y}_.$, $j = 1, \dots, J$	Not provided
`areg`: with `predict xbd`	$\widetilde{y}_.$	$\overline{y}_j$, $j = 1, \dots, J$	Not provided
`xtreg`: with `predict u`	$\widetilde{y}_.$	$\overline{y}_j - \widetilde{y}_.$, $j = 1, \dots, J$	Not provided
`xtreg`: with `predict xbu`	$\widetilde{y}_.$	$\overline{y}_j$, $j = 1, \dots, J$	Not provided
`fese` (must explicitly include constant)	$\widetilde{y}_.$	$\overline{y}_j$, $j = 1, \dots, J$	OLS std. error for unit mean $\times \sqrt{(N-J)/(N-J-1)}$
`felsdvregdm`	$\overline{y}_.$	$\overline{y}_j - \overline{y}_.$, $j = 1, \dots, J$	OLS std. error for effect

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 $\overline{y}_j$; the average of the unit means is $\overline{y}_{.}$; and the mean of the individual values is $\widetilde{y}_{.}$. The unit means for the vector of person-level predictors is $\overline{x}_j$; the average of the unit means is $\overline{x}_{.}$; and the mean of the individual values is $\widetilde{x}_{.}$.

Stata procedure	Constant	Unit effect	Std. error
`regress`: with `xi` command	$\overline{y}_1 - \overline{x}_1'\widehat{\boldsymbol{\beta}}$	$(\overline{y}_j - \overline{x}_j'\widehat{\boldsymbol{\beta}}) - (\overline{y}_1 - \overline{x}_1'\widehat{\boldsymbol{\beta}})$ $j = 2, \ldots, J$	OLS std. error for contrast
`areg`: with `predict d`	$\widetilde{y}_{.} - \widetilde{x}_{.}'\widehat{\boldsymbol{\beta}}$	$(\overline{y}_j - \overline{x}_j'\widehat{\boldsymbol{\beta}}) - (\widetilde{y}_{.} - \widetilde{x}_{.}'\widehat{\boldsymbol{\beta}})$ $j = 1, \ldots, J$	Not provided
`areg`: with `predict xbd`	$\widetilde{y}_{.} - \widetilde{x}_{.}'\widehat{\boldsymbol{\beta}}$	$\overline{y}_j$ $j = 1, \ldots, J$	Not provided
`xtreg`: with `predict u`	$\widetilde{y}_{.} - \widetilde{x}_{.}'\widehat{\boldsymbol{\beta}}$	$(\overline{y}_j - \overline{x}_j'\widehat{\boldsymbol{\beta}}) - (\widetilde{y}_{.} - \widetilde{x}_{.}'\widehat{\boldsymbol{\beta}})$ $j = 1, \ldots, J$	Not provided
`xtreg`: with `predict xbu`	$\widetilde{y}_{.} - \widetilde{x}_{.}'\widehat{\boldsymbol{\beta}}$	$\overline{y}_j$ $j = 1, \ldots, J$	Not provided
`fese`	$\widetilde{y}_{.} - \widetilde{x}_{.}'\widehat{\boldsymbol{\beta}}$	$\overline{y}_j - \overline{x}_j'\widehat{\boldsymbol{\beta}}$ $j = 1, \ldots, J$	OLS std. error[a]
`felsdvregdm`	$\overline{y}_{.} - \overline{x}_{.}'\widehat{\boldsymbol{\beta}}$	$(\overline{y}_j - \overline{x}_j'\widehat{\boldsymbol{\beta}}) - (\overline{y}_{.} - \overline{x}_{.}'\widehat{\boldsymbol{\beta}})$ $j = 1, \ldots, J$	OLS std. error for effect

[a] `fese` does not require the explicit inclusion of a constant when there are other variables in the model and will give the OLS standard error provided that the variables included in the model are linearly independent.

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 $\overline{y}_{gj}$; the average of the unit means for reference collection g is $\overline{y}_{g.}$; and the mean of the individual values is $\widetilde{y}_{\cdot}$. For all procedures, unit means are estimated for all reference collections.

Stata procedure	Constant	Coefficient for ref. coll. mean	Unit effect	Std. error
`regress`: with `xi` command	No set pattern; arbitrary units and group means dropped			OLS std. error for estimates
`areg`: with `predict d`	$\widetilde{y}_{\cdot}$	Not estimated	$\overline{y}_{gj} - \widetilde{y}_{\cdot}$ $j = 1, \ldots, J$	Not provided
`areg`: with `predict xbd`	$\widetilde{y}_{\cdot}$	Not estimated	$\overline{y}_{gj}$ $j = 1, \ldots, J$	Not provided
`xtreg`: with `predict u`	$\widetilde{y}_{\cdot}$	Not estimated	$\overline{y}_{gj} - \widetilde{y}_{\cdot}$ $j = 1, \ldots, J$	Not provided
`xtreg`: with `predict xbu`	$\widetilde{y}_{\cdot}$	Not estimated	$\overline{y}_{gj}$ $j = 1, \ldots, J$	Not provided
`fese`	$\widetilde{y}_{\cdot}$	Not estimated	$\overline{y}_{gj}$ $j = 1, \ldots, J$	OLS std. error for unit mean $\times \sqrt{(N-J)/(N-J-G)}$
`felsdvregdm`	Not provided	$\overline{y}_{g.}$ $g = 1, \ldots, G$	$\overline{y}_{gj} - \overline{y}_{g.}$ $j = 1, \ldots, J_g$	OLS std. error for effect

2.3 Example code for commands for one-level fixed-effects model

Below is example code for using the commands for one-level fixed effects. Each command uses the simulated dataset `statafetest`, available through the *Stata Journal* archive. The data include student test scores from students taught by 12 teachers, four from each of three years. There are 40 students, each with one score from each year. Between 1 to 16 students link to each teacher. The data include student identifiers (`sid`), teacher identifiers (`tchid`), the year (`year`), test scores (`y`), a time-varying student-level covariate (`x`), indicator variables for each year (`year1` to `year3`), and indicator variables for each student (`csid01` to `csid40`).

The example includes code for using each command to fit a model with the unit-level predictor `year`. For most commands, the code to fit models without covariates is the same as the example code without the year indicator variables, and the code to fit models with the student-level covariate replaces the year indicator variables with `x`.

For `fese`, we present the code for the model with no covariates along with the code for the model with year indicators because `fese` cannot fit a model with no specified covariates and requires the inclusion of the `con` variable, which equals 1 for every observation in the data, to fit the no covariate model. For models with specified covariates, a constant is included by default, and the `con` variable is not required.

For `felsdvregdm`, we present code for fitting all three models because the `reff()` parameter differs between models with and without a discrete unit-level predictor. For models with a discrete unit-level predictor, `reff()` must specify reference collections corresponding to the level of the predictor because the mean of the units with a common value of the unit-level predictor cannot be estimated independently of the effect of the unit-level predictor. Details on the use of reference collections in `felsdvregdm` can be found in Mihaly et al. (2010).

Both `areg` and `xtreg` require using the postestimation command `predict` to capture the fixed-effects estimates. `felsdvregdm` and `fese` have options available to save the estimated fixed effects and corresponding standard errors. For `regress`, we use the `estimates` command to extract and store the model coefficients, including the estimated fixed effects and their standard errors. Refer to the help files for each command for additional details on its syntax and options.

Example code for one-level fixed-effects models:

```
. use statafetest
. * generate a constant 1 to use with some routines
. generate con = 1
. * regress
. xi: regress y year1 year2 year3 i.tchid
  (output omitted)
. estimates store reg1L
. * areg
. areg y year1 year2 year3, absorb(tchid)
  (output omitted)
. predict aregestd1L if e(sample), d
. predict aregestxbd1L if e(sample), xbd
. * xtreg
. xtreg y year1 year2 year3, i(tchid) fe
  (output omitted)
. predict xtregestu1L if e(sample), u
. predict xtregestxbu1L if e(sample), xbu
```

```
. * fese
. * Model without covariates -- requires explicit inclusion of con variable
. fese y con, absorb(tchid) s(fese1L1)
```

(*output omitted*)

```
. * Model with covariates -- does not require inclusion of con variable
. fese y year1 year2 year3, absorb(tchid) s(fese1L2)
```

(*output omitted*)

```
. * felsdvregdm
. * Model without covariates -- reff() specifies a single reference collection
. felsdvregdm y, ivar(sid) jvar(tchid) feff(felsest1L1) peff(peffhat1)
> reff(con) feffse(felsse1L1) mover(mover1) group(group) xb(xb1) res(res1)
> mnum(mnum1) pobs(pobs1) onelevel noisily
```

(*output omitted*)

```
. * Model with unit-level covariates -- reff() specifies reference collections
. * equal to values of the covariates
. felsdvregdm y year1 year2 year3, ivar(sid) jvar(tchid) feff(felsest1L2)
> peff(peffhat1) reff(year) feffse(felsse1L2) mover(mover1) group(group)
> xb(xb1) res(res1) mnum(mnum1) pobs(pobs1) onelevel noisily
```

(*output omitted*)

```
. * Model with time-varying person-level covariates -- reff() specifies
. * a single reference collection
. felsdvregdm y x, ivar(sid) jvar(tchid) feff(felsest1L3) peff(peffhat1)
> reff(con) feffse(felsse1L3) mover(mover1) group(group) xb(xb1) res(res1)
> mnum(mnum1) pobs(pobs1) onelevel noisily
```

(*output omitted*)

3 Two levels of fixed effects

3.1 Two-level fixed-effects model

When there are repeated measures on persons (for example, 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) to

$$y_{it} = \mu + u^{'}_{j(i,t)}\boldsymbol{\gamma} + v^{'}_{j(i,t)t}\boldsymbol{\eta} + x^{'}_{i}\boldsymbol{\beta} + z^{'}_{it}\boldsymbol{\delta} + \psi_{j(i,t)} + \theta_i + \epsilon_{it} \tag{4}$$

where the index $j(i,t)$ denotes the unit of person i during time t; u_j, as in (1), is a vector of time-invariant unit variables; v_{jt} are time-varying unit variables; x_i and z_{it} are the corresponding vectors of time-invariant and time-varying person covariates; θ_i is an effect for person i; and ϵ_{it} is the mean zero error term. N persons are in J units. Because there are one or more measures for each person, there are N^* observations in the sample.[6]

6. 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.

Again least squares is the standard approach for estimating the parameters of (4). As in the model with one level of fixed effects, estimates for the two-level model are found as solutions to the normal equations. However, unlike the solutions of section 2.2, the solutions for models with two levels of fixed effects cannot be expressed as simple functions of the unit means because of the controls for person effects. The model remains overparameterized with no unique solution to the normal equations. The various estimation procedures again use different reparameterizations to remove the indeterminacies, yielding different solutions and different estimates. However, the relationships among the estimates from some procedures are simple and described in the next section.

The model for the outcomes from a typical sample of data from N^* observations from N persons is given by

$$\boldsymbol{Y} = \mathbf{1}\mu + \boldsymbol{U\gamma} + \boldsymbol{V\eta} + \boldsymbol{X\beta} + \boldsymbol{Z\delta} + \boldsymbol{F\psi} + \boldsymbol{D\theta} + \boldsymbol{\epsilon} \tag{5}$$

where $\mathbf{1}$, $\boldsymbol{U}$, $\boldsymbol{X}$, and $\boldsymbol{F}$ are defined as they were with (2), and $\boldsymbol{V}$ and $\boldsymbol{Z}$ are similarly defined matrices of the time-varying unit and person variables. $\boldsymbol{D}$ is an $N^* \times N$ matrix of indicator variables for persons and $\boldsymbol{\theta}$ is a vector of person-level fixed effects. The normal equations for this model are

$$\begin{pmatrix} \mathbf{1}'(\mathbf{1}\boldsymbol{\mu} + \boldsymbol{U\gamma} + \boldsymbol{V\eta} + \boldsymbol{X\beta} + \boldsymbol{Z\delta} + \boldsymbol{F\psi} + \boldsymbol{D\theta}) \\ \boldsymbol{U}'(\mathbf{1}\boldsymbol{\mu} + \boldsymbol{U\gamma} + \boldsymbol{V\eta} + \boldsymbol{X\beta} + \boldsymbol{Z\delta} + \boldsymbol{F\psi} + \boldsymbol{D\theta}) \\ \boldsymbol{V}'(\mathbf{1}\boldsymbol{\mu} + \boldsymbol{U\gamma} + \boldsymbol{V\eta} + \boldsymbol{X\beta} + \boldsymbol{Z\delta} + \boldsymbol{F\psi} + \boldsymbol{D\theta}) \\ \boldsymbol{X}'(\mathbf{1}\boldsymbol{\mu} + \boldsymbol{U\gamma} + \boldsymbol{V\eta} + \boldsymbol{X\beta} + \boldsymbol{Z\delta} + \boldsymbol{F\psi} + \boldsymbol{D\theta}) \\ \boldsymbol{Z}'(\mathbf{1}\boldsymbol{\mu} + \boldsymbol{U\gamma} + \boldsymbol{V\eta} + \boldsymbol{X\beta} + \boldsymbol{Z\delta} + \boldsymbol{F\psi} + \boldsymbol{D\theta}) \\ \boldsymbol{F}'(\mathbf{1}\boldsymbol{\mu} + \boldsymbol{U\gamma} + \boldsymbol{V\eta} + \boldsymbol{X\beta} + \boldsymbol{Z\delta} + \boldsymbol{F\psi} + \boldsymbol{D\theta}) \\ \boldsymbol{D}'(\mathbf{1}\boldsymbol{\mu} + \boldsymbol{U\gamma} + \boldsymbol{V\eta} + \boldsymbol{X\beta} + \boldsymbol{Z\delta} + \boldsymbol{F\psi} + \boldsymbol{D\theta}) \end{pmatrix} = \begin{pmatrix} \mathbf{1}'\boldsymbol{Y} \\ \boldsymbol{U}'\boldsymbol{Y} \\ \boldsymbol{V}'\boldsymbol{Y} \\ \boldsymbol{X}'\boldsymbol{Y} \\ \boldsymbol{Z}'\boldsymbol{Y} \\ \boldsymbol{F}'\boldsymbol{Y} \\ \boldsymbol{D}'\boldsymbol{Y} \end{pmatrix} \tag{6}$$

We let $\widetilde{\mu}$, $\widetilde{\boldsymbol{\gamma}}$, $\widetilde{\boldsymbol{\eta}}$, $\widetilde{\boldsymbol{\beta}}$, $\widetilde{\boldsymbol{\delta}}$, $\widetilde{\boldsymbol{\psi}}$, and $\widetilde{\boldsymbol{\theta}}$ be an arbitrary solution to the normal equations (6). Because the sum of the columns of $\boldsymbol{F}$ and the sum of the columns of $\boldsymbol{D}$ equal $\mathbf{1}$, we can add a constant to $\widetilde{\mu}$ 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 or Frisch–Waugh–Lovell decomposition (Lovell 2008) to remove person-level means from $\boldsymbol{Y}$ and the unit indicators (that is, the columns of $\boldsymbol{F}$), 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, 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 (for example, `a2reg`, `gpreg`, or `reg2hdfe`), but this simple method yields a commonly used, intuitive solution.

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 (for example, 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 solution. For instance, dropping one unit indicator from every reference collection would provide a solution.

3.2 Stata commands for two-level fixed-effects model

We consider seven commands for estimating unit effects in the two-level fixed-effects model: `areg`, `xtreg`, and the user-written `felsdvreg`, `felsdvregdm`, `a2reg`, `gpreg`, and `reg2hdfe`. As discussed above, `areg` and `xtreg` were written to improve computational efficiency for models with many fixed effects at one level, but they can be slow for models with many fixed effects at two levels, such as a model with many teachers and many students.[7] Although `felsdvreg`, `felsdvregdm`, `a2reg`, `gpreg`, and `reg2hdfe` were all written to overcome the computational limitations of `areg` and `xtreg`, they take different approaches to achieve computational 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} 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 variables and the person fixed effects. Hence, we do not include a table for this model.

7. `regress` could again be used to estimate the parameters of models with two levels of fixed effects, but both the unit and the 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 `regress` for parameter estimation with two-level models.

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 regress 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 regress in the one-level model.

As shown in table 4 (at the end of this section), 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 ψ for each unit and the holdout. The procedures provide standard errors for the estimated contrasts.

As shown in table 5 (at the end of this section), like regress in the one-level fixed-effects model, areg and xtreg rely on Stata defaults for estimation with collinear predictors because the unit effects and time-invariant predictors are collinear with the indicator variables for the units or the persons. As a result, in models with unit effects and time-invariant unit-level predictors, both commands 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 dataset and reordering of unit identification numbers. The estimates will, in general, be difficult to interpret as any specific function of the ψ parameters.

felsdvreg

For the model with 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. 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 fitted model because it is collinear with the person effects. However, felsdvreg does provide the overall sample mean as an estimate of μ. 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 datasets than the other two commands.

For models that include time-invariant unit-level predictors [U in (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 $\psi^* = \psi + u^{'}\boldsymbol{\gamma} - \psi_1 - u_1^{'}\boldsymbol{\gamma}$. 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 command.[8]

felsdvregdm

Because felsdvregdm uses the same computational algorithm as felsdvreg to estimate the parameters of models with two levels of fixed effects, it provides efficient estimation for large datasets. However, it uses a different model parameterization from 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 it uses for the one-level fixed-effects model: it replaces ψ with $\psi^* = \psi - \overline{\psi}_r$, where $\overline{\psi}_r$ is the average of the unit means. For models without time-invariant unit-level covariates, the mean is over all units. For models with time-invariant unit-level covariates, the mean is over all units with the same value of the time-invariant unit-level covariate. As shown in tables 4 and 5 (at the end of this section), 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.

a2reg

For models without time-invariant unit-level covariates, a2reg uses a model parameterization analogous to the one used by areg, xtreg, and felsdvreg except that the holdout unit is the unit with the highest value of the ID variables. 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 fit 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

8. Whether or not the model includes time-invariant unit-level covariates, the unit effect estimates from felsdvreg are solutions to the same normal equations, but formulas for the estimates are different in tables 4 and 5. The reason is that when the model includes time-invariant unit-level covariates, they are part of the estimate even if they are not used in estimation. When the model does not include these variables, they are not part of the estimate.

fixed effects. The algorithm also does not recognize that variables in the model are collinear with the intercept, which `a2reg` also includes in the model. Consequently, the command estimates one extra model degree of freedom because it counts the redundant variables as separate variables in the model. F tests produced by the command use the wrong reference distribution. The error with the model degrees of freedom can be easily corrected provided the user knows it exists.

`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 command treats time-invariant unit-level covariates like any other covariates. Solutions are found by 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 in readily discernible ways 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 `a2reg` estimates of unit effects in this context do not appear to provide estimates of any parameters of interest.

gpreg and reg2hdfe

Like `a2reg`, `gpreg` and `reg2hdfe` use an iterative procedure to solve the normal equations and find least-squares estimates. Unlike most of the other procedures, `gpreg` and `reg2hdfe` use the algorithm of Guimarães and Portugal (2010) to directly solve the normal equations (6) without any explicit reparameterization of the model. Because `gpreg` and `reg2hdfe` use the same algorithm to solve the normal equations and estimate the fixed effects, the two commands yield nearly identical estimates of unit fixed effects with our test dataset; differences are in the fifth decimal place. With some datasets that we examined, for models without time-invariant unit-level covariates, the estimates produced by `gpreg` or `reg2hdfe` equaled those produced by `felsdvregdm`; thus the estimands matched the reparameterization used by that procedure. With other datasets, the estimates produced by `gpreg` and `reg2hdfe` differed from those produced by `felsdvregdm` by a small constant (denoted C in table 4). We could not map this constant onto any straightforward combination of unit-level quantities; hence, we could not map the estimand of `gpreg` or `reg2hdfe` to a well-defined, straightforward combination of the ψ.

For models with time-invariant unit-level covariates, `gpreg` and `reg2hdfe` explicitly fold these variables into the unit effect by definition. Consequently, the command produces the same estimates for models with and without these covariates. Again the estimates do not have a simple interpretation as contrasts of the ψ.

Neither `gpreg` nor `reg2hdfe` provides standard errors for unit or person fixed effects. Both provide standard errors for coefficients on other variables that they retain in the model, and `reg2hdfe` has an option to provide cluster-adjusted standard errors whereas `gpreg` does not.[9]

Table 4. Description of estimates from various Stata procedures for unit means and standard errors of unit means for a two-level model with only unit means, person means, and an intercept. The $\widetilde{\mu} = 0$; $\widetilde{\psi}_j$, $j = 1, \ldots, J$; and $\widetilde{\theta}_k$, $k = 1, \ldots, K$ are an arbitrary solution to the normal equations for the ordinary least-squares (OLS) estimates of these parameters, and $\overline{\widetilde{\psi}}_.$ is the average of the $\widetilde{\psi}_j$. For `gpreg` and `reg2hdfe`, C is an unspecified constant that is not a simple function of unit-level quantities.

Stata procedure	Unit effect	Std. error
`areg`	$\widetilde{\psi}_j - \widetilde{\psi}_1$ $j = 2, \ldots, J$	OLS std. error for contrast
`xtreg`	$\widetilde{\psi}_j - \widetilde{\psi}_1$ $j = 2, \ldots, J$	OLS std. error for contrast
`felsdvreg`	$\widetilde{\psi}_j - \widetilde{\psi}_1$ $j = 2, \ldots, J$	OLS std. error for contrast
`felsdvregdm`	$\widetilde{\psi}_j - \overline{\widetilde{\psi}}_.$ $j = 1, \ldots, J$	OLS std. error for effect
`a2reg`	$\widetilde{\psi}_j - \widetilde{\psi}_J$ $j = 1, \ldots, J-1$	Not provided
`gpreg`	$\widetilde{\psi}_j - C$ $j = 1, \ldots, J$	Not provided
`reg2hdfe`	$\widetilde{\psi}_j - C$ $j = 1, \ldots, J$	Not provided

9. The `regress`, `fese`, and `felsdvreg` procedures provide cluster-adjusted standard errors for the unit fixed effects. The remaining procedures do not. However, like `reg2hdfe`, `areg` and `xtreg` include an option for estimating clustered-adjusted standard errors for coefficients other than the unit or person fixed effects.

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 $\widetilde{\psi}_{j_g}$ and $\widetilde{\gamma}$ form an arbitrary solution for the OLS normal equations for the unit means $j_g = 1, \ldots, J_g$ of reference collection $g = 1, \ldots, G$ and coefficients for the parameters on the time invariant variables u. The $\widetilde{\psi}_{g.}, 1, \ldots, G$ equal the refererence collection means of the is $\widetilde{\psi}_{j_g}$. For all procedures, unit means are estimated for all reference collections. For `gpreg` and `reg2hdfe`, C is an unspecified constant that is not a simple function of unit-level quantities.

Stata procedure	Coefficient for ref. coll. mean	Unit effect	Std. error
`areg`	No set pattern; arbitrary units and group means dropped		OLS std. errors for estimate
`xtreg`	No set pattern; arbitrary units and group means dropped		OLS std. errors for estimate
`felsdvreg`	Not provided	$\widetilde{\psi}_j + u_j'\widetilde{\gamma} - \widetilde{\psi}_1 - u_1'\widetilde{\gamma}$ $j = 2, \ldots, J$	OLS std. error for contrast
`felsdvregdm`	$\overline{\overline{\psi}}_{g.} - \overline{\overline{\psi}}_{G.}$ $g = 1, \ldots, G-1$	$\widetilde{\psi}_{gj} - \overline{\overline{\psi}}_{g.}$ $j_g = 1, \ldots, J_g$	OLS std. error for effect
`a2reg`	Not provided	Does not correspond to a combination of unit values	Not provided
`gpreg`	Not provided	$\widetilde{\psi}_j + u_j'\widetilde{\gamma} - C$ $j = 1, \ldots, J$	Not provided
`reg2hdfe`	Not provided	$\widetilde{\psi}_j + u_j'\widetilde{\gamma} - C$ $j = 1, \ldots, J$	Not provided

3.3 Example code for commands for two-level fixed-effects model

Below is example code for using the commands for two-level fixed effects. Each command again uses the simulated dataset `statafetest`. As with the example for one-level fixed-effects models, we present code for using each command to fit a model with the unit-level predictor `year` because, for most commands, the code for fitting models with different covariates requires only a change in specification of the covariates. For `a2reg` and `gpreg`, we present the code for the model with no covariates along with the code for the model with year indicators because these commands cannot fit a model with no specified covariates and require the inclusion of the `con` variable to fit the no-covariate model. For models with specified covariates, a constant is included by default in both these commands, and the `con` variable is not required. For **`felsdvregdm`**, we present code for fitting all three models because, as with the one-level fixed-effects model, the `reff()` parameter differs between models with and without a discrete unit-level predictor. For **`areg`** and **`xtreg`**, the teacher indicator variables are explicitly included in the model through the use of `` `teacherfes' `` and are extracted and saved along with their standard errors with the `estimates` command. For all other commands, the fixed effects and standard errors (when available) are assigned to variables via options in the call to the command. Refer to the help files for each command for details on its syntax and options.

Example code for two-level fixed-effects models:

```
. use statafetest, clear
. * generate a constant 1 to use with some routines
. generate con = 1
. * generate teacher indicators to use in some routines
. xi, prefix(T) i.tchid
  (output omitted)
. unab teacherfes: T*
. * areg
. areg y `teacherfes´ year1 year2 year3, absorb(sid)
  (output omitted)
. estimates store areg2L
. * xtreg
. xtreg y `teacherfes´ year1 year2 year3, i(sid) fe
  (output omitted)
. estimates store xtreg2L
. * felsdvreg
. felsdvreg y year1 year2 year3, cons ivar(sid) jvar(tchid) feff(felsest2L)
> peff(peffhat) feffse(felsse2L) mover(mover1) group(group) xb(xb1) res(res1)
> mnum(mnum1) pobs(pobs1)
  (output omitted)
. * felsdvregdm
. * Model without covariates -- reff() specifies a single reference collection
. felsdvregdm y, ivar(sid) jvar(tchid) feff(felsdmest2L1) peff(peffhat1)
> reff(con) feffse(felsdmse2L1) mover(mover1) group(group) xb(xb1) res(res1)
> mnum(mnum1) pobs(pobs1)
  (output omitted)
. * Model with unit-level covariates -- reff() specifies reference collections
. * equal to values of the covariates
. felsdvregdm y year1 year2 year3, ivar(sid) jvar(tchid) feff(felsdmest2L2)
> peff(peffhat1) reff(year) feffse(felsdmse2L2) mover(mover1) group(group)
> xb(xb1) res(res1) mnum(mnum1) pobs(pobs1)
  (output omitted)
```

```
. * Model with time-varying person-level covariates -- reff() specifies
. * a single reference collection
. felsdvregdm y x, ivar(sid) jvar(tchid) feff(felsdmest2L3) peff(peffhat1)
> reff(con) feffse(felsdmse2L3) mover(mover1) group(group) xb(xb1) res(res1)
> mnum(mnum1) pobs(pobs1)
  (output omitted)
. * a2reg
. * Model without covariates -- requires explicit inclusion of con variable
. a2reg y con, individual(sid) unit(tchid) uniteffect(a2regest2L1)
  (output omitted)
. * Model with covariates -- does not require inclusion of con variable
. a2reg y year1 year2 year3, individual(sid) unit(tchid) uniteffect(a2regest2L2)
. * gpreg
. * Model without covariates -- requires explicit inclusion of con variable
. gpreg y con, ivar(sid) jvar(tchid) ife(tmp) jfe(gpregest2L1)
  (output omitted)
. drop tmp
. * Model with covariates -- does not require inclusion of con variable
. gpreg y year1 year2 year3, ivar(sid) jvar(tchid) ife(tmp) jfe(gpregest2L)
  (output omitted)
. drop tmp
. * reg2hdfe
. reg2hdfe y year1 year2 year3, id1(sid) id2(tchid) fe1(tmp) fe2(reg2hdfeest2L)
  (output omitted)
. drop tmp
```

4 Computational efficiency

In addition to the requisite estimates and parameterization, the choice of which Stata command to use will also depend on the size of the sample, speed considerations, and available computing resources. The available Stata commands use various computational algorithms to reduce the computational time required to estimate the parameters of interest. The various commands, however, are not equally efficient at computing estimates, and some commands can take very long to produce estimates and standard errors. In some instances, some commands cannot compute estimates 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 commands and some guidance about the utility of alternative procedures for conducting analyses. The timing data reported in table 6 (at the end of this section) are based on running Stata/MP 12 with four processors under Windows 7 Professional x64 on a machine with two 6-core Intel Xeon CPUs (X5660) running at 2.80 GHz and 48 GB of RAM.

For these comparisons, we estimated teacher effects by grade for grade 4 to 8 mathematics teachers teaching in Palm Beach County, Dade County, or the eight largest counties in Florida during the 2001–2002 to 2004–2005 school years. All are large school systems, and the datasets range from 99,397 records on 39,888 students to 774,156 records on 304,006 students and 2,587 teachers to 14,831 teacher-by-grade units.

We fit models with both one level (teacher) and two levels (teachers and students) of fixed effects so that we could compare the performance of all commands. Models with one level of fixed effects also include 16 time-varying teacher variables, 14 time-constant student variables, 3 time-varying student variables, and 5 grade indicators. Models with two levels of fixed effects also included all the covariate variables from the one-level models except the 14 time-constant student variables, which were dropped (and replaced by the student fixed effect).

For the case of one-level models, `areg` and `xtreg` are roughly equivalent in their speed and use of resources because both difference out the single fixed effect. The use of `regress` with the `xi` command is inferior in terms of speed and memory usage because it requires generating explicit indicator variables for the fixed effects. The use of `regress` is further constrained by the 11,000 variable limit inherent in Stata. Estimation with the `regress` 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 large samples. `regress` is slower than `felsdvregdm`, even in samples of modest size, and does not scale up like `felsdvregdm`.

Computational efficiency is a greater concern for fitting two-level models, and the choice of estimation command is less clear cut. There are tradeoffs in terms of the parameters that are estimated, parameterization of the fixed effects, computational speed, and memory requirements. As in the one-level case, the 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 because it is much faster than any of the alternatives. It can be used with very large datasets and does not require vast amounts of memory. However, as noted above, one must be careful to check for collinearities ex-ante because `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 because computational time is directly proportional to the number of bootstraps. Even a 100-repetition bootstrap would make `a2reg` much slower than other alternatives.

For small and moderately large samples ($<$ 750,000 observations), `felsdvreg` and `felsdvregdm` are faster than `gpreg` and `reg2hdfe`. However, as the sample size grows, the speed advantage of `felsdvreg` and `felsdvregdm` dissipates. Because they are iterative commands, `gpreg` and `reg2hdfe` have a time to fit models that grows in direct proportion with sample size whereas estimation time for `felsdvreg` or `felsdvregdm` grows as approximately the cube of the number of units due to the matrix inversion. `felsdvreg` and `felsdvregdm` are probably not practical for samples larger than 1.5 million observations, whereas Guimarães and Portugal (2010) report fitting a two-level model on a sample of over 30 million observations with `reg2hdfe`. With our test data, `gpreg` using the default algorithm found estimates in about half of the time that `reg2hdfe` required. However, `gpreg` using other algorithms actually required more time than `reg2hdfe`.

`felsdvreg`, `felsdvregdm`, `gpreg`, and `reg2hdfe` 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. It also means that relatively little is gained from devoting more processors to a job. While many Stata commands have been optimized for multithreading, Mata-based programs generally do not yield significant performance gains from additional processors. When we devoted eight processors rather than four to each of the four Mata-based programs, estimation times improved by 0% to 16%.

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

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 County, Dade County, 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.

	Palm Beach	Dade	8 largest counties
Number of student-year obs.	99,397	250,698	774,156
Number of students	39,888	100,077	304,006
Number of teacher-by-grade units	2,587	5,239	14,831
One-level models:			
`regress`	0:02:44	0:21:02	N/F
`areg`	0:00:01	0:00:01	0:00:02
`xtreg`	0:00:04	0:00:10	0:00:21
`fese`	0:43:04	3:41:34	N/F
`felsdvregdm`	0:01:10	0:09:09	4:10:56
Two-level models:			
`areg`	0:13:38	2:05:07	N/F
`xtreg`	0:34:31	4:50:46	N/F
`felsdvreg`	0:00:46	0:06:30	3:45:34
`felsdvregdm`	0:01:17	0:10:02	4:11:53
`a2reg`	0:00:20	0:01:00	0:02:37
`gpreg`	0:07:40	0:22:43	0:57:48
`reg2hdfe`	0:14:36	0:53:52	2:00:24

5 Conclusion

The recent increase in large administrative datasets for the estimation of firm, teacher, or school effects has been matched by an increase in the development of Stata commands 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 commands differ along three important dimensions: the estimands that they estimate (for example, 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 mean and a holdout unit mean; 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 `regress` with the one-level fixed-effects model and for `areg`, `xtreg`, and `a2reg` with the two-level fixed-effects model. These procedures should be avoided in these contexts because 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 computationally demanding. For this reason, many procedures do not provide standard-error estimates for the unit fixed effects. In addition, commands 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. However, 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 grow faster than linearly with the 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, if standard errors are desired, then procedures such as `gpreg` or `reg2hdfe` might be the only approach to estimation. Alternatively, if only point estimates are required, `a2reg` can be used; 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 estimating 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 and 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

Abowd, J. M., F. Kramarz, and S. Roux. 2006. Wages, mobility and firm performance: Advantages and insights from using matched worker–firm data. *Economic Journal* 116: F245–F285.

Bazzoli, G. J., H.-F. Chen, M. Zhao, and R. C. Lindrooth. 2008. Hospital financial condition and the quality of patient care. *Health Economics* 17: 977–995.

Cornelissen, T. 2008. The Stata command felsdvreg to fit a linear model with two high-dimensional fixed effects. *Stata Journal* 8: 170–189.

Goldhaber, D., B. Gross, and D. Player. 2011. Teacher career paths, teacher quality, and persistence in the classroom: Are public schools keeping their best? *Journal of Policy Analysis and Management* 30: 57–87.

Guimarães, P., and P. Portugal. 2010. A simple feasible procedure to fit models with high-dimensional fixed effects. *Stata Journal* 10: 628–649.

Harris, D. N., and T. R. Sass. 2011. Teacher training, teacher quality and student achievement. *Journal of Public Economics* 95: 798–812.

Jacob, B. A., and L. Lefgren. 2008. Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics* 26: 101–136.

Lovell, M. C. 2008. A simple proof of the FWL theorem. *Journal of Economic Education* 39: 88–91.

Mihaly, K., D. F. McCaffrey, J. R. Lockwood, and T. R. Sass. 2010. Centering and reference groups for estimates of fixed effects: Modifications to felsdvreg. *Stata Journal* 10: 82–103.

Nichols, A. 2008. fese: Stata module to calculate standard errors for fixed effects. Statistical Software Components S456914, Department of Economics, Boston College. http://ideas.repec.org/c/boc/bocode/s456914.html.

Ouazad, A. 2008. a2reg: Stata module to estimate models with two fixed effects. Statistical Software Components S456942, Department of Economics, Boston College. http://ideas.repec.org/c/boc/bocode/s456942.html.

Schmieder, J. F. 2009. gpreg: Stata module to estimate regressions with two dimensional fixed effects. Statistical Software Components S457048, Department of Economics, Boston College. http://ideas.repec.org/c/boc/bocode/s457048.html.

Searle, S. R. 1971. *Linear Models*. New York: Wiley.

About the authors

Daniel F. McCaffrey is a senior statistician at RAND, where he holds the PNC Chair in Policy Analysis. His current research focuses on value-added modeling for the estimation of teacher effects and the use of those estimates in teacher evaluation and school reform. He is currently working on several projects to study the relationship between value-added and other forms of teacher evaluations. His other research interests include causal modeling, primarily with application to adolescent substance abuse treatment.

J. R. Lockwood is a senior statistician at RAND. His education research has focused on longitudinal modeling of student achievement, value-added models for estimating teacher effects, and experimental and quasi-experimental methods in educational evaluation. He has led two projects to develop enhanced statistical models for estimating teacher effects and to develop computational methods for implementing complex models with large datasets.

Kata Mihaly is an associate economist at RAND. Her research on education focuses on value-added models of teacher quality and the effect of peers on student achievement. Her other research interests include peer effects on adolescent substance abuse and the role of social networks on the decision of adults to save for retirement.

Tim Sass is a professor of economics at Georgia State University, where he teaches courses in applied microeconomics at the undergraduate and graduate levels. He is also a member of the National Center for Analysis of Longitudinal Data in Education Research. His current research focuses on various aspects of education policy, including teacher quality, charter schools, and peer effects.