

Centering and reference groups for estimates of fixed effects: Modifications to `felsdvreg`

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

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

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

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

Abstract. Availability of large, multilevel longitudinal databases in various fields including labor economics (with workers and firms observed over time) and education research (with students and teachers observed over time) has increased the application of panel-data models with multiple levels of fixed-effects. Existing software routines for fitting fixed-effects models were not designed for applications in which the primary interest is obtaining estimates of any of the fixed-effects parameters. Such routines typically report estimates of fixed effects relative to arbitrary holdout units. Contrasts to holdout units are not ideal in cases where the fixed-effects parameters are of interest because they can change capriciously, they do not correspond to the structural parameters that are typically of interest, and they are inappropriate for empirical Bayes (shrinkage) estimation. We develop an improved parameterization of fixed-effects models using sum-to-zero constraints that provides estimates of fixed effects relative to mean effects within well-defined reference groups (e.g., all firms of a given type or all teachers of a given grade) and provides standard errors for those estimates that are appropriate for shrinkage estimation. We implement our parameterization in a Stata routine called `felsdvregdm` by modifying the `felsdvreg` routine designed for fitting high-dimensional fixed-effects models. We demonstrate our routine with an example dataset from the Florida Education Data Warehouse.

Keywords: `st0185`, `felsdvreg`, `felsdvregdm`, fixed effects, linked employer–employee data, longitudinal achievement data

1 Introduction

The implementation of models that include a large number of fixed effects for individual units, such as employers and employees or teachers and students, has grown with the increased availability of large longitudinal datasets.¹ This is especially evident

1. Units do not need to be synonymous with physical units. A unit could be a firm or teacher during a period of time (e.g., year) or could be a subset of its function (e.g., a firm’s manufacturing division versus its research division or a teacher’s first period versus his second period class) so that the same physical entity could be associated with multiple units. The example in section 4 defines units as teacher-years so that teachers can be associated with multiple units.

in the context of employer–employee linked data (Abowd, Kramarz, and Roux 2006; Menezes-Filho, Muendler, and Ramey 2008) and student–teacher linked data (Aaronson, Barrow, and Sander 2007; Clotfelter, Ladd, and Vigdor 2006; Harris and Sass 2006; McCaffrey, Han, and Lockwood 2008; Rockoff 2004). It is usually not possible to estimate all the models’ parameters because inclusion of a fixed effect for every individual unit generally makes the model overparameterized because of linear dependence among the independent variables in the model (e.g., the indicator or “dummy” variables for the individual units). Consequently, estimation requires identifying assumptions and reparameterization of the model. In cases where the fixed effects are used to control for unobserved heterogeneity of units (e.g., workers, firms, students, teachers) and interest lies in the estimation of the effects of other time-varying covariates (e.g., exposure to policies or programs), then the reparameterization is irrelevant because parameters and their estimates for time-varying covariates are invariant to different parameterizations of the fixed effects. In fact, when the model contains only one level of fixed effects, it is typically not explicitly included as regressors but rather is “absorbed” out prior to regression through the within-subjects transformation (Greene 2008; Wooldridge 2002). This is how several commonly used packages such as `areg`, `xtreg`, `fe`, and SAS[®] PROC GLM (SAS Institute, Inc. 2004) implement single-level fixed-effects estimation. Even if multiple levels of fixed effects are in the model, if the effects at all levels are nuisances, then the absorption can be applied with “spell” fixed effects that allow one effect for each unique combination of the multiple levels of indicators (e.g., each unique worker–firm combination) (Andrews, Schank, and Upward 2006).

However, there are some applications where the fixed effects are of direct interest, either by themselves or in conjunction with the effects of time-varying covariates. This is particularly true in education research, where there is growing interest in using large, longitudinal achievement databases with students linked to teachers and schools to estimate the effects of individual teachers on student achievement (Goldhaber and Hansen 2008; Harris and Sass 2006, 2007; Kane, Rockoff, and Staiger 2006; McCaffrey et al. 2004; McCaffrey et al. 2009) and increasing calls to use such estimates for high-stakes purposes such as pay or tenure decisions (Gordon, Kane, and Staiger 2006). Similarly, in labor economics, there is interest in estimating the effects of individual firms (Abowd, Creecy, and Kramarz 2002; Abowd, Kramarz, and Margolis 1999).

In cases where estimates of unit fixed effects are of interest, the parameterization matters and it is essential for the analyst to be able to control it. However, software for modeling with fixed effects (e.g., `areg`, `xtreg`, `fe`, or SAS[®] PROC GLM) all solve the problem of overparameterization by deleting the indicator variables for one or more units until the remaining variables are linearly independent. Although this parameterization suffices in allowing for estimation of the remaining model parameters, it changes the definition of the fixed-effects parameters for indicator variables to be contrasts between the units that were not deleted and the arbitrary holdout unit or units whose indicators were deleted.

Comparison to a holdout unit results in estimates that are not invariant to the chosen holdout unit, and estimates can change as a result of immaterial changes to the data, such as the labeling or ordering of units. In more complex settings and in models

where more than one indicator variable must be deleted to solve the overparameterization of fixed effects, it can be difficult to determine the specific interpretation of the reported contrasts. This situation is unacceptable in cases where obtaining estimates of the fixed-effects parameters is a primary objective. In addition, estimation of contrasts to arbitrary holdout units can result in instability of estimates across time or other settings due to differences in the holdouts. The problems of interpretation and instability potentially can be overcome by using linear combinations of model parameters to estimate individual units' contributions (e.g., combining the intercept and the individual-unit fixed effects to estimate the unit-specific mean, or post hoc mean, centering the estimated effects to estimate the effect of the unit relative to the average unit); however, the standard errors of such combinations are not easily obtained with available software.

Hence, there is a need for software that estimates the fixed effects using a parameterization that is interpretable, not capricious, and that provides standard errors for these estimates. A parameterization meeting these criteria is one in which effects for individual units are defined as deviations from the average unit or the average units within specified reference groups. These contrasts more directly correspond to the parameters of structural (causal) models; for example, in education settings, the notion of an “effective teacher” is a teacher whose students perform better with that teacher than they would with the average teacher in the population. These estimates are obtained by enforcing sum-to-zero constraints among the fixed effects in the model.

Users could code independent variables to correspond to the sum-to-zero parameterization and then include these as variables in standard linear model software. However, coding the independent variables is tedious, especially if multiple constraints are necessary to remove overparameterization. In addition, when there are many units or there are two levels of fixed effects (e.g., firms and workers or teachers and students), creating the complete matrix of independent variables can be computationally burdensome or infeasible.

To address the need for software designed to estimate fixed effects when they are of interest, we introduce the `felsdvregdm` command, which automatically parameterizes the fixed effects using sum-to-zero constraints. The command allows the user to define the collection of units to be compared (called “reference collections”) and calculates the sum-to-zero parameterization so that the fixed-effects parameters are deviations from the reference collection means. For example, in the education setting, it is possible to calculate teacher fixed effects that are deviations from the average teacher in the school, the average teacher in the grade across schools, or any other grouping of teachers of interest to researchers. The program works with a single level of fixed effects (firms or teachers) and with two levels of effects (workers and firms or students and teachers). The program builds on the efficient computational methods of `felsdvreg` (Cornelissen 2008), which allow it to handle datasets with a large number of units and individuals.

`felsdvregdm` complements several commands that already exist in Stata to handle fixed effects. For a model with one level of fixed effects, `felsdvregdm` produces estimates and inferences for effects of time-varying covariates that are identical to `areg`,

`xtreg`, or `fese` (Nichols 2008). All four of these procedures will produce the same overall model fit and residual error because the alternative parameterizations of fixed effects do not affect the overall model fit. The only difference in the results from `felsdvregdm` and the other procedures is the estimates of the fixed-effects parameters and their standard errors. Similarly, for a model with two levels of fixed effects, the estimates and inferences for time-varying covariates and model fit from `felsdvregdm` are identical to those produced by `felsdvreg`, `a2reg` (based on Abowd, Creecy, and Kramarz [2002]) or `gpreg` (based on Guimarães and Portugal [2009]), with differences only in the estimates of the unit fixed-effects parameters and their standard errors.² In neither case does the fixed-effects parameterization used by `felsdvregdm` add any additional constraints to the model.

In this article, `felsdvregdm` is introduced and implemented in an example dataset from the Florida Education Data Warehouse in the context of educational achievement of students linked to teachers.

2 Sum-to-zero constraints

2.1 One level of fixed effects

We roughly follow the notation established by Cornelissen (2008). The basic model with a single level of fixed effects assumes that the outcome for a “person” i with time-varying covariates x_{it} associated with “unit” j at time t is given by

$$y_{it} = x'_{it}\boldsymbol{\beta} + \psi_{j(i,t)} + \epsilon_{it}$$

where ϵ_{it} is a mean zero error term and there is a separate factor ψ_j for each unit, with the index $j(i, t)$ indicating the unit j to which person i is linked at time t . The fundamental problem with this model is that without additional assumptions, the average of the ψ_j cannot be distinguished from the average person outcome, or more generally, the model with separate factors for every unit is overparameterized and model parameters cannot be estimated uniquely.

The model for the outcomes from a typical sample of panel data from N persons, with a total of N^* person-time observations is given by

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{F}\boldsymbol{\psi} + \boldsymbol{\epsilon} \quad (1)$$

where \mathbf{F} is an $N^* \times J$ incidence matrix consisting of only zeros and ones and satisfying $\mathbf{F}\mathbf{1}_J = \mathbf{1}_{N^*}$ where $\mathbf{1}_k$ is a k -vector of ones. That is, each observation is linked to exactly one of the J units. \mathbf{X} is $N^* \times K$ containing time-varying covariates, as well as time effects or an overall intercept. Overparameterization is then characterized by the

2. The two-level model can also be fit via `areg` by explicitly including indicator variables for the unit fixed effects via the `xi` command. This will yield the same estimates of time-varying covariates and model fit as `felsdvregdm`, but it will not match the estimates for the fixed effects because the `xi` command automatically deletes dummy variables to identify the model. Using `areg` for two-level models will not be feasible when there are large numbers of units because of memory constraints and Stata’s 11,000-variable matrix size limit.

matrix (\mathbf{X}, \mathbf{F}) being less than full column rank. Under the assumptions about \mathbf{F} , this will always be true when \mathbf{X} contains either an intercept or separate means for every time point, because in either case it is possible to recover $\mathbf{1}_{N^*}$ as a linear combination of columns of \mathbf{X} or columns of \mathbf{F} separately, and thus the columns of (\mathbf{X}, \mathbf{F}) are not linearly independent. As noted previously, the most common approach to this overparameterization is to remove as many columns of \mathbf{F} as there are linear constraints among the columns of (\mathbf{X}, \mathbf{F}) .³ In the simplest case, $\text{rank}\{(\mathbf{X}, \mathbf{F})\} = \text{rank}(\mathbf{X}) + J - 1$; that is, the introduction of \mathbf{F} results in only one linear dependency. For instance, there is one linear dependency when \mathbf{X} consists of only the intercept or includes only indicators for time-point means.

Our solution to the problem is to replace \mathbf{F} with an $N^* \times (J - 1)$ matrix \mathbf{F}^* with associated parameter vector $\boldsymbol{\psi}^*$. \mathbf{F}^* is obtained by dropping an arbitrary column j of \mathbf{F} and setting all rows of \mathbf{F}^* corresponding to observations linked to unit j to a vector of negative ones. That is, we define the effect of unit j to be $1 - \sum_{i=1, \dots, J; i \neq j} \psi_i$ so that $\boldsymbol{\psi}^*$ contains $J - 1$ free parameters rather than J free parameters because the elements of $\boldsymbol{\psi}^*$ sum to zero. Unlike the default holdout parameterization, these contrasts and their estimates are invariant to which unit- j column is eliminated from \mathbf{F} , and without loss of generality, we eliminate the first column of \mathbf{F} .⁴ In this case, we can express $\mathbf{F}^* = \mathbf{F}\mathbf{C}$, where \mathbf{C} is the $J \times (J - 1)$ matrix $(-\mathbf{1}, \mathbf{I})^T$ (Venables and Ripley 2001).

More generally, analysts may be interested in controlling for differences among collections of firms or teachers. For example, firms may have classification factors such as size or sector that the analysts want to remove from the individual unit effects. In education applications, partitioning of teachers into comparison groups is commonplace. For datasets following a single cohort of students across grades, it is typical to compare teachers of the same grade or same group of grades (i.e., elementary versus middle grades). In more complex datasets, where multiple cohorts of students are followed and teachers are observed for multiple years across different cohorts, it may be desirable to estimate separate teacher-year effects for each teacher and to compare teachers within the same year and perhaps the same grade or grade range within year (McCaffrey et al. 2009).

In the context of estimating firm or teacher effects, controlling for a classification variable creates additional redundant parameters and requires additional constraints on the unit-effect parameters for identification. We cannot distinguish average performance for units in a group from the effects of being in the group. One approach that yields interpretable estimates for the firms and teachers is to parameterize the model so that units are compared within groups defined by the classification variables but not across those groups. In general, establishing the comparison groups is an analyst decision that depends on the substantive application.

3. Sometimes, depending on the specific routine and the order in which regressors are specified, elements of \mathbf{X} rather than \mathbf{F} may be dropped. This is equally problematic even though no elements of \mathbf{F} are dropped because the estimates of the fixed effects contain the omitted intercept or time effect(s) and are thus still subject to arbitrary choices made by the regression routine.

4. This parameterization is invariant to the number of units, J , and is equally effective on a small or large scale.

We refer to such comparison groups as “reference collections” to avoid confusion with the “groups” described by Cornelissen (2008). Cornelissen’s groups are relevant in models with two levels of fixed effects and are discussed in the next section. Specifically, we define reference collections as any partitioning of the J units into G mutually exclusive and exhaustive categories, and we parameterize the effects within each of these $g = 1, \dots, G$ categories with sum-to-zero constrained effects. Thus there are $J - G$ parameters for the effects of the J units. If the columns of \mathbf{F} are sorted by reference collection, then the $N^* \times (J - G)$ design matrix for the fixed effects is given by $\mathbf{F}^* = \mathbf{F}(\oplus \mathbf{C}_g)$ where \mathbf{C}_g is the matrix \mathbf{C} defined previously, restricted to reference collection g .

We would expect that analysts would almost always include reference collection means in the model, because the motivation for creating a reference collection will typically be to control for mean differences between those collections. However, some analysts might not follow this convention; in those cases, to ensure that our parameterization has the correct interpretation, we add those means to the model. We let \mathbf{G} be the $N^* \times G$ incidence matrix such that $\mathbf{G}_{GI} = 1$ if the unit to which person-time observation i is linked is in reference collection g and $\mathbf{G}_{GI} = 0$ otherwise. The columns of \mathbf{G} correspond to the reference collection means and the column space of $(\mathbf{F}^*, \mathbf{G})$ is identical to \mathbf{F} . We define $\mathbf{X}^* = (\mathbf{X}, \mathbf{G})$ with parameter β^* , and as discussed in more detail later in the article, our routine uses \mathbf{X}^* rather than the user-supplied \mathbf{X} to ensure that our parameterization of unit fixed effects introduces no other changes to the model.⁵ The resulting equation estimated in the case of one-level fixed effects is given by

$$\mathbf{Y} = \mathbf{X}^* \beta^* + \mathbf{F}^* \psi^* + \epsilon$$

2.2 Two levels of fixed effects

We extend (1) to include person-level fixed effects using the following model:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{F}\psi + \mathbf{D}\theta + \epsilon \quad (2)$$

where \mathbf{D} is an $N^* \times N$ matrix of person (e.g., worker or student) fixed effects. Unlike the unit effects, in the vast majority of applications, these fixed effects are nuisance parameters used to control for person-level heterogeneity when estimating unit effects or the effects of time-varying covariates. `felsdvregdm` can implement both this two-level fixed-effects model or the one-level model (1). When fitting the two-level model, person-level fixed effects are treated as nuisance parameters and are absorbed using the within transformation (Cornelissen 2008).

The essential logic of our approach to parameterizing the unit effects is unchanged by the use of (2) and the within transformation to remove the person fixed effects. However, the elements of \mathbf{X}^* and \mathbf{F}^* reflect the within transformation. The transformed version of (2) after the reparameterization can be written as

5. If any of the columns \mathbf{G} are in the column space of \mathbf{X} , they are deleted from the model during estimation.

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}^* \boldsymbol{\beta}^* + \tilde{\mathbf{F}}^* \boldsymbol{\psi}^* + \tilde{\boldsymbol{\epsilon}}$$

The \sim reflects the within-person transformed data in which the person mean has been removed from each observation and which leads to the \mathbf{D} matrix becoming the null matrix. $\tilde{\mathbf{X}}^* = (\tilde{\mathbf{X}}, \tilde{\mathbf{G}})$ is an $N^* \times (K + G)$ matrix that includes the time-varying covariates and the reference collection indicators after the within transformation, and $\tilde{\mathbf{F}}^*$ is the design matrix for the unit fixed effects after the sum-to-zero reparameterization and within-person transformation.

Often the inclusion of the person fixed effects results in no additional lost degrees of freedom beyond those already discussed and so no additional considerations are necessary. However, there are certain cases in which additional degrees of freedom may be lost. These are cases in which multiple disconnected “groups” or “strata” exist in the data (Abowd, Creedy, and Kramarz 2002; Cornelissen 2008; McCaffrey et al. 2009). Groups or strata are subgroups of units and persons that are disconnected from other subgroups in the sense that no person from that subgroup is ever linked to units outside the subgroup, and the units in the subgroup only ever link to persons within the subgroup. The simplest example can occur in education data: if students and teachers never switched schools, then each school and the students attending it would be disconnected groups.

In two-level models, groups must be considered when defining reference collections because they create additional linear dependencies among the columns of \mathbf{D} and \mathbf{F} . If centering within a particular reference collection is desired but that reference collection spans two or more disconnected groups, sum-to-zero constraints within the reference collection would fail to identify the model parameters in the context of (2). This is because comparisons of units that are in the same reference collection but in different groups conflate differences between the groups with differences between the units. One remedy for this issue is to redefine the reference collections by intersecting them with group, i.e., splitting any reference collections that spanned groups into separate reference collections by group. Other alternatives might also be possible. The key is that for each reference collection, all the units must belong entirely to one and only one group, although groups can include multiple reference collections. The appropriate solution will depend on the substantive goals of the analysis. We discuss how `felsdvregdm` deals with the relationship between groups and reference collections for the two-level model in section 3.3, and we give an example of modeling with reference collections and groups in section 4.

An extreme case of grouping occasionally occurs in which a group contains a single firm (teacher) because every worker employed by the firm (student taught by the teacher) is linked only to this firm (teacher). Cornelissen (2008), refers to these persons (workers or students) associated with a single unit as “nonmovers” because they do not move across the units but stay at the same one. These nonmovers can occur in any group and do not contribute to the estimation of the unit effects in the two-level model. However, when the data for a unit comes only from nonmovers (i.e., the unit creates a single unit group), then the unit effect cannot be estimated in the two-level model. The

unit fixed effect is excluded from the model. The persons are included in the estimation with their fixed effects and can contribute to the estimation of the coefficients for the time-varying covariates.

3 Implementation of `felsdvregdm`

3.1 Syntax

```
felsdvregdm varlist [if] [in], ivar(varname) jvar(varname) reff(varname)
peff(name) feff(name) xb(name) res(name) mover(name) mnum(name)
pobs(name) group(name) [grouponly onelevel noisily cholsolve
nocompress feffse(name)]
```

3.2 Options

`i`var(*varname*) specifies the identification variable for the person-level effect, such as the person or student ID. `i`var() is required.

`j`var(*varname*) specifies the identification variable for the unit-level effect, such as the firm or teacher ID. `j`var() is required.

`r`eff(*varname*) specifies the identification variable for the reference collection. Reference collections specify collections of units to be compared with the unit mean for the collection. For two-level fixed-effects applications, reference collections must be nested within groupings or strata of connected units. `r`eff() is required.

`p`eff(*name*) specifies the name of the new variable in which to store the person-level effect.⁶ `p`eff() is required.

`f`eff(*name*) specifies the name of the new variable in which to store the unit effect.⁷ `f`eff() is required.

`x`b(*name*) specifies the name of the new variable in which to store the linear prediction $\mathbf{X}'\mathbf{b}$. `x`b() is required.

`r`es(*name*) specifies the name of the new variable in which to store the residual. `r`es() is required.

`m`over(*name*) specifies the name of the new variable in which to store an indicator variable for whether the person is observed with multiple firms or teachers. `m`over() is required.

`m`num(*name*) specifies the name of the new variable in which to store the number of movers per unit. In the two-level model, units for which `m`num() is zero do not have identified effects. `m`num() is required.

6. This is referred to as the first effect in the `felsdvreg` documentation.

7. This is referred to as the second effect in the `felsdvreg` documentation.

`pobs(name)` specifies the name of the new variable in which to store the number of observations per person. Persons for whom only one observation is available will have person effects computed with a residual of zero for the two-level model. `pobs()` is required.

`group(name)` specifies the name of the variable in which to store the group (strata) identifier calculated during the routine. If the variable does not already exist in the data, it will be created; if there is an existing variable *name*, it will be overwritten. `group()` is required.

`grouponly` causes `felsdvregdm` to run only the grouping algorithm and produce group indicator variables. No estimates are computed in this case.

`onelevel` omits the person fixed effects, resulting in the one-level fixed-effects model being fit.

`noisily` provides more detailed outputs with tables and summary statistics on the number of movers, units, and persons, and the mobility variable.

`cholsolve` directs Stata to use the `cholsolve()` function to invert the moment matrix instead of using the default `invsym()` function. This may result in greater internal precision.

`nocompress` specifies not to compress the dataset; by default, the `compress` command is implemented to save memory.

`feffse(name)` specifies the name of the new variable in which to store the standard errors of the estimated unit fixed effects.

3.3 Remarks

Because `felsdvregdm` extends the functionality of `felsdvreg` and uses syntax and its estimation methods as its base, the procedures for the two commands are very similar. In particular, `felsdvregdm` requires the same specification of person-related variables as `felsdvreg`. This is true for both one- and two-level applications, even though in principle, several variables might only apply to models with two levels of fixed effects (e.g., `ivar()`, `peff()`, `mover()` and `pobs()`). All variables are required regardless of whether one or two levels are specified because the computational methods developed in `felsdvreg` and used by `felsdvregdm` rely on both the unit and person levels of indicators to save time and memory during estimation. Also consistent with `felsdvreg`, when specifying the `grouponly` option, all the variable names need to be defined even though only the `group()` variable will be modified and saved.

Although the syntax of `felsdvregdm` and `felsdvreg` are very similar, there are several important differences to note between the procedures:

- The `reff()` option, which is not part of `felsdvreg`, is required by `felsdvregdm`. This variable specifies the reference collections inside which the user wishes to impose sum-to-zero constraints among the unit effects.

As noted in the previous section, maintaining correspondence between the models in the original parameterization and our sum-to-zero constrained parameterization requires that \mathbf{G} , or something equivalent to it (in the sense of having identical column space), be included in the model. We encourage users to create indicator variables corresponding to the levels of the `reff(varname)` variable and to include these as regressors in the model, if they are not already part of the model. For example, in the setting where teachers of the same grade are being compared in the reference collections, the explanatory variables should contain dummy variables for each of the grades. In the case of a single reference collection, where all of the units are being compared to one another, the regression should include a constant term. If the user-specified explanatory variables do not span \mathbf{G} , `felsdvregdm` will add variables corresponding to \mathbf{G} to the explanatory variable list as well as to the dataset after estimation, where they are labeled as `Dvarname_1, \dots, Dvarname_G`, where `varname` is the name of the `reff()` variable in the function call.⁸

- Similarly to `felsdvreg`, `felsdvregdm` checks for collinearity among the columns of \mathbf{X}^* , the matrix of explanatory variables (augmented by \mathbf{G} , if necessary), and drops redundant variables if any exist. That is, \mathbf{X}^* (or $\widetilde{\mathbf{X}}^*$ in the two-level model) is restricted to a set of its linearly independent columns. `felsdvregdm` further checks that the columns $(\mathbf{X}^*, \mathbf{F}^*)$ [or $(\widetilde{\mathbf{X}}^*, \widetilde{\mathbf{F}}^*)$] are all linearly independent and aborts estimation with an error message if they are not because any collinearity among the columns would prevent the unambiguous interpretation of the fixed-effects estimates, defeating the purpose of our reparameterization.
- Like `felsdvreg`, `felsdvregdm` fits a two-level model, but unlike `felsdvreg`, it also provides the option of fitting a one-level fixed-effects model with the sum-to-zero constraint using the `onelevel` option.
- For the two-level model, `felsdvregdm` runs a grouping algorithm prior to estimation to determine groups, stores the results in the `group()` variable, and checks that no reference collection is ever associated with more than one group. If this test fails, the program aborts and returns an error message, because in this case, the sum-to-zero constraints within reference collection are insufficient to identify all the model parameters. In practice, we suggest that prior to attempting to fit a two-level model, users call `felsdvregdm` with the `grouponly` option. In addition, users should use the group information stored in the `group()` variable to check that reference collections do not contain multiple groups and as necessary redefine reference collections to be within group boundaries and remain substantively meaningful.
- Many options that were available in `felsdvreg` can no longer be implemented. Because automatic selection of parameterizations might not coincide with the substantive interests of users, the `takegroup` option of `felsdvreg` has been elimi-

8. On occasion, the user-specified \mathbf{G} will be replaced with variables that are equivalent to it. Stata developers are working to fix this inconsistency.

nated. Similarly, because the matrix of explanatory variables is augmented by \mathbf{G} , the reference collection fixed-effects matrix, the `cons` option is no longer possible.

- The F tests conducted by `felsdvregdm` are different from those conducted by `felsdvreg`. In particular, in `felsdvregdm` we made the decision to test the unit effects only within reference collections rather than testing the unit effects along with the reference collection means. That is, if there are G reference collections and J total unit effects, we include the G reference collection means in the reduced model and test the significance of the remaining $J - G$ centered unit effects. This choice was consistent with the notion that a user may choose a particular set of reference collections because between-collections variability may not be attributable to units. For example, it is common to compare teachers within grades, because test scores may vary across grades due to scaling of tests, which users generally would not want to include as part of the teacher effects.

For both the one-level and two-level models, `felsdvregdm` uses a model that includes \mathbf{X} for the first set of F tests and $\mathbf{X}^* = (\mathbf{X}, \mathbf{G})$ for the second set of F tests as the reduced model.⁹ For the one-level model, it tests the joint significance of the within-collection unit effects relative to the reduced model. For the two-level model, it reports three tests: the joint significance of the person and unit effects relative to the reduced model, the significance of the unit effects conditional on the person effects and the covariates, and the significance of the person effects conditional on the unit effects and the covariates.

- The `normalize` option in `felsdvreg` normalizes the group means of the unit effects to zero. This normalization is equivalent to the case in `felsdvregdm` if all the observations are specified to be in a single reference collection. However, in many situations, the user may wish to compare smaller collections of units, in which case `felsdvregdm` is more desirable. In addition, the `normalize` option does not adjust the standard errors for the normalization and continues to report standard errors that are relative to an arbitrary holdout teacher. For more information on the impact of using an arbitrary holdout teacher to calculate standard errors, see the discussion surrounding table 1 in section 4.1.
- Forthcoming in later versions of `felsdvregdm` are the `robust` and `cluster(varname)` options, which compute robust and clustered standard errors; the `noadji` and `noadjj` options, which adjust the degrees of freedom for the clustered standard errors; and the `hat(varlist)` and `orig(varlist)` options, which allow for two-stage least-squares estimation.

4 Example

We illustrate the use of `felsdvregdm` with an example of longitudinally linked student–teacher data from the Florida Education Data Warehouse. The data consist of a single

9. For the two-level model, time-invariant person covariates that were included among the list of independent variables are not included in the reduced model.

cohort of 9,283 students from an anonymous district followed from grades 4 to 8. In each grade, the students are linked to their mathematics teachers, and there are a total of 1,360 unique teacher-year units in the dataset.¹⁰ For the purposes of the example, we focus on estimating separate effects for each teacher for each school year. Any teacher who teaches students from the cohort in multiple years through changes in teaching assignment across years will have a distinct effect for each year they teach the cohort; hence, our units in this study are teacher-years rather than individual teachers. To simplify our presentation, we refer to our estimated effects as teacher effects although they are more precisely teacher-year effects for the small number of teachers who teach the cohort in multiple years. Because the data include only one cohort of students, there is an exact mapping of the grade level of the cohort and year. Teacher effects from any given year correspond to effects of all the teachers instructing the students in mathematics during one grade level. For instance, the year one teacher effects are the effects of the fourth grade mathematics teachers for the cohort, and the year two effects are the effects of the fifth grade mathematics teachers, and so on.

The outcome variable of interest is `nrtgain`, measured at each grade and equal to the student's gain (current year scale score minus prior year scale score) on a statewide mathematics exam. The data also include five time-invariant student-level indicators of basic demographics: `female`, `white`, `black`, `asian`, and `hispanic`. Finally, the data include the time-varying student-level covariate `nonstrucmove` indicating whether the student made a school move between years that was not also undertaken by a significant proportion of his prior-year schoolmates (i.e., not due to matriculation from elementary to middle school). This variable proxies for student family mobility that may be related to achievement.

In this section, we demonstrate the use of `felsdvregdm` to fit two different types of models for student mathematics achievement gains as a function of teacher and student inputs: a one-level fixed-effects model with teacher fixed effects and both time-invariant and time-varying student covariates, and a two-level model with both teacher and student fixed effects as well as the time-varying student covariate. For the one-level model, we show how inferences about the relationships of the student covariates to the outcome obtained by `felsdvregdm` are identical to those reported by `areg`. For the two-level model, we compare `felsdvregdm` to `felsdvreg`, showing that the estimates of covariates are identical but that the teacher effects from `felsdvregdm` have better properties. We also demonstrate how the issue of disconnected groups needs to be addressed in `felsdvregdm`.

Student gains can differ by grade in part because of the structure of the tests and general trends in student learning. We want to control for these grade-level differences in our model and avoid conflating them with our teacher effects. Consequently, for both the one-level and the two-level models, we include grade means in the model and we define reference collections for the teacher fixed effects at the level of individual grades; i.e., grade 4 teachers are one reference collection, grade 5 teachers are another, etc.,

10. We combined students from six teachers into two "teachers" to collapse three single-student strata into one stratum to improve the exposition of the example.

for a total of five reference collections. There are 412, 399, 178, 197, and 177 teachers in grades 4 through 8, respectively, with the smaller numbers in grades 6 through 8 reflecting the period structure of middle schools in which there are fewer teachers and more students per teacher. As noted above, because grades map uniquely to years and there is a distinct teacher effect for every teacher-year unit, reference collections defined by grades fully partition our teacher effects into disjoint sets. If we were interested in estimating a single effect for every teacher that is constant across years for teachers who teach the cohort in multiple years or grades, then reference collections defined by grade would not properly partition teachers and an alternative definition would be required.

4.1 One level of fixed effects

The model used in this example regresses gains in students' mathematics achievement, `nrtgain`, on teacher fixed effects and all student-level covariates. This is an instance of (1) with \mathbf{Y} being the vector of annual mathematics gain scores, \mathbf{X} including grade means and all student-level variables, and \mathbf{F} representing the teacher effects.

This model is fit using `felsdvregdm` as follows:

```
. felsdvregdm nrtgain female white black asian hispanic nonstrucmove grade4
> grade5 grade6 grade7 grade8, ivar(student) jvar(teachergrd) reff(grade1vl)
> feff(teachfe_dm) peff(stufe_dm) feffse(teachse_dm) mover(mover) group(group)
> xb(xb) res(resid) mnum(mnum) pobs(pobs) onelevel
```

```
Memory requirement for moment matrices in GB:
.01502616
```

```
Computing generalized inverse, dimension: 1365
Start: 9 Dec 2009 10:15:40
End: 9 Dec 2009 10:15:46
```

```
N=34326
```

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
female	.3093813	.2556106	1.21	0.226	-.1916247	.8103874
white	-.0732443	.8370731	-0.09	0.930	-1.713938	1.567449
black	.2331433	.8516812	0.27	0.784	-1.436182	1.902469
asian	3.080624	1.082118	2.85	0.004	.9596337	5.201615
hispanic	1.448387	1.031368	1.40	0.160	-.5731313	3.469905
nonstrucmove	.0243751	.3639665	0.07	0.947	-.6890122	.7377625
grade4	18.43987	.8813575	20.92	0.000	16.71238	20.16736
grade5	20.66716	.8959979	23.07	0.000	18.91097	22.42335
grade6	11.972	1.242489	9.64	0.000	9.536679	14.40733
grade7	15.04787	1.232244	12.21	0.000	12.63263	17.46312
grade8	19.07987	1.260358	15.14	0.000	16.60953	21.55022

```
F-test that firm effects are equal to zero: F(1354,32961)=2.558 Prob > F = .0031
```

The student identifiers are passed using `ivar()`, and the teacher identifiers are passed using `jvar()`.¹¹ The reference collection is denoted by `grade1v1`, which takes on values of 4 through 8 and is specified in `reff()`. The `felsdvregdm` call includes indicator variables corresponding to each grade level in the list of explanatory variables. The remaining options denote variables created to store the teacher and student fixed effects, the standard errors, and other variables of interest, similar to `felsdvreg`. The `onelevel` option was included to fit the one-level model that does not include student fixed effects. As discussed previously, if the reference collection indicator variables had not been included in the explanatory variable list, `felsdvregdm` would have added them and provided a message noting this addition and the names of the added variables. The fitted model would be identical to the one above in all other respects.

We find that student covariates generally are not related to student gains in mathematics achievement; only the `asian` variable is significant, but there are large differences in grade-level means. Comparing teachers across grade levels could conflate test differences and other factors with teacher effects. Hence, we need to compare teachers within grade level, as was accomplished via our choice of reference collection. The significant value for the F test indicates that within grade levels, there are meaningful differences among teachers, which is consistent with literature on teacher effects.

The Stata `areg` command is a commonly used alternative to `felsdvregdm` with the `onelevel` option for fitting the one-level (1). For example, `areg nrtgain female white black asian hispanic nonstrucmove grade4 grade5 grade6 grade7 grade8, absorb(teachergrd)` is an alternative command for fitting the model above. The `predict` postestimation command with the `d` or `xbd` option will recover the teacher effects without standard errors. Another alternative for the one-level model is the `fese` command, which provides estimates for both the teacher effects and the standard errors (Nichols 2008).¹²

`felsdvregdm` and `areg` (or `fese`) produce similar results but with some subtle, important differences. Both models produce the same estimates and standard errors for the student-level variables (`female`, `white`, `black`, `asian`, `hispanic`, and `nonstrucmove`) and both models give the same overall model fit, test for the overall model, residuals, and root mean squared error. However, because the grade indicators (`grade4–grade8`) are collinear with the absorbed teacher effects, `areg` drops these variables from the model and makes them part of the estimated teacher effects. In particular, `predict`'s `d` option yields teacher-effect estimates that equal the mean for each teacher, adjusted for the student-level covariates, less the grand mean adjusted for student-level covariates. `predict`'s `xbd` option yields teacher-effect estimates that equal the mean for each teacher, adjusted for the student-level covariates. The estimates make no adjustment

11. Even though this model contains one level of fixed effects, both the `ivar()` and the `jvar()` variables need to be specified. This is because even with one fixed-effects model, the memory saving computation of the fixed-effects parameters in `felsdvregdm` is possible by taking advantage of the panel structure of the dataset.

12. Similarly, `xtreg`, `fe` could be used to estimate the coefficients on the time-varying and time-invariant variables and has options to recover the teacher effects with the `predict` postestimation command using the `u` or `xbu` option. Again, `xtreg` does not provide the control over that parameterization that `felsdvregdm` offers, and standard errors are difficult to estimate in this case.

for the grade-level means and are not parameterized to sum to zero overall or within grade level. As a result, the teacher-effect estimates produced by `areg` with `predict's d` option differ from the estimates from `felsdvregdm` within each grade level by a constant equal to the adjusted mean of the outcome variables for the entire sample minus the grade-level mean of the adjusted teacher means (table 1). `fese` produces the same estimates as `areg` with `predict's xbd` option, and both are estimates of the teacher means assuming the grade-level means and the overall mean are part of the parameter of interest. These are unbiased estimates of the means given the model, but they may not truly be the parameters of interest and would be inappropriate for empirical Bayes (shrinkage) estimation.

Table 1. Mean differences in teacher fixed-effects estimates, by reference collection, produced by `areg` and `felsdvreg`.

Grade	Difference
4	-1.966
5	-4.193
6	4.502
7	1.426
8	-2.606

Whether the grade-level means should be part of the teacher effects cannot be determined by the data and must be determined by the user. `felsdvregdm` provides users with the option of treating the grade-level mean differences as an external source of error and estimating the teacher effects relative to the mean of their grade-level peers. `areg` and `fese` provide the alternative approach. It would be possible to post hoc center the estimate from `areg` or `fese` by grade level, but the standard errors would not be provided as they are with `felsdvregdm`.

4.2 Two levels of fixed effects

The model used in this example regresses gains in students' mathematics achievement, `nrtgain`, on teacher fixed effects, student fixed effects, and the time-varying student covariate `nonstrucmove`. Compared with the previous example, this model uses fixed effects for each individual student rather than time-invariant student covariates, to control for student heterogeneity. This is an instance of (2). Because the model includes two levels of fixed effects, we compare `felsdvregdm` with `felsdvreg` rather than `areg`.

As discussed previously, issues of grouping must be considered in the two-level model. Because of `felsdvregdm`'s goal to maintain strict control over the interpretation of the unit fixed effects, reference collections must not attempt to compare teachers in disconnected strata. To check the grouping structure, we first run `felsdvregdm` with the `grouponly` option:

```
. felsdsvregdm nrtgain nonstrucmove grade4 grade5 grade6 grade7 grade8,
> ivar(student) jvar(teachergrd) reff(gradelvl) feff(teachfe) peff(stufe)
> feffse(teachse) mover(mover) group(groupout) xb(xb) res(resid) mnum(mnum)
> pobs(pobs) grouponly
Note: You specified `grouponly`. Only the group variable was modified and saved.
      No estimates were produced.
```

Groups of firms connected by worker mobility:

	Person-years	Persons	Movers	Firms
groupout	N(__000000)	N(__000009)	sum(__00000D)	N(__000008)
1	34,320	9,280	9280	1,357
2	6	3	3	2
Total	34,326	9,283	9283	1,359

The results of the grouping algorithm are stored in `groupout`. These data contain two groups: one with the vast majority of the observations and one with three students and two teachers. Given the groups and the reference collections, we modify the reference collections to have the five grade-specific collections in the large group and one collection for both teachers in the small group.

The output below shows the composition of the six reference groups and the results from fitting the model using `felsdsvregdm` with the new reference collections:

```
. egen newreff = group(gradelvl) if groupout==1
(6 missing values generated)
. replace newreff = 6 if groupout==2
(6 real changes made)
. preserve
. collapse groupout (sum) freq, by(newreff gradelvl)
. order newreff
. list, noobs
```

newreff	gradelvl	groupout	freq
1	4	1	6940
2	5	1	7622
3	6	1	6991
4	7	1	6894
5	8	1	5873
6	7	2	3
6	8	2	3

```
. restore
. quietly xi i.newreff, noomit
```



```

. felsdvregdm nrtgain nonstrucmove female white black asian hispanic
> _Inewreff_1 _Inewreff_2 _Inewreff_3 _Inewreff_4 _Inewreff_5 _Inewreff_6,
> ivar(student) jvar(teachergrd) reff(newreff) feff(teachfe) peff(stufe)
> feffse(teachse) mover(mover) group(groupout) xb(xb) res(resid) mnum(mnum)
> pobs(pobs)
The group variable (called groupout) that exists in the dataset will be replaced!
note: _Inewreff_6 dropped because of collinearity

Some variables were added or dropped from the independent variables
Reference collection fixed effects may not be sufficiently defined,
Variables could be collinear due to multiple groups in the dataset,
Or some variables may be collinear with the second level of fixed effects
The following variable(s) were added to the explanatory variables

The following variable(s) were ignored or dropped from the explanatory
variables
  Dnewreff_6 Dnewreff_5 Dnewreff_4 Dnewreff_3 Dnewreff_2 Dnewreff_1 _Inewreff_5
> hispanic asian black white female

Memory requirement for moment matrices in GB:
  .01489488

Computing generalized inverse, dimension: 1358
  Start: 9 Dec 2009 10:16:18
  End: 9 Dec 2009 10:16:24

N=34326

```

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
nonstrucmove	.346897	.4944743	0.70	0.483	-.6223043	1.316098
_Inewreff_1	-.0394583	1.46098	-0.03	0.978	-2.903073	2.824156
_Inewreff_2	2.52157	1.498037	1.68	0.092	-.4146783	5.457819
_Inewreff_3	-6.279251	1.882227	-3.34	0.001	-9.968537	-2.589965
_Inewreff_4	-4.169235	1.853738	-2.25	0.025	-7.80268	-.5357898

```

F-test that person and firm effects are zero: F(10635,23685)=.601 Prob > F =.7295
F-test that person effects are equal to zero: F(9281,23685)=.38 Prob > F = 1
F-test that firm effects are equal to zero: F(1353,23685)=2.196 Prob > F = 0

```

In general, in a two-level model, one reference-collection indicator will be dropped per group because of collinearity between the set of student fixed effects for students in each group and the set of reference collection means in each group (both sets add to an indicator for the group). In this example, the indicator for reference collection 6 is dropped for the small group and the indicator for reference collection 5 is dropped for the large group. However, the teacher effects are identified in both of the groups.

The crucial difference between `felsdvregdm` and other fixed-effects routines is in the estimated teacher effects. To demonstrate this, we calculated the mean estimated teacher effect, as well as the mean standard error of these estimates, by reference collection for both `felsdvregdm` and `felsdvreg`. We did this for two different datasets: the one used in the example reported above, in which the teacher with the lowest ID number has a large positive effect and many linked students, and an alternative version in which the teacher ID numbers were deliberately changed so that the teacher with the lowest ID number has a large negative effect and few linked students. These purposeful

manipulations of the teacher ID numbers are designed to demonstrate the impacts on both the estimates and their standard errors of changing the holdout teacher used by `felsdvreg`. The results are given in table 2.

Table 2. Mean estimated teacher effects by reference collection, with mean standard errors of these estimates by reference collection in parentheses. The number following the procedure name refers to the dataset used for estimation, where only the holdout teacher identifier was changed.

Reff Coll	felsdvregdm 1	felsdvregdm 2	felsdvreg 1	felsdvreg 2
1	0.000 (8.216)	0.000 (8.216)	-41.844 (11.926)	74.502 (37.146)
2	0.000 (8.649)	0.000 (8.649)	-39.283 (12.447)	77.063 (37.499)
3	0.000 (13.716)	0.000 (13.716)	-48.084 (17.222)	68.262 (40.058)
4	0.000 (14.703)	0.000 (14.703)	-45.974 (18.092)	70.372 (40.615)
5	0.000 (13.697)	0.000 (13.697)	-41.805 (17.198)	74.542 (40.065)
6	0.000 (10.385)	0.000 (10.385)	-13.667 (20.770)	-13.667 (20.770)

As intended, the mean teacher effect by reference collection for `felsdvregdm` is zero, regardless of how the teacher identifiers are labeled. For `felsdvreg`, the mean teacher effect is strongly negative for each reference collection in the first dataset, where the holdout teacher was above average, and strongly positive in the second dataset, where the holdout teacher was well below average. The estimated teacher effects from the two routines are perfectly correlated within reference collection (not shown); however, the parameterization used by `felsdvregdm` removes the instability caused by changing the holdout. An additional difference is that the fixed effect is not estimated for the holdout teacher in `felsdvreg`, and the standard error for this teacher is missing. `felsdvregdm` reports fixed effects and standard errors for all the teachers.

Because the teacher effects estimated by `felsdvregdm` are invariant to the labeling of teacher identifiers, the standard errors for those effects reported by `felsdvregdm` are also invariant to the labeling of the teacher identifiers. The standard errors of the effects reported by `felsdvreg` vary dramatically depending on the holdout teacher, with the average standard errors for the second dataset being much larger because the holdout teacher had very few students. Both sets of standard errors are larger than those reported by `felsdvregdm` because they are for contrasts of each teacher relative to the holdout instead of relative to the group mean, the former being a less precise comparison.

It would be easy to recover the desired mean-centered teacher effects reported by `felsdvreg` by post hoc centering; however, the standard errors of those contrasts cannot easily be obtained because they are a function of the full design matrix and not simple functions of the standard errors reported by `felsdvreg`. Having the appropriate standard errors is essential to many post hoc inferences about teachers made from the model. For example, in education applications, it is commonly of interest to estimate whether each teacher's effect is statistically significantly different from its

reference collection mean. These hypothesis tests are a straightforward function of the estimates and standard errors provided by `felsdvregdm` but are not easily conducted with the output from `felsdvreg`. In addition, it is often of interest to use the estimates and standard errors to estimate the true between-teacher variability in effectiveness within reference groups and then subsequently use this estimated variance to conduct empirical Bayes (shrinkage) estimators (Boyd et al. 2008; Koedel and Betts 2005; Jacob and Lefgren 2008; McCaffrey, Han, and Lockwood 2008). A simple method of moments estimator of this variance component is obtained by subtracting the average squared standard error of the estimates from the sample variance of estimates. For example, the average method of moments estimate of the between-teachers variance across reference collections obtained by the `felsdvregdm` output is 0.03, which is roughly consistent with values commonly reported in the teacher effectiveness literature.¹³ Applying the same computations to the two different `felsdvreg` specifications yields infeasible values of -0.02 and -0.78 , since the sample variance of the estimated effects is biased because of the correlation resulting from contrasting every estimated effect to the same holdout. Therefore, estimates and standard errors provided by `felsdvregdm` are appropriate for these computations, but those provided by `felsdvreg` are not.

5 Conclusion

As linked longitudinal datasets from various fields become more available, the desire to analyze these data to estimate the effects of individual units is poised to increase. These models can be computationally challenging, particularly when multiple levels of fixed effects are specified, and existing software routines for dealing with these challenges are generally not designed to provide estimates that are useful for inferences about individual fixed effects. Arbitrary decisions about how to deal with the lack of full identification of the fixed effects is at odds with the desire for unambiguous inferences about those effects. Sum-to-zero constrained effects within reference collections as implemented in `felsdvregdm` is an intuitive way for analysts to specifically control the parameterization of fixed effects when those effects are of interest and should prove valuable in many applications.

Although we implemented sum-to-zero constrained effects within reference collections in the context of longitudinal data, the logic of the parameterization applies to other types of models and data structures that are not addressed by `felsdvregdm`. For example, in education research, some analysts fit models where current year test scores or gain scores are modeled as a function of current teacher indicators and other student-level covariates such as demographic variables and prior test scores (McCaffrey, Han, and Lockwood 2008; Rothstein 2007). That is, rather than treating the data as panel data with the sequence of outcomes all being on the left-hand side of the model equation, these models fit the regression of the current outcome on past out-

13. Let $\widehat{\phi}_{gi}$ and se_{gi} equal the estimated effect and standard error of that effect for teacher $i = 1, \dots, N_g$ of grade $g = 4, \dots, 8$. The method of moment estimator for the between-teachers variance for teachers in grade g is $v_g = \sum_i (\widehat{\phi}_{gi} - \widehat{\phi}_g)^2 / (N_g - 1) - \sum_i se_{gi}^2 / N_g$, where $\widehat{\phi}_g$ equals the average of the estimates for grade g . We report the average of v_g across grade levels.

comes and other predictors. Our parameterization is equally desirable to this type of model, but `felsdvregdm` is designed for data with separate records for each person and time point rather than for data with one record per person and observations from prior time points included as additional fields instead of additional records. Future work to implement sum-to-zero constrained effects within reference collections in the context of cross-sectional models or other applications with similar data structures would be valuable.

6 References

- Aaronson, D., L. Barrow, and W. Sander. 2007. Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics* 25: 95–135.
- Abowd, J. M., R. H. Creedy, and F. Kramarz. 2002. Computing person and firm effects using linked longitudinal employer–employee data. Technical Paper No. TP-2002-06, Center for Economic Studies, U.S. Census Bureau.
<http://lehd.did.census.gov/led/library/techpapers/tp-2002-06.pdf>.
- Abowd, J. M., F. Kramarz, and D. N. Margolis. 1999. High wage workers and high wage firms. *Econometrica* 67: 251–333.
- 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.
- Andrews, M., T. Schank, and R. Upward. 2006. Practical fixed-effects estimation methods for the three-way error-components model. *Stata Journal* 6: 461–481.
- Boyd, D., P. Grossman, H. Lankford, S. Loeb, and J. Wyckoff. 2008. Who leaves? Teacher attrition and student achievement. NBER Working Paper No. 14022, National Bureau of Economic Research. <http://www.nber.org/papers/w14022>.
- Clotfelter, C., H. Ladd, and J. Vigdor. 2006. Teacher–student matching and the assessment of teacher effectiveness. NBER Working Paper No. 11936, National Bureau of Economic Research. <http://www.nber.org/papers/w11936>.
- 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., and M. Hansen. 2008. Is it just a bad class? Assessing the stability of measured teacher performance. CPRE Working Paper #2008-5, Center on Reinventing Public Education. http://www.crpe.org/cs/crpe/view/csr_pubs/249.
- Gordon, R., T. J. Kane, and D. O. Staiger. 2006. Identifying effective teachers using performance on the job. Discussion Paper 2006-01, The Brookings Institution.
http://www.brookings.edu/papers/2006/04education_gordon.aspx.
- Greene, W. H. 2008. *Econometric Analysis*. 6th ed. Upper Saddle River, NJ: Prentice Hall.

- Guimarães, P., and P. Portugal. 2009. A simple feasible alternative procedure to estimate models with high-dimensional fixed effects. IZA Discussion Paper No. 3935, Institute for the Study of Labor (IZA). <http://ideas.repec.org/p/iza/izadps/dp3935.html>.
- Harris, D. N., and T. R. Sass. 2006. Value-added models and the measurement of teacher quality. Unpublished manuscript.
- . 2007. What makes for a good teacher and who can tell? Manuscript, Florida State University, Tallahassee.
- 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.
- Kane, T. J., J. E. Rockoff, and D. O. Staiger. 2006. What does certification tell us about teacher effectiveness? Evidence from New York City. Unpublished manuscript.
- Koedel, C., and J. Betts. 2005. Re-examining the role of teacher quality in the educational production function. Working Paper 0708, Department of Economics, University of Missouri. <http://ideas.repec.org/p/umc/wpaper/0708.html>.
- McCaffrey, D., B. Han, and J. R. Lockwood. 2008. From data to bonuses: A case study of the issues related to awarding teachers pay on the basis of their students' progress. NCPI Working Paper No. 2008-14, National Center for Performance Incentives.
- McCaffrey, D. F., J. R. Lockwood, D. Koretz, T. A. Louis, and L. Hamilton. 2004. Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics* 29: 67–101.
- McCaffrey, D. F., T. R. Sass, J. R. Lockwood, and K. Mihaly. 2009. The intertemporal variability of teacher effect estimates. *Education Finance and Policy* 4: 572–606.
- Menezes-Filho, N. A., M.-A. Muendler, and G. Ramey. 2008. The structure of worker compensation in Brazil, with a comparison to France and the United States. *Review of Economics and Statistics* 90: 324–346.
- 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>.
- Rockoff, J. E. 2004. The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review: Papers and Proceedings* 94: 247–252.
- Rothstein, J. 2007. Do value-added models add value? Tracking, fixed effects, and causal inference. Working Paper 1036, Center for Economic Policy Studies, Princeton University. <http://ideas.repec.org/p/pri/cepsud/1036.html>.
- SAS Institute, Inc. 2004. *SAS 9.1.3 Help and Documentation*. Cary, NC: SAS Institute.
- Venables, W. N., and B. D. Ripley. 2001. *Modern Applied Statistics with S-plus*. 3rd ed. New York: Springer.

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

About the authors

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.

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.

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 with the National Center on Performance Incentives to study the effects of using value-added measures to determine teacher compensation in two randomized trials. He is also 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.

Tim Sass is a professor of economics at Florida 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.