AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

Check for updates

The Stata Journal (2017)
17, Number 1, pp. 89–115

Within- and between-cluster effects in generalized linear mixed models: A discussion of approaches and the xthybrid command

Reinhard Schunck
GESIS—Leibniz–Institute for the Social Sciences
Cologne, Germany
reinhard.schunck@gesis.de

Francisco Perales
Institute for Social Science Research
University of Queensland
Brisbane, Australia
f.perales@uq.edu.au

Abstract. One typically analyzes clustered data using random- or fixed-effects models. Fixed-effects models allow consistent estimation of the effects of level-one variables, even if there is unobserved heterogeneity at level two. However, these models cannot estimate the effects of level-two variables. Hybrid and correlated random-effects models are flexible modeling specifications that separate within- and between-cluster effects and allow for both consistent estimation of level-one effects and inclusion of level-two variables. In this article, we elaborate on the separation of within- and between-cluster effects in generalized linear mixed models. These models present a unifying framework for an entire class of models whose response variables follow a distribution from the exponential family (for example, linear, logit, probit, ordered probit and logit, Poisson, and negative binomial models). We introduce the user-written command `xthybrid`, a shell for the `meglm` command. `xthybrid` can fit a variety of hybrid and correlated random-effects models.

Keywords: st0468, xthybrid, correlated random effects, fixed effects, generalized linear mixed models, hybrid model, meglm, Mundlak model, random effects

1 Introduction

Researchers undertaking multilevel and panel analysis of hierarchically clustered data often face a difficult decision between random- and fixed-effects models. Random-effects models allow researchers to estimate the effect of cluster-invariant variables (that is, level-two variables) on the outcome variable but impose the assumption that the random effects (for example, the level-two error) are uncorrelated with the observed covariates. If this assumption is violated, the model coefficients are biased. Fixed-effects models, on the other hand, do not require this assumption—and can provide unbiased estimates of the level-one variables, even if there is unobserved heterogeneity

© 2017 StataCorp LLC

at the cluster level. However, fixed-effects model estimation relies only on within-cluster variation in the explanatory and outcomes variables; thus these models cannot provide effect estimates for the level-two variables.[1]

More flexible modeling specifications provide fixed-effects estimates (or estimates that are close to these) for level-one variables and allow inclusion of level-two variables, most notably the hybrid (Allison 2009) and correlated random-effects models (Wooldridge 2010). The latter is also known as the Mundlak model (Baltagi 2006; Mundlak 1978). These estimation strategies differentiate within- and between-cluster effects and combine the strengths of random- and fixed-effects models. In the linear case, they yield estimates of the level-one covariates that are unbiased by cluster-level unobserved heterogeneity, while allowing for level-two cluster-invariant covariates (Allison 2009; Mundlak 1978; Neuhaus and Kalbfleisch 1998; Rabe-Hesketh and Skrondal 2012; Raudenbush 1989; Schunck 2013; Snijders and Berkhof 2008).

Schunck (2013) described using these models for continuous outcome variables, providing a theoretical overview and a practical application in Stata. This article discusses the applicability of hybrid and correlated random-effects models within the umbrella of generalized linear mixed models (GLMM) (Brumback et al. 2010). In doing so, we show how the decomposition of within- and between-cluster effects can be extended to GLMM, which comprise popular models for binary, ordered, and count outcomes (Neuhaus and Kalbfleisch 1998; Neuhaus and McCulloch 2006; Brumback et al. 2010). Importantly, such decomposition can approximate fixed-effects estimates for specifications in which a fixed-effects estimator is not available or implemented (Neuhaus and McCulloch 2006).

In the remainder of the article, we first elaborate on the separation of within- and between-cluster effects in a GLMM framework, then present a user-written command, `xthybrid`, that builds on Stata's `meglm` command and can fit hybrid and correlated random-effects models.

1. The terms "fixed effects" and "random effects" are not used consistently across disciplines and literature. In the multilevel model literature, the term "fixed effects" denotes a model's regression coefficients, whereas the term "random effects" refers to a model's random intercepts and slopes. In this article, random-effects models refer to models for clustered data that have both random effects and fixed effects (also known as multilevel models, hierarchical models, and mixed models). In this context, a fixed-effects model refers to a model that includes only fixed effects, which is typically a pooled or cross-sectional model that does not consider that the data may be clustered. In econometric literature, however, the term "fixed effects model" refers to a model for clustered data that allows for arbitrary dependence between the unobserved effects and the covariates (Wooldridge 2010, 286). The name "fixed-effects model" emerged because these models treat the unobserved cluster-level effects as fixed rather than random (McCulloch, Searle, and Neuhaus 2008). Whether these effects are random or nonrandom is not of concern to us. Modern econometrics assumes they are random (Wooldridge 2010, 286). In this article, we adopt the econometric terminology for fixed-effects models.

2 GLMM

Generalized linear models (GLM) constitute a unifying framework for an entire class of models whose response variables follow a distribution from the exponential family. This includes many popular models such as the standard linear model, models for binary responses (for example, logit and probit models), models for ordinal responses (for example, ordered probit and logit models), and models for count responses (for example, Poisson and negative binomial models). In this section, we will provide a short overview of GLM and GLMM (for details, see McCulloch, Searle, and Neuhaus [2008], Skrondal and Rabe-Hesketh [2003], and Rabe-Hesketh and Skrondal [2012]).

Let μ_j be the expected response of y_j given the covariates $\{\mu_j = E(y_j|x_j)\}$. Then, a GLM with y_j as the response variable and x_j as a covariate is defined as

$$g\{E(y_j|x_j)\} = g(\mu_j) = \beta x_j$$

$g(\cdot)$ is the so-called link function, which transforms the mean μ_j so that it can be linearly related to the predictors. The link function therefore defines the functional relationship between the predictors and the response variable (McCullagh and Nelder 1989; McCulloch, Searle, and Neuhaus 2008; Rabe-Hesketh, Skrondal, and Pickles 2004; Skrondal and Rabe-Hesketh 2003). Specifying a GLM also requires choosing a conditional distribution for the response variable from the exponential family of distributions. Different permutations of link functions and distributions result in different models (see table 1).

GLMs can be extended to include random effects and are thus suited for analyzing clustered data, such as multilevel and panel data. These models are known as generalized linear mixed models (GLMM). Consider a situation where we have data with two hierarchical levels. Let i denote level two (for example, schools) and j denote level one (for example, students). y_{ij} is the response (dependent) variable, x_{ij} is a level-one variable that varies within and between clusters, c_i is a level-two variable that varies only between clusters, and u_i is the random intercept. A GLMM is specified as

$$g\{E(y_{ij}|x_{ij}, c_i, u_i)\} = g(\mu_{ij}) = \beta x_{ij} + \gamma c_i + u_i \tag{1}$$

The "mixing" outlined above becomes obvious: this model "mixes" a fixed part (the fixed coefficients β and γ) and a random part (the random intercept u_i). To relax the assumption that the effects of level-one covariates are the same across all clusters, we can include random slopes as follows:

$$g(\mu_{ij}) = (\beta + u_{i2})x_{ij} + \gamma c_i + u_{i1}$$

In Stata, the available link functions for GLMM comprise identity, logit, probit, log, and complementary log-log. The available distributions from the exponential family of distributions comprise the normal (Gaussian), Bernoulli, binomial, gamma, negative binomial, ordinal, and Poisson distributions (see table 1). Other link functions and distributions are theoretically possible.

Table 1. Possible combinations of link functions and distributions in `xthybrid`

	Link function				
Distribution	Identity	Log	Logit	Probit	Cloglog
Gaussian	x	x			
Bernoulli			x	x	x
Binomial			x	x	x
Gamma		x			
Negative binomial		x			
Ordinal			x	x	x
Poisson		x			

Using the identity link $g(\mu_{ij}) = \mu_{ij}$ and the Gaussian distribution for y_{ij} yields a linear random-intercept model,

$$\mu_{ij} = \beta x_{ij} + \gamma c_i + u_i \tag{2}$$

where the conditional distribution of y_{ij} is $y_{ij}|x_{ij}, c_i, u_i \sim N(\mu_{ij}, \sigma^2)$. If the outcome variable is binary, the expected value for y_{ij} is the probability that $y_{ij} = 1$; that is, $\mu_{ij} = \Pr(y_{ij} = 1|x_{ij}, c_i, u_i)$. Combining a Bernoulli or a binomial distribution for the response variable with a probit link results in the random-intercept probit model

$$\Phi^{-1}(\mu_{ij}) = \beta x_{ij} + \gamma c_i + u_i$$

where $\Phi(\cdot)^{-1}$ is the inverse function of the standard normal cumulative distribution. Here the conditional distribution of y_{ij} is $y_{ij}|x_{ij}, c_i, u_i \sim B(1, \pi_{ij})$. We specify a random-effects logit model by choosing the logit link:

$$\text{logit}(\mu_{ij}) = \beta x_{ij} + \gamma c_i + u_i$$

This rationale can extend to other, more complex models, for example, models for ordered and count outcomes (for an overview, see McCulloch, Searle, and Neuhaus [2008]; Skrondal and Rabe-Hesketh [2003, 2004]).

3 Level-two confounders (unobserved heterogeneity)

The standard assumptions in GLMM (that is, in multilevel models) are that i) the level-two error is a normally distributed random variable—that is, $u_i \sim N(0, \sigma^2_{u_i})$; and ii) the level-two error is uncorrelated with the covariates—that is, $E(u_i|x_{ij}, c_i) = 0$. The latter assumption is of particular importance and means that there are no omitted level-two confounders; that is, there is no unobserved heterogeneity at level two. If the level-two error is correlated with the covariates so that $E(u_i|x_{ij}, c_i) \neq 0$, then the effect estimates in a random-effects model, for example, as in (1), will be biased. This situation emerges if we omit a confounding variable at level two from the model.

Alternatively, we can treat unobserved level-two effects as fixed effects. In linear models, a simple way of accomplishing this is to directly estimate the fixed effects by including a dummy variable for each of the clusters:

$$y_{ij} = \beta_{\text{LSDV}} x_{ij} + \sum_{i=1}^{i} \beta_i k_i + \epsilon_{ij} \tag{3}$$

This model is the least-squares dummy variable (LSDV) estimator (Wooldridge 2010). It fits i intercepts—one for each cluster—represented by the variables k_i. This approach provides consistent estimation of the level-one covariates without the assumptions of the random-effects model that the cluster-specific intercepts are random variables and uncorrelated with the covariates. Instead, they are explicitly included in the model and estimated as fixed effects. Thus the estimated level-one effects will be unbiased by level-two unobserved heterogeneity, because there is none anymore. Note that estimates may still be biased because of unobserved heterogeneity at level one. The models still assume that $E(\epsilon_{ij}|x_{ij}, k_i) = 0$. ϵ_{ij} is the level-one error, which we will treat as white noise for the remainder of this article.

A disadvantage of the LSDV model is that we cannot retrieve the effect of level-two variables (c_i). Because the model incorporates dummy variables capturing the overall cluster effects, we cannot identify the effects of cluster-level covariates because of collinearity. Additionally, estimating each cluster effect is impractical when there is a large number of clusters. Furthermore, using maximum likelihood to fit these models leads to inconsistent parameter estimates. This is because of the incidental parameters problem; the number of fixed effects parameters (or "nuisance parameters") increases with sample size, which leads to inconsistent estimates when using maximum likelihood estimation (Andersen 1970; Chamberlain 1980; Wooldridge 2010). This makes it infeasible to use this approach for models estimated using maximum likelihood—including GLMM.

An alternative approach to fitting a fixed-effects (FE) model in the linear case is demeaning the explanatory and outcome variables. We do this by subtracting the between model

$$\overline{y}_i = \beta \overline{x}_i + \gamma c_i + u_i + \overline{\epsilon}_i$$

from the random-effects model

$$y_{ij} = \beta x_{ij} + \gamma c_i + u_i + \epsilon_{ij} \tag{4}$$

Note that (4) is equivalent to (2.2). Because $\overline{c}_i = c_i$ and $\overline{u}_i = u_i$, the subtraction leads to

$$(y_{ij} - \overline{y}_i) = \beta_{\text{FE}}(x_{ij} - \overline{x}_i) + (\epsilon_{ij} - \overline{\epsilon}_i) \tag{5}$$

This technique, also called the "within transformation", averages out all elements in (5) that do not vary within clusters, including the level-two error term, u_i. Thus this FE model does not require any assumptions on the distribution of u_i or its correlation with the covariates. The estimated effects of β in the LSDV model in (4) and the

FE model in (5) are identical ($\beta_{\text{LSDV}} = \beta_{\text{FE}}$).[2] Thus the fixed-effects model in its demeaned form (5) also provides estimates of level-one covariates, which are unbiased by unobserved heterogeneity at level two. Because any level-two characteristic—observed or unobserved—is removed from the equation, we cannot retrieve the effects of level-two covariates, just as in the LSDV model (4).

The above refers only to the linear case. For some GLMM that are estimated with maximum likelihood (for example, logit models), we can compute a similar fixed-effects estimator using a conditional likelihood approach (Chamberlain 1980; McCulloch, Searle, and Neuhaus 2008; Wooldridge 2010). The conditional likelihood approach uses a sufficient statistic to remove the level-two error from the equation (Chamberlain 1980; McCulloch, Searle, and Neuhaus 2008; Wooldridge 2010). The conditional likelihood approach can thus forgo any assumptions on the level-two error. Therefore, conditional likelihood models are also referred to as fixed-effects models. However, conditional likelihood approaches are unavailable for most GLMM, so this approach is not always a viable alternative (McCulloch, Searle, and Neuhaus 2008; Wooldridge 2010).

Altogether, fixed-effects models provide less biased estimates of the level-one covariates than random-effects models, but unlike random-effects models, they fail to retrieve the effect estimates of level-two variables. Depending on the nature of the research question, disregarding between-cluster variation can even be seen as an advantage. For example, this is sometimes regarded as focusing on the informative cases (Halaby 2004; Wooldridge 2010, 621f.). This holds in particular for longitudinal research investigating how change in an explanatory variable, x_{ij}, is associated with change in an outcome variable y_{ij}. However, failure to retrieve the effect estimates of level-two variables is a major problem in multilevel analysis, where the interest often lies in these effects, for example, how the characteristics of neighborhoods, schools, workplaces, or geographical areas influence individuals' outcomes (Sampson 2003; Sampson, Raudenbush, and Earls 1997). Because the fixed-effects approach discards all contextual (level-two) information, some argue that it is generally less preferable than the random-effects approach for multilevel analysis (Bell and Jones 2015).

4 Within- and between-decomposition

An alternative to both random and fixed-effects models within the framework of GLMM is models that separate within- and between-cluster effects (Neuhaus and McCulloch 2006; Schunck 2013), such as the hybrid model (Allison 2009) and the related correlated random-effects model (Cameron and Trivedi 2005; Wooldridge 2010).

The hybrid model (Allison 2009) splits within- and between-cluster effects for the level-one covariates:

$$g(\mu_{ij}) = \beta_W(x_{ij} - \overline{x}_i) + \beta_B\overline{x}_i + \gamma c_i + u_i \tag{6}$$

2. To be precise, both (4) and (5) are fixed-effects models. Thus we could label both estimators of β with the subscript FE.

This is accomplished by including both the deviation from the cluster-specific mean $(x_{ij} - \overline{x}_i)$ and the cluster-specific mean $\overline{x}_i$ among the model covariates. β_W gives the within-cluster effect, and β_B gives the between-cluster effect.

The correlated random-effects model (Wooldridge 2010), sometimes called the Mundlak (1978) model, is mathematically equivalent to the hybrid model. However, in contrast to (6), it includes the level-one variable (x_{ij}) in its undemeaned form:

$$g(\mu_{ij}) = \beta_W x_{ij} + \tau \overline{x}_i + \gamma c_i + v_i \tag{7}$$

In contrast to the standard random-intercept model, it introduces the assumption that the level-two error $u_i = \tau \overline{x}_i + v_i$ and $v_i \sim N(0, \sigma^2_{v_i})$. This means that the level-two error can depend on x_{ij} through its cluster means. The inclusion of $\overline{x}_i$ picks up any correlation between this variable and the unobserved random effect. In this model, $\tau = \beta_B - \beta_W$. The relationship between the two model specifications is apparent when rewriting (6) as

$$g(\mu_{ij}) = \beta_W x_{ij} + (\beta_B - \beta_W)\overline{x}_i + \gamma c_i + u_i$$

4.1 Within-cluster and between-cluster effects

The basic idea behind the hybrid and correlated random-effects models is to restrict the dependency between u_i and the level-one covariates. We can accomplish this by assuming that u_i depends on the mean values of the level-one covariates $(\overline{x}_i)$. We estimate the within-cluster effect, β_W, using only within-cluster variation. It assesses how on average a within-cluster change in x_{ij} is associated with a within-cluster change in y_{ij}. In the linear case (that is, with the identity link and the Gaussian distribution), the within-cluster effects fit by both the hybrid and correlated random-effects models are identical to the fixed-effects estimates, so that $\beta_W = \beta_{\text{FE}}$ (Goetgeluk and Vansteelandt 2008; Hsiao 2003; Mundlak 1978).

The between-cluster effect, β_B, assesses how a change in $\overline{x}_i$ is associated with a change in $\overline{y}_i$. It is estimated using only between-cluster variation. Different research traditions have different views as to whether these effects are informative. In multilevel research, the interest often lies on the level-two variables, and between-cluster effects of the level-one variables are level-two effects. The substantive interpretation of the cluster means of the level-one variables is different from the substantive interpretation of the level-one variables (Snijders and Berkhof 2008, 146). However, one should keep in mind that if the random-effects assumption is violated (that is, there is unobserved heterogeneity at level two), the between-cluster effect is biased. In panel-data analysis, the interest lies chiefly on the within-cluster effects, so it is imperative to obtain estimates of these effects, that are robust to unobserved heterogeneity at level two (Allison 2009; Halaby 2004; Wooldridge 2010). From this perspective, it is questionable whether the between-cluster effects are of substantial interest at all.

Regardless of whether one is interested in interpreting the between-cluster effect, one should include the means of the level-one variables as controls for the other level-two variables. If one includes other level-two variables, for example, c_i, without controlling

for $\overline{x}_i$, then the estimated effect of c_i is not adjusted for between-cluster differences in x_{ij}. Note that consistent estimation of the effects of level-two variables still rests on the assumption that there is no correlation between these and the level-two error $\{E(u_i|c_i) = 0\}$.

In both multilevel and panel-data models, it is helpful to compare between- and within-cluster effects for pragmatic reasons (Allison 2009; Schunck 2013). The random-effects model (1) assumes that both effects are the same and uses a weighted average of within- and between-cluster variation in estimation. However, if this assumption does not hold, the estimates of the random-effects model are biased. The comparison of within- and between-cluster effects is in fact a regression-based alternative to the Hausman specification test (Baltagi 2013, 76–77). In the hybrid model (6), one can test whether $\beta_W = \beta_B$ using a Wald test. In the correlated random-effects model (7), one can test whether $\tau = 0$. Note that both tests are mathematically equivalent and yield the same test statistics. This is because $\tau = \beta_B - \beta_W$. If the between-cluster effect β_B and the within-cluster effect β_W are not statistically significantly different from each other (which implies that $\tau = 0$), this suggests that $\beta_W = \beta_B = \beta$. In this case, (6) and (7) simplify to (1), the standard random-intercept model. Substantively, this means that the random-effects model's assumption of a zero correlation between the level-two error and the level-one covariates holds. In contrast to the Hausman test, this test can also be used when we estimate (cluster) robust standard errors (SEs). Furthermore, it also works if the difference of the covariance matrices in the Hausman test is not positive definite. An additional advantage of this test is that it works at the level of individual variables. Thus one can use the more efficient random-effects estimate for those variables for which within and between effects do not differ significantly and retain both within and between effects for those variables for which they are significantly different.

4.2 Within-cluster effects in nonlinear models and nonlinear dependencies

Equivalence between effects estimates from standard fixed-effects models and effects estimates from hybrid or correlated random-effects models holds only in the linear case. In nonlinear models, the estimated within-cluster effects are typically similar, though not identical to the fixed-effects estimates, that is, the conditional likelihood estimates. In contrast to conditional likelihood approaches, which use a sufficient statistic to condition the cluster-level effects away (McCulloch, Searle, and Neuhaus 2008, 295), hybrid and correlated random-effects models take a parametric approach to the unobserved heterogeneity problem at level two (Wooldridge 2010, 286), placing certain restrictions on the conditional distribution of such heterogeneity, given the level-one covariates. While the conditional likelihood approach does not require any distributional assumption on the level-two error and its correlation with the covariates, the hybrid and correlated random-effects models assume that u_i depends on the mean values of x_{ij}.

Differences in effect estimates of nonlinear hybrid and correlated random-effects models and standard fixed-effects models can emerge because of a violation of this

assumption, that is, if the level-two error and the independent variables at level one are not completely but only linearly uncorrelated. Failure to meet this assumption may result in biased estimates (Brumback et al. 2010; Brumback, Dailey, and Zheng 2012). However, note that this assumption is still less restrictive than the assumption of complete independence between the level-two error and the level-one variables of (1).

If we implement a conditional likelihood approach for a model belonging to the family of GLMM, we can easily compare the estimates from (6) and (7) against actual fixed-effects estimates. For example, this is possible for logit models. Unfortunately, conditional likelihood approaches are not available for many GLMM. Differences between fixed-effects estimates and within-cluster estimates from hybrid and correlated random-effects models may suggest that u_i depends on x_{ij} through functional forms other than the cluster means of the cluster-varying covariates (Brumback et al. 2010, 1652).[3] Given enough observations within clusters, one can explicitly model other dependencies. For instance, one can do so by adding polynomial functions of the cluster means of the level-one covariates to the model (Allison 2014). A correlated random-effects model with additional quadratic and cubic terms is given by

$$g(\mu_{ij}) = \beta_W x_{ij} + \tau \overline{x}_i + \delta \overline{x}_i^2 + \eta \overline{x}_i^3 + \gamma c_i + v_i \tag{8}$$

Here we assume that $u_i = \tau \overline{x}_i + \delta \overline{x}_i^2 + \eta \overline{x}_i^3 + v_i$ and $v_i \sim N(0, \sigma^2_{v_i})$. Just as in (7), a test of $\tau = 0$, $\delta = 0$, and $\eta = 0$ can serve as inference regarding dependencies between u_i and x_{ij}. If the estimates of δ and η are not statistically significant, we can take this as evidence that the assumption of the correlated random-effects model is not violated. The equivalent hybrid model is

$$g(\mu_{ij}) = \beta_W (\overline{x}_i - x_{ij}) + \beta_B x_{ij} + \delta \overline{x}_i^2 + \eta \overline{x}_i^3 + \gamma c_i + v_i \tag{9}$$

Note that the estimated effects of the nonlinear dependencies are the same for both (8) and (9). In the hybrid model, however, inference regarding dependencies between u_i and x_{ij} requires a test of $\beta_W = \beta_B$, $\delta = 0$, and $\eta = 0$. Thus the only difference between a correlated random-effects model with nonlinear dependencies and a hybrid model with nonlinear dependencies lies in τ.

One may be tempted to compare the estimated within-cluster effects (β_W) from (6) or (7) with the effect estimates obtained by analogous models, including nonlinear dependencies. If β_W does not differ substantially when we account for additional nonlinear dependencies, we could take this as evidence that the assumption of the hybrid and correlated random-effects models holds. However, such a comparison is complicated by the fact that estimates in nonlinear models with fixed error variance at level one are not directly comparable because of the "rescaling problem" (Allison 1999; Kohler, Karlson, and Holm 2011). This extends to comparisons of β_W from (6) and β_W from (9) and of β_W from (7) and β_W from (8). Including additional nonlinear dependencies will affect the estimate of β_W, even if these are orthogonal to x_{ij} because of the rescaling problem. Consequently, β_W from (6) cannot be identical to β_W from

3. See also Chamberlain (1982).

(9), and the same applies to the β_W estimates from (7) and (8). Thus we should exert caution when comparing estimates from nonlinear models with different functional forms of the dependency between u_i and x_{ij}. A strict comparison of coefficients is not possible (Allison 1999; Kohler, Karlson, and Holm 2011). Instead, one should carefully inspect if the additional nonlinear dependencies are statistically significant or use a likelihood-ratio (LR) test to decide whether to remove or retain them.

4.3 Random slopes

Just like standard GLMMs, these hybrid and correlated random-effects models can include random slopes that allow the (within) effects of the level-one variables to vary between clusters. A hybrid model with a random slope on (β_W) is

$$g(\mu_{ij}) = (\beta_W + u_{2i})(x_{ij} - \overline{x}_i) + \beta_B \overline{x}_i + \gamma c_i + u_{1i}$$

What is the advantage of using a hybrid model with random slopes over a standard random-slope model? Again, in the standard random-effects model, we assume the random effects (including the random slope) to be uncorrelated with any unobserved characteristics at level two. Using a hybrid model relaxes this assumption. One can also specify correlated random-effects models with random slopes. However, while correlated random-effects and hybrid models without random slopes produce equivalent results (both in terms of fixed coefficients and variance components), this is not the case if random slopes are present (Kreft, de Leeuw, and Aiken 1995).

5 Hybrid and correlated random-effects models in Stata: The xthybrid command

Correlated random-effects and hybrid models have been increasingly discussed in the methodological literature (Allison 2009; Bell and Jones 2015; Neuhaus and Kalbfleisch 1998; Neuhaus and McCulloch 2006; Rabe-Hesketh and Skrondal 2012; Schunck 2013). However, their use in the wider scientific community is not widespread. The `xthybrid` command in this article simplifies the specification of hybrid as well as correlated random-effects models. It builds on the existing `mundlak` command (Perales 2013). However, unlike `mundlak`, it allows for all models belonging to the class of GLMM.

5.1 Syntax

`xthybrid` relies on Stata's `meglm` command to estimate hybrid and correlated random-effects versions of any two-level specification that can be fit with `meglm`. The syntax for `xthybrid` is

`xthybrid` *depvar indepvars* [*if*] [*in*], `clusterid(`*varname*`)` [`family(`*type*`)` `link(`*type*`)` `cre` `nonlinearities(`*type*`)` `randomslope(`*varlist*`)` `use(`*varlist*`)` `percentage(`#`)` `test` `full` `stats(`*list*`)` `se` `t` `p` `star` `vce(`*vcetype*`)` `iterations` `meglmoptions(`*list*`)`]

5.2 Options

`clusterid(`*varname*`)` specifies the cluster or grouping variable. `clusterid()` is required.

`family(`*type*`)` specifies the distribution of the outcome variable. *type* may be `gaussian`, `bernoulli`, `binomial`, `gamma`, `nbinomial`, `ordinal`, or `poisson`. The default is `family(gaussian)`.

`link(`*type*`)` specifies the link function. *type* may be `identity`, `log`, `logit`, `probit`, or `cloglog`. The default is `link(identity)`.

`cre` requests a correlated random-effects model instead of a hybrid model.

`nonlinearities(`*type*`)` adds polynomial functions of the cluster means to the model. *type* may be `quadratic`, `cubic`, or `quartic`.

`randomslope(`*varlist*`)` requests random slopes on the random-effect and within-group coefficients of selected variables.

`use(`*varlist*`)` splits between- and within-cluster effects only for selected explanatory variables.

`percentage(`#`)` sets the minimum percent within-cluster variance for explanatory variables to be considered cluster varying.

`test` presents test results of the random-effects assumption for separate model variables.

`full` prints the full model output (`meglm`).

`stats(`*list*`)` allows users to select which model summary statistics are reported.

`se` requests SEs for the parameters on model variables.

`t` requests *t*-values for the parameters on model variables.

`p` requests *p*-values for the parameters on model variables.

`star` requests stars to denote statistically significant parameters on model variables.

`vce(`*vcetype*`)` specifies the type of SE to be reported. *vcetype* may be `oim`, `robust`, or `cluster` *clustervar*.

`iterations` requests that the command be executed noisily.

`meglmoptions(`*list*`)` enables the user to request options from the `meglm` command.

5.3 Stored results

`xthybrid` returns in `e()` the same results as `meglm` because `xthybrid` relies on `meglm`. See [ME] **meglm** for details.

5.4 Applications

We now illustrate the `xthybrid` command through practical examples. We use Stata's `nlswork.dta`, which contains unbalanced panel data on a sample of 4,711 young, working, American women, observed up to 15 times between 1968 and 1988. The data are hence nested so that the level-one units are person-year observations and the level-two units are individuals. The cluster variable is `idcode`. Suppose that our interest is on the relationships between several socioeconomic factors (`age`, `msp`, and `race`) and the number of weekly hours worked (`hours`). Two of these factors are level-one variables that vary both within and between clusters (`age` and `msp`), whereas one is a level-two variable that varies only between clusters (`race`).

We first open the dataset, describe its contents, and create dummy variables out of the `race` variable:

```
. webuse nlswork
(National Longitudinal Survey.  Young Women 14-26 years of age in 1968)
. describe idcode hours age msp race
              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------
idcode          int     %8.0g                 NLS ID
hours           int     %8.0g                 usual hours worked
age             byte    %8.0g                 age in current year
msp             byte    %8.0g                 1 if married, spouse present
race            byte    %8.0g      racelbl    race
. generate black = race==2 if race!=.
. generate other = race==3 if race!=.
```

We can estimate the relationships between the variables of interest as a linear hybrid model using the `xthybrid` command. The `se` option requests SEs to be displayed below model coefficients, while the `test` option requests separate tests of the random-effects assumption ($\tau = 0$ or $\beta_W = \beta_B$ depending on the specification) for individual regressors.

```
. xthybrid hours age msp black other, clusterid(idcode) se test
The variable ´black´ does not vary sufficiently within clusters
and will not be used to create additional regressors.
[0% of the total variance in ´black´ is within clusters]
The variable ´other´ does not vary sufficiently within clusters
and will not be used to create additional regressors.
[0% of the total variance in ´other´ is within clusters]
```

```
Hybrid model. Family: gaussian. Link: identity.
```

Variable	model
hours	
R_black	0.5470
	0.2278
R_other	-0.0404
	0.9396
W_age	-0.0236
	0.0096
W_msp	-1.1661
	0.1529
B_age	0.0552
	0.0200
B_msp	-3.3647
	0.2685
_cons	36.5595
	0.5957
var(_cons[idcode])	
_cons	30.2310
	0.9900
var(e.hours)	
_cons	68.6073
	0.6334
Statistics	
ll	-1.032e+05
chi2	259.5143
p	0.0000
aic	2.065e+05
bic	2.066e+05

```
                                legend: b/se
Level 1: 28428 units. Level 2: 4709 units.

Tests of the random effects assumption:
  _b[B_age] = _b[W_age]; p-value: 0.0004
  _b[B_msp] = _b[W_msp]; p-value: 0.0000
```

In the `xthybrid` output, variables with the `W_` prefix denote within-cluster effects, variables with the `B_` prefix denote between-cluster effects, and variables with the `R_` prefix are those for which their effects are estimated the same as those in a standard random-effects model.[4]

As expected, `xthybrid` estimates two separate effects for the level-one variables `age` and `msp`. The coefficients `W_age` (−0.024) and `W_msp` (−1.166) give the within-cluster effects. Within-cluster increases in `age` are associated with a within-cluster decrease in `hours`, as are within-cluster increases in `msp`. That is, women in this sample work fewer hours in those years in which they are younger and unmarried relative to those years in which they are older and married, all else being equal.

4. `xthybrid` generates these variables in the background during operation. Users should be aware—if variables with the same name already exist in the active dataset, `xthybrid` will issue an error.

The coefficients on the level-one variables `B_age` and `B_msp` give their between-cluster effects. For `B_age`, the estimated coefficient (0.055) indicates that a between-individual one-year increase in age is associated with a small increase in work hours, suggesting that women from younger cohorts work more hours. For `B_msp`, the estimated coefficient (−3.365) indicates that on average, women who are never married in the data work about three hours less than women who are always married, all other things being equal.

The within-cluster effects are statistically different from the between-cluster effects, as can be seen from the small p-values in the formal tests of the random-effects assumption of orthogonality between the observables and the unobservables (`_b[B_age]= _b[W_age] p-value: 0.0004` and `_b[B_msp]=_b[W_msp] p-value: 0.0000`). This constitutes evidence in favor of rejecting such an assumption as well as using a standard random-effects model.

An analogous correlated random-effects model can be estimated using `xthybrid` by adding the `cre` option (the results are presented in table 2):

```
. xthybrid hours age msp black other, clusterid(idcode) se test cre
```
(*output omitted*)

For the sake of comparison, estimates from analogous, standard, random-effects, and fixed-effects models (estimated using `xtreg`) are also presented in table 2. Such models are fit as follows:[5]

```
. xtreg hours age msp black other, i(idcode) fe
```
(*output omitted*)
```
. xtreg hours age msp, i(idcode) fe
```
(*output omitted*)

5. Note that we include the level-two variables `black` and `other` in the fixed-effects model although they are omitted in the estimation. This ensures that the fixed-effects model uses the same sample as the random-effects model. There are, obviously, more elegant ways to define the analysis sample (see, for example, Schunck [2013]).

Table 2. Coefficients from linear models (identity link and Gaussian distribution)

	(1) Hybrid model	(2) Correlated random-effects model	(3) Random-effects model	(4) Fixed-effects model
`W_age`	-0.024^{*}	-0.024^{*}		
	(0.010)	(0.010)		
`W_msp`	-1.166^{***}	-1.166^{***}		
	(0.153)	(0.153)		
`B_age`	0.055^{**}			
	(0.020)			
`B_msp`	-3.365^{***}			
	(0.269)			
`D_age`		0.079^{***}		
		(0.022)		
`D_msp`		-2.199^{***}		
		(0.309)		
`R_black`	0.547^{*}	0.547^{*}		
	(0.228)	(0.228)		
`R_other`	-0.040	-0.040		
	(0.940)	(0.940)		
`age`			-0.010	-0.024^{*}
			(0.009)	(0.010)
`msp`			-1.627^{***}	-1.166^{***}
			(0.134)	(0.152)
`black`			0.860^{***}	
			(0.243)	
`other`			-0.035	
			(1.021)	
`_cons`	36.560^{***}	36.560^{***}	37.303^{***}	37.944^{***}
	(0.596)	(0.596)	(0.286)	(0.284)
N (Level 2)	4709	4709	4709	4709
N (Level 1)	28428	28428	28428	28428

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In the correlated random-effects model, the coefficients `W_age` (-0.024) and `W_msp` (-1.166) give the within-cluster effects on the `age` and `msp` and are identical to those fit in the hybrid model. In these linear models, the within-cluster effects fit by both the hybrid and correlated random-effects models are the same as those fit by a standard fixed-effects model.

The coefficients `D_age` (0.079) and `D_msp` (−2.199) give the difference between the between- and within-cluster effects. For example, using the estimated between- and within-cluster effects for the variable `age` from the hybrid model (that is, `B_age` and `W_age`) and the `D_age` coefficient in the correlated random-effects model, we see that $0.055 - (-0.024) = 0.079$.

In the correlated random-effects model, the coefficients on the cluster-invariant variables `R_black` (0.547) and `R_other` (−0.040) are estimated like those in a standard random-effects regression model and are identical to those in the hybrid model. For these to be unbiased, the random-effects assumption of orthogonality between observables and unobservables at level two must still hold. Note that these coefficients are not identical to those in the standard random-effects model. The inclusion of the cluster-mean variables accounts for additional sources of between-cluster variation, which affects the estimated effects of these level-two variables (Schunck 2013, 71).

We can also use `xthybrid` to fit a model with random slopes for level-one variables. For instance, if we wanted to allow the slope of `age` to vary across clusters, we would specify

```
. xthybrid hours age msp black other, clusterid(idcode) se randomslope(age)
> iterations
The variable ´black´ does not vary sufficiently within clusters
and will not be used to create additional regressors.
[0% of the total variance in ´black´ is within clusters]
The variable ´other´ does not vary sufficiently within clusters
and will not be used to create additional regressors.
[0% of the total variance in ´other´ is within clusters]
Fitting fixed-effects model:
Iteration 0:   log likelihood = -105142.73
Iteration 1:   log likelihood = -105142.73
Refining starting values:
Grid node 0:   log likelihood = -104177.16
Fitting full model:
Iteration 0:   log likelihood = -104177.16  (not concave)
Iteration 1:   log likelihood = -104034.22
Iteration 2:   log likelihood =  -102461.4
Iteration 3:   log likelihood = -102325.48
Iteration 4:   log likelihood = -102298.87
Iteration 5:   log likelihood = -102298.83
Iteration 6:   log likelihood = -102298.83
```

```
Mixed-effects GLM                               Number of obs     =     28,428
Family:                Gaussian
Link:                  identity
Group variable:          idcode                 Number of groups  =      4,709
                                                Obs per group:
                                                              min =          1
                                                              avg =        6.0
                                                              max =         15
Integration method: mvaghermite                 Integration pts.  =          7
                                                Wald chi2(6)      =     201.16
Log likelihood = -102298.83                     Prob > chi2       =     0.0000
```

hours	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
R_black	.5174951	.2310199	2.24	0.025	.0647044	.9702859
R_other	-.1408401	.9487884	-0.15	0.882	-2.000431	1.718751
W_age	-.0142051	.0163478	-0.87	0.385	-.0462462	.0178361
W_msp	-.6070389	.1589173	-3.82	0.000	-.9185111	-.2955667
B_age	.0579946	.0199396	2.91	0.004	.0189137	.0970755
B_msp	-3.363616	.2707035	-12.43	0.000	-3.894185	-2.833047
_cons	36.4563	.5944607	61.33	0.000	35.29118	37.62142
idcode						
var(W_age)	.4968284	.0233867			.4530423	.5448465
var(_cons)	34.09215	1.030189			32.13164	36.17227
var(e.hours)	55.18712	.5573331			54.10551	56.29035

```
LR test vs. linear model: chi2(2) = 5687.80                 Prob > chi2 = 0.0000
Note: LR test is conservative and provided only for reference.
```

Note that we have additionally specified the `iterations` option, which requests `xthybrid` to display the underlying `meglm` output. This is a sensible choice to detect possible convergence problems when models become more complicated.

An LR test, which compares the restricted model (the model without the random slope) with the unrestricted model (the model with the random slope), indicates that the model with a random slope for `age` fits the data better (LR $\chi^2(1) = 1861.97$).

```
. xthybrid hours age msp black other, clusterid(idcode) se
  (output omitted)
. estimates store model1
. xthybrid hours age msp black other, clusterid(idcode) se randomslope(age)
  (output omitted)
. estimates store model2
. lrtest model1 model2
Likelihood-ratio test                                 LR chi2(1)  =   1861.97
(Assumption: model1 nested in model2)                 Prob > chi2 =    0.0000
Note: The reported degrees of freedom assumes the null hypothesis is not on the
      boundary of the parameter space.  If this is not true, then the reported
      test is conservative.
```

Because `xthybrid` relies on the Stata `meglm` command, it can easily handle nonlinear models. To illustrate this, we transform our continuous outcome variable (`hours`) into a dummy variable (`full_time`) where `1` indicates that the respondent works full-time ($>$ 30 hours) and `0` indicates that the respondent works part-time ($\leq$ 30 hours). We then fit a model using the logit link and the binomial distribution for the outcome variable, which yields a logistic regression (table 3):

```
. generate full_time = hours > 30 if hours!=.
. xthybrid full_time age msp black other, clusterid(idcode) family(binomial)
> link(logit) se
  (output omitted)
```

If we wanted a correlated random-effects logit model, we would instead specify

```
. xthybrid full_time age msp black other, clusterid(idcode) family(binomial)
> link(logit) se cre
  (output omitted)
```

The results show that also in nonlinear specifications, the estimated within-cluster effects on the level-one and level-two variables are the same in the hybrid model (column 1) and the correlated random-effects model (column 2). The corresponding standard random-intercept and fixed-effects logit models are specified as (for results, see table 3)

```
. xtlogit full_time age msp black other, i(idcode) re
  (output omitted)
. xtlogit full_time age msp black other, i(idcode) fe
  (output omitted)
```

However, as explained before, within-cluster effects from hybrid and correlated random-effects models in these nonlinear specifications are not identical to those from a standard fixed-effects model, though they are very similar. For instance, the estimated within effect for `age` from the hybrid model and the correlated random-effects model is -0.020, whereas in the fixed-effects model this effect is -0.019.

Note that the number of cases differs for the fixed-effects model and the other models. This is because the fixed-effects logit model discards all clusters that have no within-cluster variation, because they do not contribute information about the parameters being estimated (these clusters' contributions to the log likelihood is zero). However, although the number of cases is higher in the hybrid and the correlated random-effects models, these models are not more efficient when it comes to estimating the within-cluster effects. Here, too, we can use only those cases that have within variation in the dependent variable and the covariates. However, we use all cases to estimate the between-cluster effects, that is, the effects of level-two covariates, and the variance components.

Table 3. Coefficients from logit models (logit link and binomial distribution)

	(1) Hybrid model	(2) Correlated random-effects model	(3) Random-effects model	(4) Fixed-effects model
`W_age`	−0.020***	−0.020***		
	(0.003)	(0.003)		
`W_msp`	−0.438***	−0.438***		
	(0.058)	(0.058)		
`B_age`	−0.004			
	(0.007)			
`B_msp`	−1.030***			
	(0.093)			
`D_age`		0.016*		
		(0.008)		
`D_msp`		−0.592***		
		(0.109)		
`R_black`	0.453***	0.453***		
	(0.079)	(0.079)		
`R_other`	0.188	0.188		
	(0.319)	(0.319)		
`age`			−0.017***	−0.019***
			(0.003)	(0.003)
`msp`			−0.602***	−0.427***
			(0.050)	(0.057)
`black`			0.542***	
			(0.078)	
`other`			0.207	
			(0.320)	
`_cons`	2.637***	2.637***	2.730***	
	(0.204)	(0.204)	(0.102)	
N (Level 1)	28428	28428	28428	15036
N (Level 2)	4709	4709	4709	2042

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We can use the `xthybrid` command to fit hybrid and correlated random-effects models using specifications for which a conditional likelihood (fixed-effects) estimator does not exist or has not been implemented in Stata. The different columns in table 4

show the results of hybrid models fit using `xthybrid` for three specifications for which a fixed-effects approach is not readily available:[6]

- a probit model of the binary variable capturing full-time work (column 1)
- an ordered logit model of an ordinal variable capturing work hours (column 2)
- a negative binomial model of weeks spent in unemployment (column 3)

Table 4. Coefficients from hybrid models

	(1) Probit	(2) Ordered logit	(3) Negative binomial
`W_age`	−0.009***	−0.004	0.002
	(0.002)	(0.003)	(0.005)
`W_msp`	−0.234***	−0.320***	−0.240**
	(0.032)	(0.048)	(0.078)
`B_age`	−0.002	0.028***	−0.004
	(0.004)	(0.006)	(0.006)
`B_msp`	−0.572***	−0.824***	−0.494***
	(0.052)	(0.076)	(0.073)
`R_black`	0.248***	−0.077	0.559***
	(0.044)	(0.064)	(0.062)
`R_other`	0.109	−0.155	0.380
	(0.179)	(0.261)	(0.250)
`_cons`	1.484***		1.082***
	(0.114)		(0.164)
`cut1`		-2.458***	
		(0.170)	
`cut2`		3.013***	
		(0.171)	
`cut3`		6.719***	
		(0.193)	
`cut4`		10.302***	
		(0.530)	
`lnalpha`			2.472***
			(0.024)
N (Level 1)	28428	28428	22794
N (Level 2)	4709	4709	4709

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6. Note that a conditional likelihood approach for the negative binomial model is implemented in Stata, but this method does not control for unobserved heterogeneity at the cluster level (Allison and Waterman 2002; Green 2007; Guimarães 2008).

The requisite syntax to fit these models is

```
. recode hours 0/20=1 21/40=2 41/60=3 61/100=4 101/200=5, gen(hour_categories)
(28454 differences between hours and hour_categories)
. xthybrid full_time age msp black other, clusterid(idcode) family(binomial)
> link(probit) se
```

(*output omitted*)

```
. xthybrid hour_categories age msp black other, clusterid(idcode) family(ordinal)
> link(logit) se
```

(*output omitted*)

```
. xthybrid wks_ue age msp black other, clusterid(idcode) family(nbinomial)
> link(log) se
```

(*output omitted*)

If a conditional likelihood estimator is available, as with the logit model, it is easy to judge whether the within effect in the GLMM corresponds to the fixed-effects estimate. If it does not, this may be an indication that u_i depends on x_{ij} through other functions of the cluster means of the level-one variables. If a conditional likelihood estimator is not available, an LR test can assess whether including other functions of $\overline{x}_i$ is sensible. As an example, the following syntax requests a probit hybrid model with quadratic polynomials of the cluster means:

```
. xthybrid full_time age msp black other, clusterid(idcode) family(binomial)
> link(probit) se test nonlinearities(quadratic)
The variable ´black´ does not vary sufficiently within clusters
and will not be used to create additional regressors.
[0% of the total variance in ´black´ is within clusters]
The variable ´other´ does not vary sufficiently within clusters
and will not be used to create additional regressors.
[0% of the total variance in ´other´ is within clusters]
```

```
Hybrid model. Family: binomial. Link: probit.
```

Variable	model
full_time	
R_black	0.2474
	0.0436
R_other	0.1282
	0.1773
W_age	-0.0093
	0.0019
W_msp	-0.2390
	0.0325
B_age	0.0761
	0.0335
B_age_2	-0.0013
	0.0006
B_msp	0.2753
	0.2024
B_msp_2	-0.8287
	0.1879
_cons	0.2231
	0.4845
var(_cons[idcode])	
_cons	0.8631
	0.0418
Statistics	
ll	-1.203e+04
chi2	316.8241
p	0.0000
aic	24082.5895
bic	24165.1408

```
                              legend: b/se
Level 1: 28428 units. Level 2: 4709 units.

Tests of the random effects assumption:
  _b[B_age] = _b[W_age]; p-value: 0.0108
  _b[B_msp] = _b[W_msp]; p-value: 0.0125
```

The quadratic terms of the cluster means for the variables `msp` and `age` are statistically significant (`B_age_2` $= -0.001$, SE $= 0.001$; `B_msp_2` $= -0.829$, SE $= 0.188$), but their inclusion has negligible effects (from a substantial significance standpoint) on the estimated within-cluster effects. However, an LR test suggests that the model with the quadratic terms of the cluster means fits the data better than a model without. Including a cubic function does not further improve the model fit:[7]

```
. quietly xthybrid full_time age msp black other, clusterid(idcode)
> family(binomial) link(probit) se test

. estimates store model3
```

7. Note that a Wald test of joint insignificance of the quadratic and the cubic terms comes to the same conclusion. This is not surprising, considering that the LR test and the Wald test are asymptotically equivalent (Johnston and DiNardo 1997, 150).

```
. quietly xthybrid full_time age msp black other, clusterid(idcode)
> family(binomial) link(probit) se test nonlinearities(quadratic)
. estimates store model4
. quietly xthybrid full_time age msp black other, clusterid(idcode)
> family(binomial) link(probit) se test nonlinearities(cubic)
. estimates store model5
. lrtest model3 model4
Likelihood-ratio test                                 LR chi2(2)  =     28.64
(Assumption: model3 nested in model4)                 Prob > chi2 =    0.0000
. lrtest model4 model5
Likelihood-ratio test                                 LR chi2(2)  =      3.06
(Assumption: model4 nested in model5)                 Prob > chi2 =    0.2166
```

In this case, we would choose the model that includes quadratic terms of the cluster means. Finally, as with the linear model, `xthybrid` can fit nonlinear models with random slopes. For instance, a hybrid GLMM with a logit link and binomial distribution (a hybrid logit model) with a random slope on `age` is specified as

```
. xthybrid full_time age msp black other, clusterid(idcode) family(binomial)
> link(logit) se randomslope(age)
  (output omitted)
```

Note that we fit GLMMs with maximum likelihood estimation (Skrondal and Rabe-Hesketh 2004; McCulloch, Searle, and Neuhaus 2008; Wooldridge 2010). This applies also to models fit using Stata's `meglm` command, on which the `xthybrid` command is based. Stata uses numerical integration to calculate and maximize the likelihood. Because these methods are computationally intensive, the models may take some time to converge. It is therefore sensible to start by fitting simple models, with few random terms.

Maximum likelihood estimates are consistent, asymptotically normally distributed, and asymptotically efficient (McCulloch and Neuhaus 2013; Wooldridge 2010). However, maximum-likelihood estimation does not perform well with small samples, often providing biased estimates (Neuhaus and McCulloch 2006, 79). The minimum number of observations required for robust maximum likelihood estimation of GLMMs depends on the model specification (for discussions on sample sizes in mixed models, see Maas and Hox [2005]; Moineddin, Matheson, and Glazier [2007]; Schunck [2016]).

6 Conclusion

We have discussed the rationale behind hybrid and correlated random-effects models (Allison 2009; Schunck 2013; Wooldridge 2010) for clustered data, focusing on how we can adapt these to specifications that fall under the umbrella of GLMM (McCullagh and Nelder 1989; McCulloch, Searle, and Neuhaus 2008; Rabe-Hesketh, Skrondal, and Pickles 2004; Skrondal and Rabe-Hesketh 2003). We introduced the user-written `xthybrid` command as an accessible and flexible tool to fit these models using Stata.

When a researcher's main goal is to account for unobserved heterogeneity at the cluster level, fixed effects (for example, conditional likelihood) approaches are the best choice. However, these approaches are not feasible for some models, including probit models, ordered logit or probit models, and certain count models (Wooldridge 2010). In these situations, hybrid and correlated random-effects models are a good solution to obtain proximate fixed-effects estimates (Cameron and Trivedi 2005, 719; Neuhaus and McCulloch 2006, 886).

One limitation of nonlinear hybrid and correlated random-effects models is that the within-cluster effect estimates are not identical to those from standard fixed-effects models, which may lead to bias (Brumback et al. 2010). When possible, we advise comparisons of the estimated within-cluster effects from hybrid and correlated random-effects models and standard fixed-effects models. When not, including nonlinear dependencies in the model can be insightful as to whether the correlated random-effects assumption is violated. Despite this shortcoming relative to traditional conditional likelihood approaches, estimates from correlated random-effects or hybrid models are still preferable to those from standard random-effects models that impose very strict assumptions on the distributions of the unobserved effects and their correlation with the covariates. Correlated random-effects and hybrid models can also be more useful than standard fixed-effects models, because the latter fail to retrieve the effects on the outcome variable of level-two variables. Moreover, correlated random-effects and hybrid models can include random slopes; thus they can relax the assumption of effect homogeneity of level-one variables across clusters. Overall, decomposing within and between effects in GLMM is a flexible alternative to standard random and fixed-effects models.

7 Acknowledgments

The authors thank Marco Giesselmann, Philipp Lersch, and an anonymous reviewer for their constructive and helpful feedback and Yangtao Huang for assistance testing the `xthybrid` command. This research was supported by the Australian Research Council Centre of Excellence for Children and Families over the Life Course (project number CE140100027).

8 References

Allison, P. D. 1999. Comparing logit and probit coefficients across groups. *Sociological Methods and Research* 28: 186–208.

———. 2009. *Fixed Effects Regression Models*. Thousand Oaks, CA: Sage.

———. 2014. Problems with the hybrid method. http://statisticalhorizons.com/problems-with-the-hybrid-method.

Allison, P. D., and R. P. Waterman. 2002. Fixed-effects negative binomial regression models. *Sociological Methodology* 32: 247–265.

Andersen, E. B. 1970. Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society, Series B* 32: 283–301.

Baltagi, B. H., ed. 2006. *Panel Data Econometrics: Theoretical Contributions and Empirical Applications*. Amsterdam: Elsevier.

Baltagi, B. H. 2013. *Econometric Analysis of Panel Data*. 5th ed. New York: Wiley.

Bell, A., and K. Jones. 2015. Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods* 3: 133–153.

Brumback, B. A., A. B. Dailey, L. C. Brumback, M. D. Livingston, and Z. He. 2010. Adjusting for confounding by cluster using generalized linear mixed models. *Statistics & Probability Letters* 80: 1650–1654.

Brumback, B. A., A. B. Dailey, and H. W. Zheng. 2012. Adjusting for confounding by neighborhood using a proportional odds model and complex survey data. *American Journal of Epidemiology* 175: 1133–1141.

Cameron, A. C., and P. K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Chamberlain, G. 1980. Analysis of covariance with qualitative data. *Review of Economic Studies* 47: 225–238.

———. 1982. Multivariate regression models for panel data. *Journal of Econometrics* 18: 5–46.

Goetgeluk, S., and S. Vansteelandt. 2008. Conditional generalized estimating equations for the analysis of clustered and longitudinal data. *Biometrics* 64: 772–780.

Green, W. 2007. Functional form and heterogeneity in models for count data. *Foundations and Trends in Econometrics* 1: 113–218.

Guimarães, P. 2008. The fixed effects negative binomial model revisited. *Economics Letters* 99: 63–66.

Halaby, C. N. 2004. Panel models in sociological research: Theory into practice. *Annual Review of Sociology* 30: 507–544.

Hsiao, C. 2003. *Analysis of Panel Data*. 2nd ed. Cambridge: Cambridge University Press.

Johnston, J., and J. DiNardo. 1997. *Econometric Methods*. 4th ed. New York: McGraw–Hill.

Kohler, U., K. B. Karlson, and A. Holm. 2011. Comparing coefficients of nested nonlinear probability models. *Stata Journal* 11: 420–438.

Kreft, I. G. G., J. de Leeuw, and L. S. Aiken. 1995. The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research* 30: 1–21.

Maas, C. J. M., and J. J. Hox. 2005. Sufficient sample sizes for multilevel modeling. *Methodology* 1: 86–92.

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*. 2nd ed. London: Chapman & Hall/CRC.

McCulloch, C. E., and J. M. Neuhaus. 2013. Generalized linear mixed models: Estimation and inference. In *The SAGE Handbook of Multilevel Modeling*, ed. M. A. Scott, J. S. Simonoff, and B. D. Marx, 271–286. London: Sage.

McCulloch, C. E., S. R. Searle, and J. M. Neuhaus. 2008. *Generalized, Linear, and Mixed Models*. 2nd ed. Hoboken, NJ: Wiley.

Moineddin, R., F. I. Matheson, and R. H. Glazier. 2007. A simulation study of sample size for multilevel logistic regression models. *BMC Medical Research Methodology* 7: 34.

Mundlak, Y. 1978. On the pooling of time series and cross section data. *Econometrica* 46: 69–85.

Neuhaus, J. M., and J. D. Kalbfleisch. 1998. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 54: 638–645.

Neuhaus, J. M., and C. E. McCulloch. 2006. Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *Journal of the Royal Statistical Society, Series B* 68: 859–872.

Perales, F. 2013. mundlak: Stata module to estimate random-effects regressions adding group-means of independent variables to the model. Statistical Software Components S457601, Department of Economics, Boston College.
http://econpapers.repec.org/software/bocbocode/s457601.htm.

Rabe-Hesketh, S., and A. Skrondal. 2012. *Multilevel and Longitudinal Modeling Using Stata*. 3rd ed. College Station, TX: Stata Press.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles. 2004. Generalized multilevel structural equation modeling. *Psychometrika* 69: 167–190.

Raudenbush, S. 1989. "Centering" predictors in multilevel analysis: Choices and consequences. *Multilevel Modelling Newsletter* 1(2): 10–12.

Sampson, R. J. 2003. The neighborhood context of well-being. *Perspectives in Biology and Medicine* 46: 53–64.

Sampson, R. J., S. W. Raudenbush, and F. Earls. 1997. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science* 277: 918–924.

Schunck, R. 2013. Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models. *Stata Journal* 13: 65–76.

———. 2016. Cluster size and aggregated level 2 variables in multilevel models: A cautionary note. *Methods, Data, Analyses* 10: 97–108.

Skrondal, A., and S. Rabe-Hesketh. 2003. Some applications of generalized linear latent and mixed models in epidemiology: Repeated measures, measurement error and multilevel modeling. *Norwegian Journal of Epidemiology* 13: 265–278.

———. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman & Hall/CRC.

Snijders, T. A. B., and J. Berkhof. 2008. Diagnostic checks for multilevel models. In *Handbook of Multilevel Analysis*, ed. J. de Leeuw and E. Meijer, 141–175. Berlin: Springer.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

About the authors

Reinhard Schunck is the head of GESIS Training at GESIS—Leibniz–Institute for the Social Sciences in Cologne, Germany. His research interests span multilevel and longitudinal data analysis. He works primarily in the field of social stratification and inequality.

Francisco (Paco) Perales is a research fellow at the Life Course Centre at the Institute for Social Science Research (The University of Queensland). His research relies heavily on the analysis of large-scale panel surveys. His current interests include the impact of life-course transitions on gender-based socioeconomic inequality, differences in life outcomes by sexual identity, and the effect of early life-course family structure on children's development.