The Stata Journal (2013)
13, Number 1, pp. 65–76

Within and between estimates in random-effects models: Advantages and drawbacks of correlated random effects and hybrid models

Reinhard Schunck
Department of Sociology
University of Bielefeld
Bielefeld, Germany
reinhard.schunck@uni-bielefeld.de

Abstract. Correlated random-effects (Mundlak, 1978, *Econometrica* 46: 69–85; Wooldridge, 2010, *Econometric Analysis of Cross Section and Panel Data* [MIT Press]) and hybrid models (Allison, 2009, *Fixed Effects Regression Models* [Sage]) are attractive alternatives to standard random-effects and fixed-effects models because they provide within estimates of level 1 variables and allow for the inclusion of level 2 variables. I discuss these models, give estimation examples, and address some complications that arise when interaction effects are included.

Keywords: st0283, xtreg, xtmixed, multilevel data, panel data, fixed effects, random effects, correlated random effects, hybrid model

1 Introduction

It is widely recognized that fixed-effects models have an advantage over random-effects models when analyzing panel data because they control for all level 2 characteristics, measured or unmeasured (Allison 2009; Halaby 2004; Wooldridge 2010). This also applies in a multilevel framework. However, a major drawback of fixed-effects models is their inability to estimate the effect of any variable that does not vary within clusters, which holds for all level 2 variables. To circumvent this disadvantage, it has been proposed to estimate within effects in random-effects models (Allison 2009; Neuhaus and Kalbfleisch 1998; Rabe-Hesketh and Skrondal 2008; Raudenbush 1989a; Wooldridge 2010).

2 Models

In the linear case,[1] the random-intercept model is given by

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 c_i + \mu_i + \epsilon_{it} \tag{1}$$

1. The strategy presented here also extends to nonlinear models (Allison 2009; Neuhaus and Kalbfleisch 1998, Wooldridge 2010).

 st0283

where subscript i denotes level 2 (for example, subjects) and t denotes level 1 (for example, occasions). x_{it} is a level 1 variable that varies between and within clusters, and c_i is a level 2 variable that varies only between clusters. μ_i is the level 2 error and the random intercept, and ϵ_{it} is the level 1 error. Throughout the article, ϵ_{it} will be treated as white noise and not considered further.

The standard distributional assumption regarding the level 2 error is $\mu_i|x_{it}, c_i \sim N(0, \sigma^2_\mu)$. The model provides consistent effect estimates if $E(\mu_i|x_{it}, c_i) = 0$. Subtracting the between model

$$\overline{y}_i = \beta_0 + \beta_1\overline{x}_i + \beta_2 c_i + \mu_i + \overline{\epsilon}_i$$

from (1) provides the fixed-effects model in its demeaned form:

$$(y_{it} - \overline{y}_i) = \beta_1(x_{it} - \overline{x}_i) + (\epsilon_{it} - \overline{\epsilon}_i) \tag{2}$$

The subtraction removes the level 2 error (μ_i) from the equation. As a result, the model's estimate of β_1 is unbiased even if $E(\mu_i|x_{it}) \neq 0$. But this comes at a cost. The subtraction also removes all variables that do not vary at level 1. Fixed-effects models therefore cannot estimate the effect of level 2 variables. This may not be seen as a problem in panel-data analysis. But it is definitely a major drawback in multilevel analysis, where interest often lies particularly in estimating the effect of level 2 variables.

However, it is possible to estimate within effects in random-effects models by decomposing level 1 variables into a between ($\overline{x}_i = n_i^{-1}\sum_{t=1}^{n_i} x_{it}$) and a cluster ($x_{it} - \overline{x}_i$) component. This hybrid model (Allison 2009) is given by

$$y_{it} = \beta_0 + \beta_1(x_{it} - \overline{x}_i) + \beta_2 c_i + \beta_3\overline{x}_i + \mu_i + \epsilon_{it} \tag{3}$$

Using (3) to estimate β_1 gives the within-effect estimate, that is, the fixed-effects estimate (Mundlak 1978; Neuhaus and Kalbfleisch 1998). Hence, the estimates of β_1 from (2) and (3) are identical. Because (3) is a random-effects model, we can use it to estimate effects of level 2 variables. However, for the estimate of β_2 to be unbiased, $E(\mu_i|x_{it}, c_i) = 0$ and $\mu_i|x_{it}, c_i \sim N(0, \sigma^2_\mu)$ still have to hold. Moreover, even though (3) is a random-effects model, its estimate of β_1 is not more efficient than the one obtained through estimating (2), because both estimates are solely based on within variation. β_3 estimates the between effect (Mundlak 1978; Neuhaus and Kalbfleisch 1998). While it is not necessary to include the cluster mean ($\overline{x}_i$) to obtain the within estimate of β_1, its inclusion ensures that effect estimates of level 2 variables are corrected for between-cluster differences in x_{it}.

The idea to decompose between and within variation and to estimate the respective effects in a single model is not new (Kaufman 1993; Kreft, de Leeuw, and Aiken 1995; Neuhaus and Kalbfleisch 1998; Raudenbush 1989a), but it seems to have become increasingly popular in panel data (Burnett and Farkas 2009; Phillips 2006; Ousey and Wilcox 2007; Teachman 2011; Zhou 2011) as well as in multilevel analysis (Curran and Bauer 2011; Epstein et al. 2012; Landale, Gorman, and Oropesa 2006; Nomaguchi and Brown 2011; Park, Lee, and Epstein 2009; Schempf et al. 2011).

This approach offers several additional advantages. First, it allows us to test for equivalence of within and between estimates. This test, which is referred to as an augmented regression test (Jones et al. 2007, 217), can be used as an alternative to the Hausman specification test (Baltagi 2008, 73). If between and within effects are the same—that is, $\beta_1 = \beta_3$—then (3) collapses to (1), the random-intercept model. Second, a decomposition into between and within effects can be used with generalized estimating equations, which enables us to specify less restrictive within-cluster error structures. Third, this approach can be extended to include random slopes, allowing effects of level 1 variables to vary between clusters.

A hybrid model that includes a random slope for $(x_{it} - \overline{x}_i)$ is given by

$$y_{it} = \beta_0 + (\beta_1 + \mu_{2i})(x_{it} - \overline{x}_i) + \beta_2 c_i + \beta_3 \overline{x}_i + \mu_{1i} + \epsilon_{it}$$

The hybrid model is closely related to the correlated random-effects model (Wooldridge 2010), first proposed by Mundlak (1978). The correlated random-effects model relaxes the assumption of zero correlation between the level 2 error and the level 1 variables. In particular, it introduces the assumption $\mu_i = \pi \overline{x}_i + \nu_i$, so (1) becomes

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 c_i + \pi \overline{x}_i + \nu_i + \epsilon_{it} \tag{4}$$

The cluster mean of x_{it} picks up any correlation between this variable and the level 2 error. Including the cluster mean of a level 1 variable in a random-effects model is therefore an alternative to cluster mean centering (Halaby 2003, 519). Thus β_1 from (4) is the fixed-effects estimate (Mundlak 1978; Wooldridge 2010), and it is identical to the estimate obtained from (3). But the estimated effect of $\overline{x}_i$ will differ. In the hybrid model, this is the between effect. In the correlated random-effects model, this is the difference of the within and between effects (Mundlak 1978); that is, $\pi = \beta_3 - \beta_1$.

The relation between the hybrid model and the correlated random-effects model becomes obvious if we rewrite (3) as

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 c_i + (\beta_3 - \beta_1)\overline{x}_i + \mu_i + \epsilon_{it} \tag{5}$$

The correlated random-effects model allows for the inclusion of level 2 variables, and it can be used with generalized estimating equations, just like the hybrid model. It is also possible to perform an augmented regression test. But the test takes a different form. Because π already estimates the difference of the within and between effects, the null hypothesis is $\pi = 0$ (Baltagi 2008, 133). In principle, the correlated random-effects model can also include random slopes. However, a correlated random-effects model with random slopes is not equivalent to the corresponding hybrid model with random slopes, and it is advisable to use the hybrid model (Raudenbush 1989a; 1989b).[2]

2. In particular, because $\pi = \beta_3 - \beta_1$, the random part of the slope, μ_{2i}, will appear in the estimated effect of x_{it} and $\overline{x}_i$, which makes it hard to interpret (see Raudenbush [1989a, 12]).

3 Estimation

The example presented below uses infant birth weight data (Abrevaya 2006; Rabe-Hesketh and Skrondal 2008). The data comprise 8,604 infants clustered within 3,978 mothers. We will examine how the level 1 variables, mother's age (`mage`, continuous) and smoking behavior (`smoke`, binary), and the level 2 variable, race (`black`, binary), affect a child's birth weight (`birwt`, continuous). To estimate between and within effects in one model, we must first generate the cluster-specific mean of x_{it}. The second step is to create the deviation scores, which is also known as group mean centering. We have to ensure that the means are generated on the multivariate sample, that is, by using listwise deletion to handle missings. This is done using `mark` (Jann 2007b).

```
. use http://www.stata-press.com/data/mlmus2/smoking
. mark nonmiss
. markout nonmiss birwt smoke mage black
. egen msmoke = mean(smoke) if nonmiss==1, by(momid)
. generate dsmoke = smoke-msmoke
```

The cluster-specific means and the deviation scores can also be computed easily with the `center` command (Jann 2007a).

```
. by momid, sort: center mage if nonmiss==1, prefix(d) mean(m)
```

Once the variables have been created, `xtreg, re` can be used to fit the hybrid model.

```
. xtreg birwt dsmoke dmage msmoke mmage black, i(momid) re
. estimates store hybrid
```

The correlated random-effects model is fit in a similar way, but it includes uncentered versions of the level 1 variables.

```
. xtreg birwt smoke mage msmoke mmage black, i(momid) re
. estimates store corr_re
```

We also fit a random-intercept and a fixed-effects model so that we can compare their estimates with those of the hybrid model and the correlated random-effects model.

```
. xtreg birwt smoke mage black, i(momid) re
. estimates store re
. xtreg birwt smoke mage, i(momid) fe
. estimates store fe
```

Estimation results are shown in table 1. Model 1 presents the estimates from the random-intercept model, model 2 from the fixed-effects model, model 3 from the hybrid model, and model 4 from the correlated random-effects model. Comparing the random-intercept model (model 1) with the fixed-effects model (model 2), we see that there are considerable differences in effect estimates. The estimated effect on birth weight of smoking during pregnancy, for instance, is considerably smaller when estimated as a within-mother effect, that is, once all time-constant (observed and unobserved) differences between mothers are accounted for. The estimated effect of mother's age, on the other hand, is considerably larger in the fixed-effects model than in the random-effects model.

If we compare the estimates from the fixed-effects model (model 2) with those from the hybrid model (model 3) and the correlated random-effects model (model 4), we see that all three models estimate the same (within) effects of the level 1 variables.[3] A comparison of between and within effects from the hybrid model (model 3) provides additional insight. The model, for instance, estimates the between-mother effect of smoking as −332.93. Accordingly, the average birth weight for a mother who smokes and a mother who does not smoke will differ by 332.928 grams. The within-mother effect is estimated to be −105.70. Thus for a given mother, smoking decreases the birth weight of her children by 105.70 on average. A Wald test can be used to test for equivalence of within and between estimates.

```
. estimates restore hybrid
. test dsmoke=msmoke
. test dmage=mmage
. hausman fe re
```

In the present case, the test statistics suggest that the null hypothesis of equality for within and between estimates should be rejected [smoking: Wald $\chi^2(1) = 38.68$, age: Wald $\chi^2(1) = 22.09$], which can be considered evidence against the random-effects model. A Hausman test reaches the same conclusion [$\chi^2(2) = 58.63$]. How can we explain the differences of between and within effects? Smoking is likely to be correlated with other mother-specific unobserved variables (Abrevaya 2006) that adversely impact birth weight. Therefore, the between effect (and the estimate from the random-intercept model, which is a weighted average of the between and the within estimate) overestimates the effect of smoking. Certainly, this raises the question whether there is a meaningful interpretation of the between effect. The estimate is obviously biased because it is confounded with the level 2 error. However, a comparison with the within estimate can inform us how much of the observed relation in birth weight and a mother's smoking is due to unobserved heterogeneity between smoking and nonsmoking mothers, which is not accounted for by our model.

3. The estimated standard errors differ slightly because the data used here are unbalanced (Allison 2009, 27).

Table 1. Random-effects, fixed-effects, hybrid, and correlated random-effects linear regression models for birth weight data

	Model 1: random effects	Model 2: fixed effects	Model 3: hybrid	Model 4: correlated re	Model 5: hybrid, random slope
`smoke`	−249.06	−105.70		−105.70	
	(17.42)	(29.53)		(29.52)	
`mage`	10.70	23.12		23.12	
	(1.20)	(3.05)		(3.05)	
`dsmoke`			−105.70		−106.08
			(29.52)		(32.33)
`dmage`			23.12		23.10
			(3.05)		(3.05)
`black`	−255.03		−260.03	−260.03	−260.03
	(26.66)		(26.62)	(26.62)	(26.56)
`msmoke`			−332.93	−227.23	−332.92
			(21.53)	(36.54)	(21.48)
`mmage`			7.52	−15.59	7.52
			(1.31)	(3.32)	(1.30)
Sqrt. Variance: `dsmoke`					234.96
Sqrt. Variance: Level 2	341.66	442.92	341.71	341.66	341.32
Sqrt. Variance: Level 1	374.73	374.73	374.73	374.73	372.47
`xtmixed_sigma_u`					
`xtmixed_sigma_e`					
`xtmixed_N_g`					
Level 2: Mothers	3978	3978	3978	3978	3978
Level 1: Infants	8604	8604	8604	8604	8604

Standard errors in parentheses
Constant omitted

As argued above, the estimated effects of the cluster means will differ between the hybrid model and the correlated random-effects model. In the correlated random-effects model, this is $\pi = \beta_3 - \beta_1$, as shown in (5). Indeed, the estimated effect of -227.23 for the cluster mean from the correlated random-effects model (model 4) corresponds exactly to the difference of the estimated within-mother and between-mother effects from the hybrid model (model 3): $-332.93 - (-105.70) = -227.23$. The test of the null hypothesis that the difference of within and between estimates is equal to 0 provides the same results as for the hybrid model [smoking: Wald $\chi^2(1) = 38.68$, age: Wald $\chi^2(1) = 22.09$].

```
. estimates restore corr_re
. test msmoke
. test mmage
```

As columns 4 and 5 in table 1 show, the hybrid model and the correlated random-effects model also provide (identical) effect estimates of the level 2 variable `black`. The estimated effect of this variable is similar, albeit not identical, to the one obtained from the random-intercept model (model 1). This is because including $\overline{x}_i$ controls more encompassingly for between-cluster differences in x_{it}.

As pointed out above, a decomposition into between and within effects also allows us to incorporate random slopes. Let us assume we have reasons to believe that the within effect of smoking varies across mothers. In that case, we would want to specify a hybrid model with a random slope for the variable `dsmoke`. This can be done with `xtmixed`.

```
. xtmixed birwt dsmoke dmage msmoke mmage black || momid: dsmoke
```

The results are shown in the last column of table 1 (model 5).

4 Interactions

There are pitfalls to the application of these models when including interactions. Let us say we are interested in including an interaction of smoking by mother's age. Thus our comparison model is the following fixed-effects model:

$$\begin{aligned}\left(\texttt{birwt}_{it} - \overline{\texttt{birwt}}_i\right) &= \beta_1 \left(\texttt{mage}_{it} - \overline{\texttt{mage}}_i\right) + \beta_2 \left(\texttt{smoke}_{it} - \overline{\texttt{smoke}}_i\right) \\ &+ \beta_3 \left(\texttt{smoke}_{it}\texttt{mage}_{it} - \overline{\texttt{smoke}_i\texttt{mage}_i}\right) + (\epsilon_{it} - \overline{\epsilon}_i) \end{aligned} \tag{6}$$

We can fit this easily by using the operator `#` to specify the interaction, the `c.` operator to indicate continuous variables, and `i.` to indicate factor variables.

```
. xtreg birwt i.smoke##c.mage, i(momid) fe
. estimates store fe_inter
```

However, specifying the hybrid model as[4]

```
. xtreg birwt c.dsmoke##c.dmage c.msmoke##c.mmage black, i(momid) re
. estimates store hybrid_inter_incorrect
```

is incorrect. Why is that? What we want to estimate is $\beta_k(x_{it}z_{it} - \overline{x_i z_i})$. But if we specify the interaction in the hybrid model as above, we estimate $\beta_k\{(x_{it} - \overline{x}_i)(z_{it} - \overline{z}_i)\} = \beta_k(x_{it}z_{it} - x_{it}\overline{z}_i - \overline{x}_i z_{it} + \overline{x}_i\overline{z}_i)$, which produces a completely different result. Therefore, we first have to generate the interaction term $x_{it}z_{it}$, cluster mean center the new variable, and then enter it into the model.

```
. generate smokeXmage = smoke*mage
. by momid, sort: center smokeXmage if nonmiss==1, prefix(d) mean(m)
. xtreg birwt dsmoke dmage dsmokeXmage msmoke mmage msmokeXmage black, i(momid) re
. estimates store hybrid_inter_correct
```

Because the correlated random-effects model does not include deviations from the cluster means but does include the variables in their uncentered form, it is possible to include the interaction of these variables with the operator #. However, to obtain the correct within effect of the interaction, we still have to include the respective interaction terms of the cluster means. This, again, cannot be done with the operator #. We have to control for $\overline{x_i z_i}$, but using # instead results in $\overline{x}_i\overline{z}_i$.

The following specification therefore results in incorrect estimates:

```
. xtreg birwt i.smoke##c.mage c.msmoke##c.mmage black, i(momid) re
. estimates store corr_re_inter_incorrect
```

The correct estimates are obtained through the following specification:

```
. xtreg birwt i.smoke##c.mage msmoke mmage msmokeXmage black, i(momid) re
. estimates store corr_re_inter_correct
```

An estimation example is provided in table 2. Model 1 shows the estimates from the standard fixed-effects model. This model estimates the interaction effect of smoking by mother's age as 3.79. Models 2 and 3 present estimates from hybrid models, and models 4 and 5 present estimates from correlated random-effects models. Models 2 and 4 include the incorrect interaction, and models 3 and 5 include the correct interactions. If we compare the estimated interaction effects of models 2 and 4 with the benchmark, model 1, we see that both are incorrect. Moreover, the estimated main effects of models 2 and 4 are, of course, also incorrect. Models 3 and 5, on the other hand, estimate the correct main and interaction effects.[5]

4. Note that group mean-centered variables are continuous.
5. The magnitude of the difference between the correct and the incorrect estimates is considerably larger in the hybrid model than in the correlated random-effects model. This is what we would expect, considering that the difference between the correct and the incorrect interaction terms in the hybrid model, that is, the difference between $(x_{it}z_{it} - \overline{x_i z_i})$ and $\{(x_{it} - \overline{x}_i)(z_{it} - \overline{z}_i)\}$, is larger than the difference in the correlated random-effects model, that is, the difference between $\overline{x_i z_i}$ and $\overline{x}_i\overline{z}_i$.

Table 2. Fixed-effects, hybrid, and correlated random-effects linear regression models with interactions for birth weight data

	Model 1: fixed effects	Model 2: hybrid, incorrect interaction	Model 3: hybrid, correct interaction	Model 4: corr re, incorrect interaction	Model 5: corr re, correct interaction
`1.smoke`	−205.66			−198.40	−205.66
	(134.01)			(133.04)	(133.94)
`mage`	22.75			22.78	22.75
	(3.09)			(3.09)	(3.09)
`dsmoke`		−103.98	−205.66		
		(29.54)	(133.94)		
`dmage`		23.12	22.75		
		(3.05)	(3.09)		
`1.smoke#c.mage`	3.79			3.52	3.79
	(4.96)			(4.92)	(4.96)
`c.dsmoke#c.dmage`		−43.82			
		(31.12)			
`dsmokeXmage`			3.79		
			(4.96)		
`black`		−258.03	−258.43	−258.57	−258.43
		(26.73)	(26.73)	(26.73)	(26.73)
`msmoke`		−266.03	−265.73	−70.78	−60.07
		(104.91)	(104.70)	(169.16)	(170.00)
`mmage`		7.89	7.93	−14.86	−14.82
		(1.45)	(1.45)	(3.41)	(3.41)
`c.msmoke#c.mmage`		−2.51		−5.91	
		(3.84)		(6.24)	
`msmokeXmage`			−2.52		−6.31
			(3.84)		(6.27)
Sqrt. Variance: Level 2	443.37	341.83	341.72	341.75	341.69
Sqrt. Variance: Level 1	374.75	374.66	374.75	374.75	374.75
Level 2: Mothers	3978	3978	3978	3978	3978
Level 1: Infants	8604	8604	8604	8604	8604

Standard errors in parentheses
Constant omitted

If we want to exploit the possibility of estimating within effects in random-effects models via the hybrid specification, we have to include interaction terms in the old-fashioned way by generating interaction variables. If we use the correlated random-effects setup, we still have to generate interactions of the cluster mean variables by hand.

Unfortunately, in the case of the hybrid model, it becomes impossible to use postestimation commands such as `margins` in the usual manner. `margins` relies on factor-variable notation. If interactions are not specified via `#`, Stata will not automatically take into account both main and interaction effects.

Suppose we are interested in the marginal effect of mother's age, that is, in the partial derivate of (6) with respect to mother's age. Based on the fixed-effects model (model 1) this is 23.28. However, using `margins` after the hybrid model (model 3) gives 22.75, which is only the main effect. Using `margins` after the correlated random-effects model (model 5), on the other hand, gives the correct estimate of 23.28.[6]

```
. estimates restore fe_inter
. margins, dydx(mage)
. estimates restore hybrid_inter_correct
. margins, dydx(dmage)
. estimates restore corr_re_inter_correct
. margins, dydx(mage)
```

5 Conclusion

I discussed some advantages that correlated random-effects and hybrid models offer. In either case, a decomposition of within and between effects in a single model increases flexibility in model setup because it combines advantages of fixed- and random-effects models. It allows us to estimate the effect of level 2 variables while providing effect estimates of level 1 variables that are unbiased by a possible correlation with the level 2 error. Moreover, a comparison of within and between effects (or their difference) provides an assessment of the degree to which unobserved heterogeneity in level 2 characteristics is responsible for an observable relation between the outcome and a level 1 variable, which is not accounted for by our model. Importantly, these advantages apply when handling any clustered data, whether it is panel data or multilevel data.

Yet there are some aspects to consider. First, within-effect estimates obtained through random-effects models are not more efficient than those obtained from fixed-effects models. Second, the approach described here offers no remedy for a possible correlation of a level 2 variable and the level 2 error (this would require an instrumental-variables approach; see Hausman and Taylor [1981] when handling panel data). Third, handling interactions in these models may be cumbersome. Nevertheless, these models are useful extensions to the standard random-effects and fixed-effects approaches.

6. Note that depending on what `margins` is used for, predictions based on fixed-effects models may still differ from those based on correlated random-effects models because the estimated intercepts differ.

6 Acknowledgments

I would like to thank Silvia Maja Melzer, Klaus Pforr, Carsten Sauer, and Peter Valet for their comments and suggestions. I would also like to thank Julia Harand for her help in preparing the manuscript.

7 References

Abrevaya, J. 2006. Estimating the effect of smoking on birth outcomes using a matched panel data approach. *Journal of Applied Econometrics* 21: 489–519.

Allison, P. D. 2009. *Fixed Effects Regression Models*. Thousand Oaks, CA: Sage.

Baltagi, B. H. 2008. *Econometric Analysis of Panel Data*. 4th ed. New York: Wiley.

Burnett, K., and G. Farkas. 2009. Poverty and family structure effects on children's mathematics achievement: Estimates from random and fixed effects models. *The Social Science Journal* 46: 297–318.

Curran, P. J., and D. J. Bauer. 2011. The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology* 62: 583–619.

Epstein, A. J., J. D. Ketcham, S. S. Rathore, and P. W. Groeneveld. 2012. Variations in the use of an innovative technology by payer: the case of drug-eluting stents. *Medical Care* 50: 1–9.

Halaby, C. N. 2003. Panel models for the analysis of change and growth in life course studies. In *Handbook of the Life Course*, ed. J. T. Mortimer and M. J. Shanahan, 503–527. New York: Springer.

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

Hausman, J. A., and W. E. Taylor. 1981. A generalized specification test. *Economics Letters* 8: 239–245.

Jann, B. 2007a. center: Stata module to center (or standardize) variables. Statistical Software Components S444102, Department of Economics, Boston College. http://ideas.repec.org/c/boc/bocode/s444102.html.

———. 2007b. Stata tip 44: Get a handle on your sample. *Stata Journal* 7: 266–267.

Jones, A., N. Rice, T. B. d'Uva, and S. Balia. 2007. *Applied Health Economics*. Abingdon, UK: Routledge.

Kaufman, R. L. 1993. Decomposing longitudinal from cross-unit effects in panel and pooled cross-sectional designs. *Sociological Methods and Research* 21: 482–504.

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.

Landale, N. S., B. K. Gorman, and R. S. Oropesa. 2006. Selective migration and infant mortality among Puerto Ricans. *Maternal and Child Health Journal* 10: 351–360.

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.

Nomaguchi, K. M., and S. L. Brown. 2011. Parental strains and rewards among mothers: The role of education. *Journal of Marriage and Family* 73: 621–636.

Ousey, G. C., and P. Wilcox. 2007. The interaction of antisocial propensity and life-course varying predictors of delinquent behavior: Differences by method of estimation and implications for theory. *Criminology* 45: 313–354.

Park, C. Y., M. A. Lee, and A. J. Epstein. 2009. Variation in emergency department wait times for children by race/ethnicity and payment source. *Health Services Research Journal* 44: 2022–2039.

Phillips, J. A. 2006. Explaining discrepant findings in cross-sectional and longitudinal analyses: An application to U.S. homicide rates. *Social Science Research* 35: 948–974.

Rabe-Hesketh, S., and A. Skrondal. 2008. *Multilevel and Longitudinal Modeling Using Stata*. 2nd ed. College Station, TX: Stata Press.

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

———. 1989b. A Response to Longford and Plewis. *Multilevel Modelling Newsletter* 1: 8–11.

Schempf, A. H., J. S. Kaufman, L. C. Messer, and P. Mendola. 2011. The neighborhood contribution to black–white perinatal disparities: An example from two North Carolina counties, 1999–2001. *American Journal of Epidemiology* 174: 744–752.

Teachman, J. 2011. Modeling repeatable events using discrete-time data: Predicting marital dissolution. *Journal of Marriage and Family* 73: 525–540.

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

Zhou, M. 2011. Intensification of geo-cultural homophily in global trade: Evidence from the gravity model. *Social Science Research* 40: 193–209.

About the author

Reinhard Schunck is a research and teaching associate in the Department of Sociology at the University of Bielefeld in Germany. His research interests span multilevel and longitudinal data analysis. He works primarily in the field of social stratification and inequality.