

Consequences of Unmodeled Nonlinear Effects in Multilevel Models

Daniel J. Bauer

University of North Carolina at Chapel Hill

Li Cai

University of California, Los Angeles

Applications of multilevel models have increased markedly during the past decade. In incorporating lower-level predictors into multilevel models, a key interest is often whether or not a given predictor requires a random slope, that is, whether the effect of the predictor varies over upper-level units. If the variance of a random slope significantly differs from zero, the focus of the analysis may then shift to explaining this heterogeneity with upper-level predictors through the testing of cross-level interactions. As shown in this article, however, both the variance of the random slope and the cross-level interaction effects may be entirely spurious if the relationship between the lower-level predictor and the outcome is nonlinear in form but is not modeled as such. The importance of conducting diagnostics to detect nonlinear effects is discussed and demonstrated via an empirical example.

Keywords: *multilevel models; hierarchical linear models; mixed models; nonlinear effects; spurious effects*

Multilevel models are commonly used to model data with two or more levels of sampling. We generically refer to the case of two-level data as consisting of observations within independent sampling units (ISUs). For instance, one might be interested in modeling variability in the academic achievement of students (observations) sampled from many different high schools (ISUs). In the simplest multilevel models, there is one random residual term associated with each level of sampling, for instance, one term capturing unexplained variability within ISUs and another capturing unexplained variability between ISUs. In more complex multilevel models, the effect of one or more predictors may also be specified as random, indicating that the magnitude of the effect varies over ISUs. One may then seek to explain this heterogeneity on the basis of observed ISU-level predictors. Moderation effects of this sort are modeled as product interaction terms, known as *cross-level interactions*, between predictors that vary at the observation level and those that vary at the ISU level.

Correspondence concerning this article should be addressed to Daniel J. Bauer, Psychology Department, University of North Carolina, Chapel Hill NC 27599-3270.

As is well known, specification errors of various kinds can affect the inferences made with multilevel models. In particular, much research has been done to evaluate the impact of misspecifying the random effects of the model (e.g., Berkhof & Kampen, 2004; Moerbeek, 2004). In addition, major textbooks on multilevel modeling (e.g., Goldstein, 2003; Hox, 2002; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002; Snijders & Bosker, 1999) emphasize the use of diagnostic procedures for checking distributional assumptions (e.g., normality, homoscedasticity) and identifying outlying values. Little attention, however, has been focused on the detection of nonlinear effects (for a recent exception, see Snijders & Berkhof, 2008), nor has much consideration been given to the possible consequences of failing to model nonlinear trends, perhaps on the assumption that, if the relationship is at least monotonic, not much information will be lost by fitting a linear approximation. To the contrary, we demonstrate here that the failure to adequately identify and model even relatively small nonlinear trends for lower-level predictors can lead to the estimation of spurious random slopes and cross-level interactions.

1. Spurious Random Slopes

As a case-in-point, let us posit that there is a predictor X and an outcome Y , both of which vary across observations within ISUs. We will also assume as a convenience that X is a stochastic predictor that is normally distributed within ISUs such that

$$X_{ij} = \mu_j + e_{ij}, \tag{1}$$

where $e_{ij} \sim N(0, \phi)$ and $\mu_j \sim N(\mu, \psi)$. That is, X varies within ISUs and the mean of X also varies across ISUs. Now, let us suppose that the effect of X on Y is well approximated by a quadratic function. Using the notation of Raudenbush and Bryk (2002), we may then write the equation for the model in mixed-model form as

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{20}X_{ij}^2 + u_{0j} + r_{ij}. \tag{2}$$

As is customary, $r_{ij} \sim N(0, \sigma)$, $u_{0j} \sim N(0, \tau_{00})$, $\text{CORR}(r_{ij}, u_{0j}) = 0$, $\text{CORR}(X_{ij}, r_{ij}) = 0$ and $\text{CORR}(X_{ij}, u_{0j}) = 0$. In total, we will refer to Equations 1 and 2 as the *generating model*.

Now, let us suppose that the *fitted model* fails to include the quadratic effect. Instead, the specified model includes a random slope for the effect of X on Y . Using a superscript * to differentiate the parameters of the fitted model from the corresponding parameters of the generating model, the fitted model is

$$Y_{ij} = \gamma_{00}^* + \gamma_{10}^*X_{ij} + u_{0j}^* + u_{1j}^*X_{ij} + r_{ij}^*, \tag{3}$$

where $r_{ij}^* \sim N(0, \sigma^*)$ and

$$\begin{pmatrix} u_{0j}^* \\ u_{1j}^* \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00}^* & \\ & \tau_{11}^* \end{pmatrix} \right] \quad (4)$$

We may now ask the question, how would this misspecification of the fitted model affect the parameter estimates? Solving for the fixed effects, we obtain the following results:¹

$$\gamma_{00}^* = \gamma_{00} + \gamma_{20}(\phi - \psi - \mu^2) \quad (5)$$

$$\gamma_{10}^* = \gamma_{10} + 2\gamma_{20}\mu \quad (6)$$

In addition, the covariance parameters for the random effects are

$$\tau_{00}^* = \tau_{00} + \gamma_{20}^2(4\mu^2\psi + 2\psi^2) \quad (7)$$

$$\tau_{11}^* = 4\gamma_{20}^2\psi \quad (8)$$

$$\tau_{01}^* = -4\gamma_{20}^2\psi\mu \quad (9)$$

Of particular interest is the spurious slope variance, τ_{11}^* , which can be seen to be a simple function of the degree of curvature of the generating function and the variability in the ISU means for X . Figure 1 illustrates these results graphically, showing apparent slope heterogeneity because of the omission of a quadratic effect. Note the decreasing steepness of the ISU-specific linear trends as the ISU location shifts from left to right.

To determine the degree of nonlinearity required to obtain a statistically significant spurious slope variance, we conducted a small simulation study. We manipulated three primary factors in data generation: degree of curvature, number of ISUs (50, 100, 200, or 300), and ISU size (5, 10, 25, or 50 observations in each ISU). The generated data sets were balanced on ISU size. There were two curvature conditions. In the first condition, the values for γ_{00} , γ_{10} , and γ_{20} were chosen to be 10, 2, and -0.025 , respectively. In the second condition, γ_{20} was set to -0.10 . These values were selected to produce modest curvature and to ensure that the quadratic function stays monotonic over the observed range of X . Visually, the larger of the two values for γ_{20} results in the curve depicted in Figure 1. We set $\mu = 0$, $\tau_{00} = 2$, $\sigma = 10$, $\psi = 3$, and $\phi = 17$ in all cells of the design, whereby the choice of ψ and ϕ results in an intraclass correlation coefficient (ICC) of 0.15 in X .² For each cell of the $2 \times 4 \times 4$ factorial design, we simulated 500 data sets and then fitted the linear model in Equation 3 by full maximum likelihood using an expectation-maximization (EM) algorithm (Laird, Lange, & Stram, 1987).

In examining the relationship between our analytical derivations and the results of the simulation, one unexpected finding emerged. Specifically, the number of observations per ISU exerted a strong effect on the obtained estimates, as can be seen in Table 1. Relative to the analytical results, the intercept variance tended to

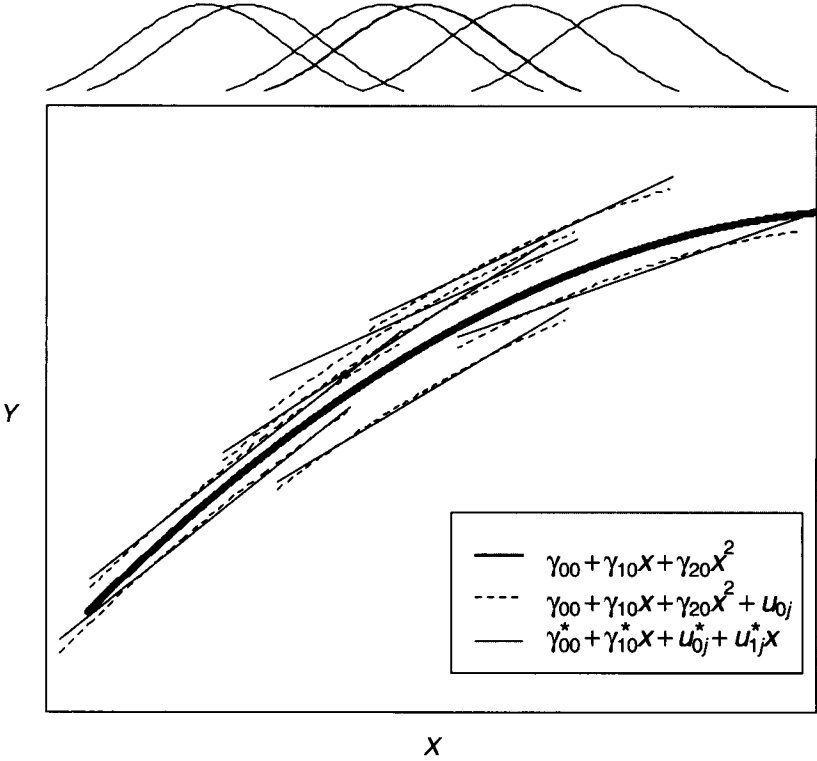


FIGURE 1. Nonlinear generating model: $\gamma_{00} = 10$, $\gamma_{10} = 2$, and $\gamma_{20} = -0.10$.

be too small, particularly as the ISU size decreased. In contrast, the slope variance was larger than expected when the ISU size was small. However, the correspondence between the empirical and analytical results became quite close as the ISU size increased.³ More important, the pattern of effects was entirely consistent with the analytical derivations: The estimated random slope variance increased with the degree of curvature. The fact that the slope variance estimate was also larger when the ISU size was small raised the possibility that this spurious effect would be judged substantively meaningful. We looked into the latter issue by plotting empirical power curves for the likelihood ratio test (LRT) of the significance of the slope variance.

To conduct the LRT, a reduced model having the same fixed effects as the fitted model but with only a random intercept was estimated for each replication. Following Stram and Lee (1994), a critical value from a 50–50 mixture of one and two degrees of freedom chi-squared distributions was used to determine the significance of the LRT statistic. Power curves for the LRT are plotted in Figure 2. Two findings are clearly conveyed here. First, as shown in the panel on

TABLE 1
Results for Spurious Slope Variance (Collapsed Over # Independent Sampling Unit [ISU])

γ_{20}	ISU Size	Fixed Effects		Covariance Parameters		
		$\hat{\gamma}_{00}^*$	$\hat{\gamma}_{10}^*$	$\hat{\tau}_{00}^*$	$\hat{\tau}_{01}^*$	$\hat{\tau}_{11}^*$
-0.025	5	9.55	2.00	1.91	0.00	0.020
	10	9.54	2.00	1.96	0.00	0.014
	25	9.55	2.00	1.96	0.00	0.010
	50	9.58	2.00	1.98	0.00	0.009
	Analytic results	9.65	2.00	2.01	0.00	0.008
-0.10	5	8.62	2.00	1.89	0.00	0.222
	10	8.58	2.00	1.98	0.00	0.183
	25	8.58	2.00	2.08	0.00	0.151
	50	8.58	2.00	2.11	0.00	0.137
	Analytic results	8.60	2.00	2.18	0.00	0.120

the right, when the extent of nonlinearity equaled that shown in Figure 1 (i.e., our greatest curvature condition), the power to detect the spurious slope variance for the linear effect of X approached 1 at essentially all sample sizes. Second, the left panel shows that the power to detect the spurious slope variance was less, but still nontrivial, when there was less curvature present, especially with larger ISU sizes.

Overall, the results suggest that ignoring even relatively minor nonlinear effects may lead to the identification of significant, but spurious, random slope variance. One example in which this occurs may be found in Snijders and Bosker (1999, pp. 112–113). A second example is provided later in this article. We next consider the implications of including specific ISU-level covariates in the model to try to explain this apparent heterogeneity in the effect of X .

2. Spurious Cross-Level Interactions

Let us now consider the generating model

$$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{10}X_{ij} + \gamma_{20}X_{ij}^2 + u_{0j} + r_{ij}, \tag{10}$$

where X is generated as before, $W_j \sim N(\nu, \lambda)$ is an ISU-level predictor, $\text{CORR}(W_j, \mu_j) = \rho$, and W (like X) is uncorrelated with r and u_0 .

We will similarly augment our fitted model by supposing that W has been identified as a candidate moderator of the effect of X :

$$Y_{ij} = \gamma_{00}^* + \gamma_{01}^*W_j + \gamma_{10}^*X_{ij} + \gamma_{11}^*W_jX_{ij} + u_{0j}^* + u_{1j}^*X_{ij} + r_{ij}^*. \tag{11}$$

The term γ_{11}^* is the cross-level interaction of interest.

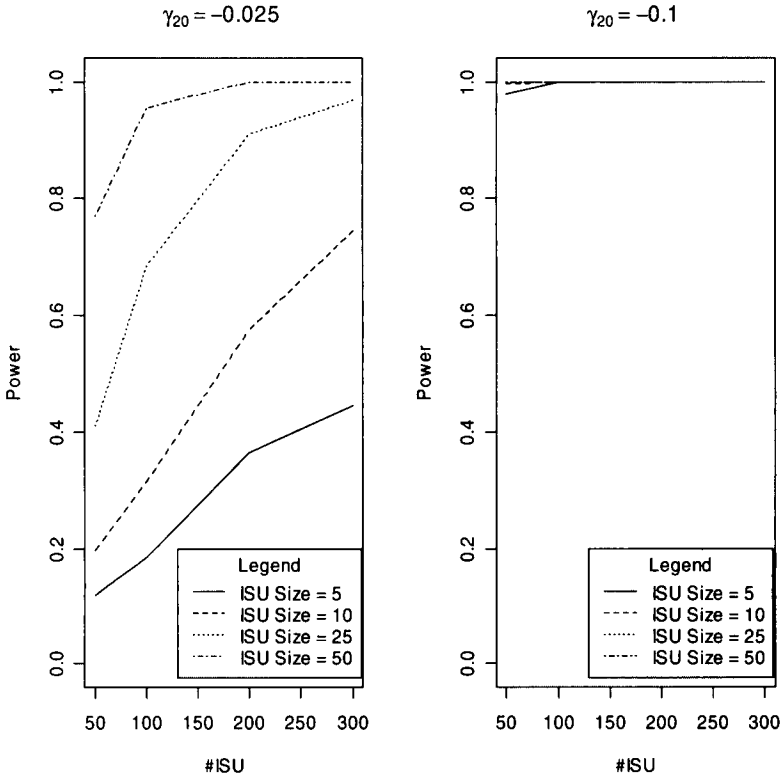


FIGURE 2. Power of the likelihood ratio test of the random slope.
 Note: ISU = independent sampling unit.

Solving for the cross-level interaction, we obtain the expression

$$\gamma_{11}^* = \frac{2\gamma_{20}\rho\sqrt{\Psi}}{\sqrt{\lambda}}. \tag{12}$$

Thus, the magnitude of the spurious cross-level interaction effect is a direct function of the omitted quadratic effect, γ_{20} , and the correlation between W and the ISU means for X , or ρ (Ψ and λ serve as little more than scaling factors). This result is quite intuitive if we consider that the variance in the random slopes was completely determined by the ISU mean differences in X when W was not included in the model, as shown in Equation 8. The correlation of W with the ISU means of X allows W to serve as an imperfect proxy for the ISU means, producing the spurious cross-level interaction effect.

Similarly, if we solve for the variance of the random slope ($\text{VAR}(u_{1j}^*) = \tau_{11}^*$) for the fitted model we obtain

$$\tau_{11}^* = 4\gamma_{20}^2\psi(1 - \rho^2). \quad (13)$$

Comparing this expression to what we obtained earlier in Equation 8 for the model excluding W , we can see that the variance of the random slope has been reduced by a factor of precisely $1 - \rho^2$. Again, W explains variance in the ISU slopes to the extent that it is correlated with the ISU means for X .⁴

We again conducted a small simulation study to determine the conditions under which the spurious cross-level interaction effect might attain statistical significance. As before, we varied the magnitude of the nonlinear effect ($\gamma_{20} = -.025$ vs. $\gamma_{20} = -.10$), the number of ISUs (50, 100, 200, and 300), and size of each ISU (5, 10, 25, and 50). In addition, we set γ_{01} to 1 for all conditions and manipulated the size of the correlation ρ between W and X . For the low correlation condition, $\rho = .3$, and for the high correlation condition, $\rho = .6$. All other parameters were held at the same values used previously. The ISU means of X and W were drawn from a bivariate normal distribution with zero means, equal variances ($\lambda = \psi$), and correlation ρ . For each cell of the $2 \times 4 \times 4 \times 2$ design, 500 replications were simulated, and the model in Equation 11 was fit using the same EM algorithm as before (Laird et al., 1987).⁵

The empirical results are shown in Table 2, with the analytically derived expected values noted for comparison. Some discrepancies between the empirical results and analytical values were again obtained at smaller ISU sizes. Nevertheless, the overall pattern of results became quite consistent with the analytical predictions at larger ISU sizes, showing that the magnitude the spurious cross-level interaction effect depends on the size of the omitted nonlinear effect and the correlation between W and the ISU means for X .

We next considered the power to detect the spurious cross-level interaction by the Wald test, one of the most commonly used tests for fixed effects. Figure 3 displays the proportion of replications reaching statistical significance by condition. Overall, the findings are predictable: Power increases with sample size (ISUs and number of observations within ISU) and with effect size (as determined by the magnitude of γ_{20} and ρ). What is of most interest, however, is the fact that the power is nontrivial at most sample sizes, even when the omitted nonlinear effect of X on Y is relatively weak. When the correlation between W and the ISU means for X is high, the power is substantial, even at relatively low sample sizes.

3. Identifying Nonlinear Effects

As shown above, failing to model the nonlinear effects of lower-level predictors may result in the identification of spurious random slopes and spurious cross-level interactions under a variety of realistic conditions. These problems could be ameliorated if, through careful diagnostics, data analysts could identify

TABLE 2
Results for Cross-Level Interaction (Collapsed Over # Independent Sampling Unit [ISU])

γ_{20}	ρ	ISU Size	Fixed Effects				Slope Variance
			$\hat{\gamma}_{00}^*$	$\hat{\gamma}_{10}^*$	$\hat{\gamma}_{01}^*$	$\hat{\gamma}_{11}^*$	$\hat{\tau}_{11}^*$
-0.025	.3	5	9.55	2.00	1.00	-0.014	0.019
		10	9.55	2.00	1.00	-0.015	0.013
		25	9.55	2.00	1.00	-0.015	0.009
		50	9.57	2.00	1.00	-0.015	0.008
		Analytic results	9.65	2.00	1.00	-0.015	0.007
-0.025	.6	5	9.58	2.00	1.00	-0.029	0.016
		10	9.58	2.00	1.00	-0.029	0.011
		25	9.58	2.00	1.00	-0.029	0.007
		50	9.59	2.00	1.00	-0.030	0.006
		Analytic results	9.65	2.00	1.00	-0.030	0.005
-0.10	.3	5	8.62	2.00	1.00	-0.051	0.213
		10	8.58	2.00	1.00	-0.054	0.173
		25	8.57	2.00	1.00	-0.057	0.140
		50	8.58	2.00	1.00	-0.058	0.126
		Analytic results	8.60	2.00	1.00	-0.060	0.109
-0.10	.6	5	8.65	2.00	1.00	-0.103	0.186
		10	8.60	2.00	1.01	-0.110	0.146
		25	8.58	2.00	1.00	-0.116	0.112
		50	8.58	2.00	1.00	-0.118	0.095
		Analytic results	8.60	2.00	1.00	-0.120	0.077

these effects and modify their models accordingly. Several popular texts on multilevel modeling, however, provide little or no guidance on diagnosing nonlinear effects (Bickel, 2007; Goldstein, 2003; Heck & Thomas, 2000; Kreft & de Leeuw, 1998; Luke, 2004; Twisk, 2006). Raudenbush and Bryk (2002, p. 258) and Hox (2002) both mention that smoothed residual plots could be used to detect nonlinear trends, whereas Snijders and Bosker (1999, pp. 128–131) provide the most extended treatment of the subject. In contrast, Goldstein (2003) and Luke (2004) demonstrate residual plots, but primarily for the purpose of detecting heteroscedasticity.

In agreement with Raudenbush and Bryk (2002) and Snijders and Bosker (1999), we suggest using residual plots to identify potentially omitted nonlinear effects. Specifically, residuals can be calculated and plotted against the predicted values to judge global misspecifications of the model. If the model includes multiple predictors, the residuals can also be plotted against each lower-level predictor to localize the source of the specification error. We use empirical Bayes's residuals

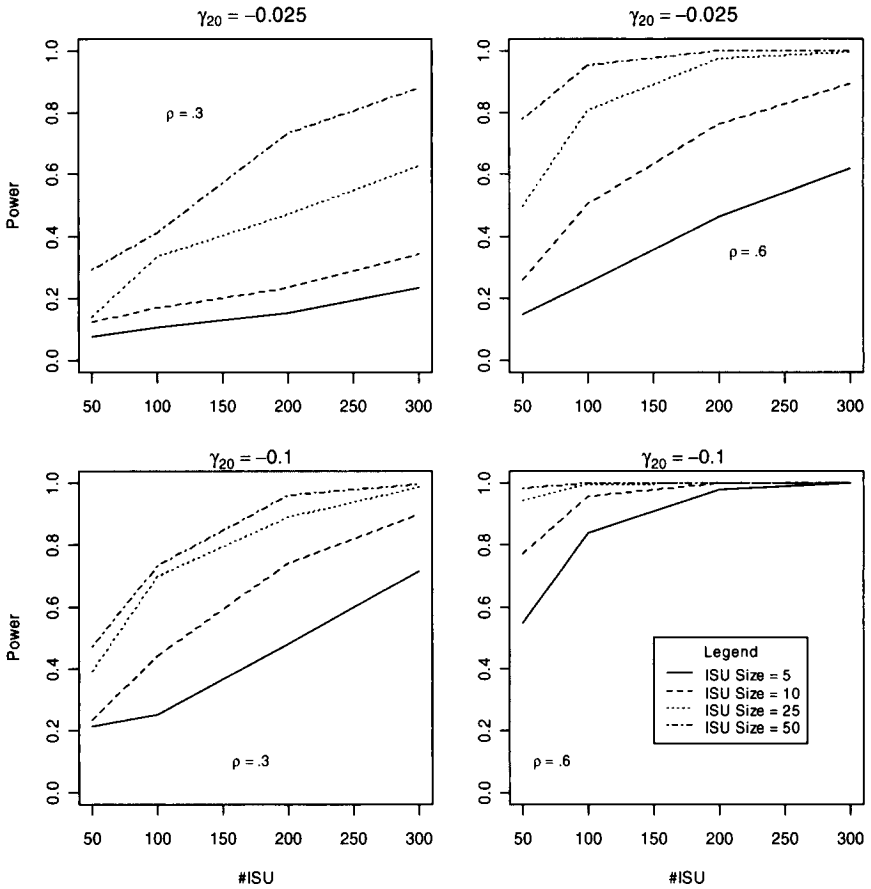


FIGURE 3. Power of the Wald test of the cross-level interaction.
 Note: ISU = independent sampling unit.

(defined by Raudenbush & Bryk, 2002, p. 50, Equation 3.62) for this purpose, although least squares residuals could also be examined (Snijders & Bosker, 1999, p. 131).

As an example, we examine a single simulated data set from our first simulation study ($\gamma_{20} = -.025$; ICC for X .15, 100 ISUs, 50 observations per ISU). A linear model including random slopes was fit to the data, and the empirical Bayes's residuals were obtained as

$$\hat{r}_{ij}^* = Y_{ij} - \hat{Y}_{ij} \tag{14}$$

$$= Y_{ij} - \hat{\gamma}_{00}^* + \hat{\gamma}_{10}^* X_{ij} + \hat{u}_{0j}^* + \hat{u}_{1j}^* X_{ij}, \tag{15}$$

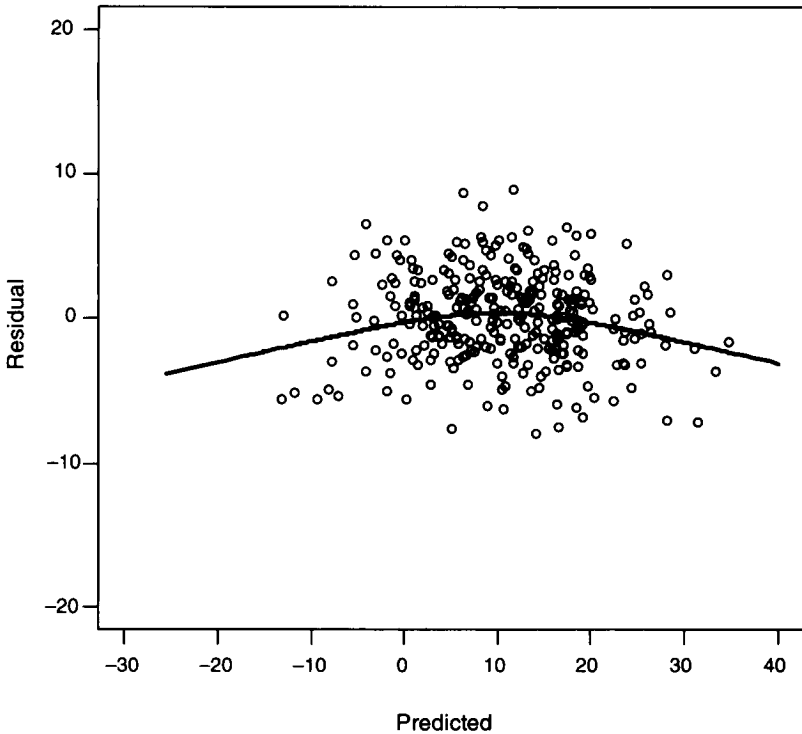


FIGURE 4. Residual plot for simulated data (300 randomly selected cases out of 5,000).

where \hat{u}_{0j}^* and \hat{u}_{1j}^* are the empirical Bayes's estimates of the random effects. Despite the erroneous inclusion of the random slope effect, the omitted nonlinear trend should still be detectable in residual plots (e.g., the curvature around each ISU-specific regression line in Figure 1). Indeed, Figure 4 shows a slight but noticeable nonlinear trend in the relationship of the residuals to the predicted values, highlighted by superimposing a LOWESS curve on the plot (Cleveland, 1981).⁶ Adding X^2 to the model would not only remove this nonlinear trend from the residual plot, indicating a correctly specified model, but also render superfluous the spurious random slope for X . Several approaches to modeling more complex nonlinear trends are described in Snijders and Berkhof (2008).

4. Empirical Example

In this section, we provide an empirical demonstration of both the potential consequences of failing to detect nonlinear effects and the application of diagnostics useful for identifying nonlinear effects early on in an analysis. For sake

of brevity, we do not provide details on other diagnostics that would normally be conducted to evaluate other standard assumptions of the multilevel model.

The data we analyze were gathered in 1980 from the original sophomore cohort of the High School and Beyond Study (U.S. Department of Education, National Center for Education Statistics, 1980). Our interest is in predicting students' mathematic achievement test scores ($M = 12.78$, $SD = 9.83$) using their annual family income as a Level 1 predictor and school sector (public vs. private) as a Level 2 predictor. Family income was measured on a 7-point graded scale (<7,000, 7,000–11,999, 12,000–15,999, 16,000–19,999, 20,000–24,999, 25,000–37,999, >38,000) and was standardized for the analysis.⁷ The proportion of private schools was 14.7% in the sophomore sample. After deleting cases with missing data, there were 22,672 students nested in 971 schools. All models were fit using SAS PROC MIXED with the maximum likelihood estimator.

We began our analysis by fitting models in which the effect of family income on math achievement was specified as linear. In Model 1, family income was included as the sole predictor of math achievement. Both the intercept and slope terms were allowed to randomly vary across schools, and their covariance was freely estimated. The results are reported in the first column of Table 3. As expected, the fixed effect of family income on math achievement scores was positive. In addition, the variance of the random slope for family income was also statistically significant, indicating that family income is more strongly related to math achievement in some schools than in others. One possible source of these differences is whether the school is public or private. To evaluate this possibility, Model 2, presented in the second column of Table 3, incorporated a main effect of school sector and Sector \times Income interaction. Following the recommendations of Bauer and Curran (2005), we graphically evaluated the significant cross-level interaction by plotting the model-implied simple regression lines for public and private schools in the first panel of Figure 5. As can be seen, the results suggest that family income is a stronger predictor of math achievement scores in public than in private schools.

We next conducted model diagnostics to determine whether family income might exert a nonlinear effect on math achievement. In Figure 6, the empirical Bayes's residuals obtained from Model 1 are plotted against the predicted values (top panel) and against the observed values of family income (bottom panel). As family income has only seven possible values and there are many observations observed for each value, the bottom panel replaces the traditional scatterplot with a side-by-side boxplot.⁸ Both panels display smoothed regression lines (dashed), which are suggestive of slight nonlinear trends. Given our earlier finding that even small omitted nonlinear trends could produce spurious effects, particularly at the sample sizes present here, we next fit several alternative models to the data that included a quadratic effect of family income. A quadratic model seemed reasonable given that, in both residual plots, a quadratic regression line (solid) reproduced the smoothed regression line fairly well.

(text continues on p. 111)

TABLE 3
Analysis of High School and Beyond Data

	Model 1	Model 2	Model 3	Model 4
Fixed effects				
Intercept	12.48****	15.14****	13.19****	15.94****
Income	1.83****	1.25****	1.82****	1.45****
Income × Income			-0.65****	-0.64****
Public		-3.06****		-3.17****
Income × Public		0.65****		0.41*
Dispersion parameters				
Var(Intercept)	15.79****	15.03****	16.08****	15.08****
Var(Income)	0.53****	0.47**		
Cov(Intercept, income)	1.17**	1.29****		
Level 1 residual	73.75****	73.72****	73.74****	73.70****
Fit statistics				
-2 log likelihood	163650.2	163581.5	163555.1	163493.7
Akaike information criterion	163662.2	163597.5	163565.1	163507.7
Bayesian information criterion	163691.5	163636.6	163589.5	163541.8

Note: * $p < .05$. ** $p < .01$. *** $p < .001$. **** $p < .0001$.

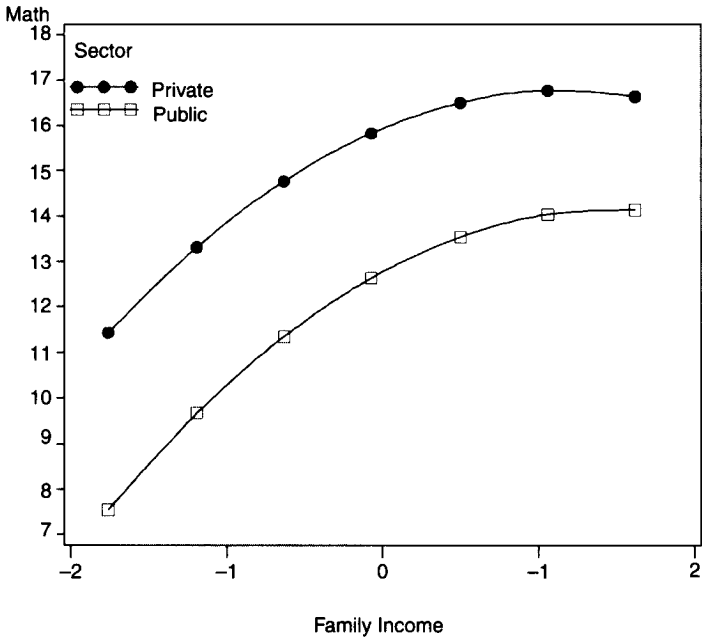
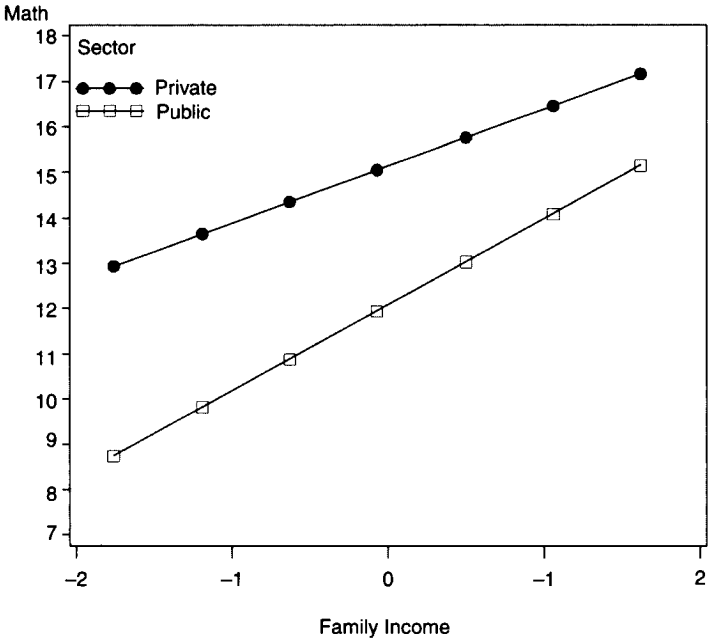


FIGURE 5. Simple regression slopes.

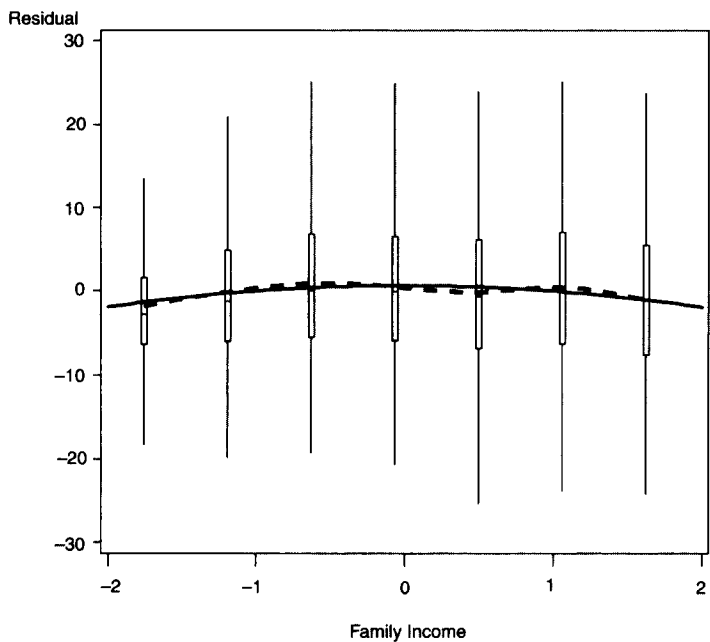
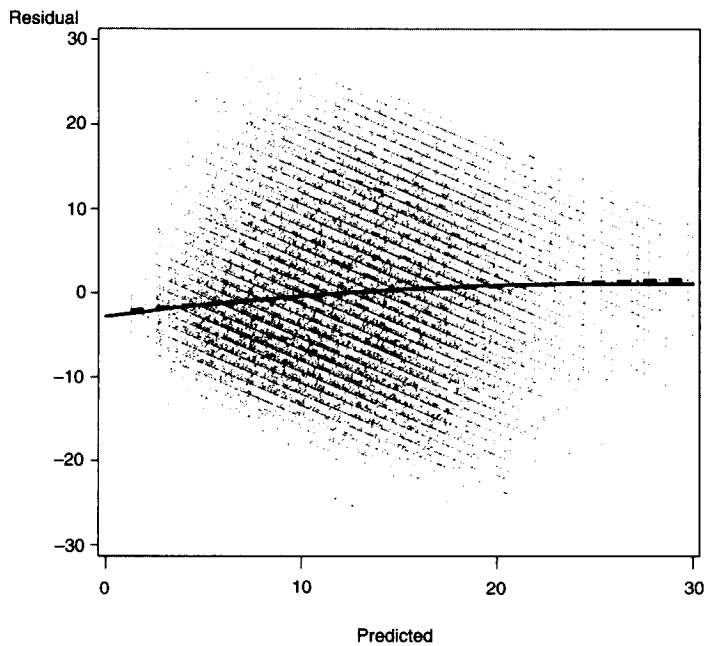


FIGURE 6. Residual plots for High School and Beyond data.

On the possibility that the random linear effect of family income might be spurious, we fit quadratic models that included only a random intercept at the school level. That is, both the linear and quadratic trends for family income were fixed. We began by fitting a quadratic model that included only these effects and not sector effects (Model 3) for comparison to Model 1. The results are reported in the third column of Table 3. The overall pattern of effects suggests that math achievement is more strongly related to family income at lower income levels and becomes a less relevant predictor at higher income levels, consistent with an asymptotic function. Also noteworthy is that the log likelihood for Model 3 is in fact smaller than the log likelihood for Model 1, despite the fact that Model 1 has one more parameter than Model 3. The Bayesian information criterion and Akaike information criterion also favor Model 3. We next reevaluated possible sector effects, including both a main effect of sector and a cross-level interaction of sector with family income (Model 4). Comparing Models 2 and 4 in Table 3, it can be seen that the cross-level interaction effect between family income and sector is still significant in Model 4, but the effect size and p value are reduced relative to Model 2. This provides evidence that the cross-level interaction in Model 2 was not entirely spurious, but the effect may have been inflated by the omission of the quadratic effect for family income. To show the difference between the two models, the second panel of Figure 5 plots the simple regression lines for public and private schools as implied by Model 4. As can be seen, though the cross-level interaction remains significant, the differences between the two simple regression lines are not as pronounced as they were in the linear model.

The conclusions we would reach from the linear and quadratic models are quite different. The linear model implies that private schools are more equitable, showing a tangibly smaller achievement gap between affluent and impoverished students than in public schools. In contrast, the quadratic model suggests that this gap is present to about the same degree in both sectors: Students in private schools score higher than do students in public schools by about the same amount at roughly all income levels. Our earlier analytical and simulation results suggest that the stronger cross-level interaction effect found in the linear model may have been a consequence of failing to model the nonlinear effect of family income on achievement (coupled with the fact that students in private schools are generally more affluent than are students in public schools).⁹

5. Conclusions

The intent of this article was to call attention to the importance of evaluating potential nonlinear effects when fitting multilevel models. We demonstrated that the omission of even relatively minor nonlinear effects of lower-level predictors could lead to the identification of spurious random slopes and cross-level interactions. Given these results, greater emphasis should be placed on the evaluation

of this type of specification error with multilevel models. We have shown one simple method that can be readily implemented with most currently available multilevel modeling software. Our primary recommendation is that diagnostics like these be conducted routinely in the course of any multilevel analysis (in addition to diagnostics for assessing other key model assumptions).

Notes

1. Further detail on the derivation of these results may be found in an online appendix available at <http://www.unc.edu/~dbauer/publications.html>.

2. Keeping the total variance of X constant, we also examined conditions with much higher intraclass correlation coefficient (ICC; $\psi = 10$ and $\phi = 10$), as often occurs with repeated measures data. The findings are similar and may be obtained at <http://www.unc.edu/~dbauer/publications.html>.

3. One reviewer offered the interesting suggestion that this pattern of results may be due to sampling variability in the distribution of X (particularly, its variance), which would be expected to be greater for small ISU sizes.

4. More detail on these derivations, and complete results for all of the parameters of the fitted models, is available at <http://www.unc.edu/~dbauer/publications.html>.

5. The results under higher ICC for X are available at <http://www.unc.edu/~dbauer/publications.html>.

6. If the fitted model had included only the fixed linear effect of X (and not the random slope for X), then this nonlinear trend would have been even more apparent.

7. We recognize that this does not represent an interval-level variable. In practice, however, predictors with as many as seven points are often assumed to have linear effects. In the present case, the increasing distances between the higher category thresholds actually make the relationship between income and achievement appear *more* linear than would be the case if income were on an interval scale.

8. Alternatively, one could “jitter” the observed values of family income to make these values distinguishable in a scatterplot (see Cohen, Cohen, West, & Aiken, 2003).

9. Several other models were also considered for these data. A model including both a random slope for the linear effect and a fixed quadratic effect fit well, illustrating that the two types of effects are not mutually exclusive. In addition, we refit Models 1 through 4 allowing for heteroscedasticity in the random effects and residuals across sectors. Public schools displayed significantly greater variances in the random effects than did private schools in each model. The overall pattern of results, however, remained stable. Hence, for expository purposes, we chose to present only the results of the simpler models.

References

- Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research*, *40*, 373–400.
- Berkhof, J., & Kampen, J. K. (2004). Asymptotic effect of misspecification in the random part of the multilevel model. *Journal of Educational and Behavioral Statistics*, *29*, 201–218.
- Bickel, R. (2007). *Multilevel analysis for applied research*. New York: Guilford.
- Cleveland, W. S. (1981). LOWESS: A program for smoothing scatter plots by robust locally weighted regression. *American Statistician*, *35*, 54.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Arnold.
- Heck, R. H., & Thomas, S. L. (2000). *An introduction to multilevel modeling techniques*. Mahwah, NJ: Lawrence Erlbaum.
- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.
- Kreft, I. G. G., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage.
- Laird, N., Lange, N., & Stram, D. (1987). Maximum likelihood computations with repeated measures: Application of the EM algorithm. *Journal of the American Statistical Association*, *82*, 97–105.
- Luke, D. A. (2004). *Multilevel modeling*. Thousand Oaks, CA: Sage.
- Moerbeek, M. (2004). The consequences of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research*, *39*, 129–149.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Snijders, T. A. B., & Berkhof, J. (2008). Diagnostic checks for multilevel models. In H. Goldstein, J. de Leeuw, & W. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 141–176). New York: Springer.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.
- Stram, D., & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, *50*, 1171–1177.
- Twisk, J. W. R. (2006). *Applied multilevel analysis*. Cambridge, UK: Cambridge University Press.
- U.S. Department of Education, National Center for Education Statistics. (1980). *High school and beyond, 1980: A longitudinal survey of students in the United States* (2nd ICPSR version) [Computer file]. Chicago: National Opinion Research Center.

Authors

DANIEL J. BAUER, is associate professor, L.L. Thurstone Psychometric Laboratory, Department of Psychology, University of North Carolina at Chapel Hill, CB #3270, Davie Hall, Chapel Hill, NC 27599–3270; e-mail: dbauer@email.unc.edu. His current

research focuses on models for repeated measures data, including multilevel models and latent variable models.

LI CAI is now an assistant professor of advanced quantitative methods in the Graduate School of Education and Information Studies at the University of California, Los Angeles, 2331 Moore Hall, Los Angeles, CA 90095-1521; e-mail: lcai@ucla.edu. His current research is centered around computational methods for latent variable models.

Manuscript received September 16, 2006

Revision received June 22, 2007

Accepted September 9, 2007