

# Modeling Multilevel Data Structures

**Marco R. Steenbergen** University of North Carolina, Chapel Hill  
**Bradford S. Jones** University of Arizona

Multilevel data are structures that consist of multiple units of analysis, one nested within the other. Such data are becoming quite common in political science and provide numerous opportunities for theory testing and development. Unfortunately, this type of data typically generates a number of statistical problems, of which clustering is particularly important. To exploit the opportunities offered by multilevel data, and to solve the statistical problems inherent in them, special statistical techniques are required. In this article, we focus on a technique that has become popular in educational statistics and sociology—multilevel analysis. In multilevel analysis, researchers build models that capture the layered structure of multilevel data, and determine how layers interact and impact a dependent variable of interest. Our objective in this article is to introduce the logic and statistical theory behind multilevel models, to illustrate how such models can be applied fruitfully in political science, and to call attention to some of the pitfalls in multilevel analysis.

Political scientists frequently work with data measured at multiple levels, for example, individual-level survey data, and aggregate contextual or demographic data. As such, the use of multilevel data structures is common in political science research. Moreover, many theories and hypotheses in political science hinge on the presumption that “something” observed at one level is related to “something” observed at another level. Yet despite the prevalence of cross-level or multilevel data and theories of political behavior, statistical methods to exploit the information found in multilevel data structures have not been widely used.

In this article, we discuss the statistical modeling issues associated with the analysis of multilevel data. Our intention is to provide both an introduction to multilevel modeling and to review the relevant statistical literature on multilevel models. The article is structured as follows. First, we consider the substantive motivation leading to models for multilevel data. Next, we discuss the statistical consequences of ignoring the multilevel structure of data. In the next section, we explain the problems of conventional approaches to analyzing these kinds of data. This is followed by a discussion of the multilevel model, its different variants and model estimation. We then provide an application of a multilevel model, examining the issue of public support for the European Union. We conclude by discussing some general considerations in multilevel modeling.

---

Marco R. Steenbergen is Associate Professor of Political Science, University of North Carolina at Chapel Hill, 308 Hamilton Hall, CB# 3265, Chapel Hill, North Carolina 27599-3265 ([Marco\\_Steenbergen@unc.edu](mailto:Marco_Steenbergen@unc.edu)). Bradford S. Jones is Associate Professor of Political Science, University of Arizona, 315 Social Sciences, Tucson, Arizona 85721 ([bsjones@email.arizona.edu](mailto:bsjones@email.arizona.edu)).

An earlier version of this paper was presented at the 14th Annual Meeting of the Political Methodology Society, Columbus, OH. The authors would like to thank the participants in this meeting, in particular, Gary King and Jeff Lewis, for their helpful comments. We also thank Bill Dixon, Bob Erikson, David Lowery, Stuart Macdonald, Bill Mishler, George Rabinowitz, Michael Sobel, and Jim Stimson for comments on various versions of this paper. Finally, we thank the anonymous reviewers for their very insightful comments.

*American Journal of Political Science*, Vol. 46, No. 1, January 2002, Pp. 218–237

©2002 by the Midwest Political Science Association

## Substantive Motivation

Multilevel data structures involve data that are ordered hierarchically. For example, a study of voting behavior may focus on voters in different places (Jones, Johnston, and Pattie 1992). In this case, there are two levels of analysis, voters and places, with the first level embedded, or nested within the second level. In general, multilevel data structures exist whenever some units of analysis can be considered a subset of other units, and one has data for both subsets. Thus, time may be nested in states (Western 1998) or in individuals (e.g., Steenbergen, Kilburn and Wolak 2001); individuals may be nested in geographic divisions (Charnock 1997; Jones, Johnston, and Pattie 1992; Quillian 1995), institutions, or groups; groups may be nested in organizations, and so forth. A multilevel data structure may comprise two, three, or more levels of data, as long as a hierarchy among the levels can be established.<sup>1</sup>

The goal of multilevel analysis is to account for variance in a dependent variable that is measured at the lowest level of analysis by considering information from all levels of analysis. There are both substantive and methodological reasons for using information from multiple levels of analysis. Here we discuss three of these reasons.<sup>2</sup> First, multilevel data make it possible for researchers to combine multiple levels of analysis in a single *comprehensive model* by specifying predictors at different levels. Because the model spans multiple levels of analysis, it is less likely to suffer from model misspecification than when compared to models comprised of a single level. The need for theoretical explanations spanning multiple levels of analysis arises frequently in political science. For example, Caldeira and Gibson (1995) used both individual-level and institutional factors to account for public support toward the judicial institutions of a particular country.

Second, multilevel analysis allows researchers to explore *causal heterogeneity* (Western 1998). By specifying cross-level interactions, it is possible to determine whether the causal effect of lower-level predictors is conditioned or moderated by higher-level predictors. In other words, is there a single, uniform causal dynamic or does the causal dynamic vary across higher levels of

<sup>1</sup>Within the multilevel modeling approach, methods exist for dealing with non-hierarchical data structures (so-called cross-classifications). These methods are beyond the scope of this paper, but a good discussion can be found in Fienberg (1980) and Goldstein (1995).

<sup>2</sup>We hold off on a discussion of a fourth reason for multilevel analysis, namely borrowing strength in the estimation of the effects of predictors. We return to this matter in the section on model estimation.

analysis? Causal heterogeneity is a concern in much of the political science literature. A good example of this concern is found in contextual analysis of political behavior. The major supposition of contextual analysis is that the “contextual effect [...] arises due to social interaction within an environment” (Huckfeldt and Sprague 1993, 289). This environment—the context—may be spatially defined (e.g., local neighborhoods; see Huckfeldt and Sprague 1987) in terms of “social networks” (e.g., Mackuen and Brown 1987) or in terms of political and social groupings (e.g., Lau 1989). Regardless of the specifics about the environment, however, a common assumption is that environmental factors interact with individual factors to shape political behavior.

Finally, multilevel analysis can provide a test of the generalizability of findings. That is, do findings obtained in one particular context (or time period) also apply to other contexts (or periods)? Generalizability is an issue that arises particularly often in comparative research. Since much of this work involves case studies of single nations or regions, some have argued that the case study approach should be augmented by a comparative method with the goal of testing generalizability (Lijphart 1971; Przeworski and Teune 1970; Rokkan 1966). Multilevel analysis can contribute to these generalizability tests because it allows researchers to explore causal heterogeneity. If the contextual units are randomly sampled, as multilevel methods typically assume, multilevel analysis may have the added benefit that it helps overcome the case selection problems that often plague comparative research (Geddes 1990; King, Keohane, and Verba 1994).

## Statistical Motivation

In addition to substantive motivations, there also is an important statistical motivation for explicitly accounting for multilevel data structures. Specifically, ignoring the multilevel character of data carries significant statistical costs in the form of possibly incorrect standard errors and inflated Type I error rates. To understand these problems, consider a situation in which a researcher has sampled  $N$  individuals from  $J$  contexts. If the researcher ignores the contextual layer in the data—a practice that is sometimes called “naïve pooling” (Burton, Gurrin, and Sly 1998)—then it appears as if there are  $N$  independent observations in the design. But this impression will generally be misleading. If individual-level factors, for example attitudes, are influenced by contextual factors, then individuals sampled from the same context,  $j$ , share

common influences. Hence, the observations in context  $j$  are not truly independent; they are *clustered* and duplicate one another to some extent. As Kreft and De Leeuw put it, "the more individuals share common experiences due to closeness in space and/or time, the more they are similar, or to a certain extent, duplications of each other" (1998: 9).

In terms of statistical models, the duplication of observations violates the assumption that the errors are independent. This assumption underlies the standard models for data analysis used in political science (e.g., analysis of variance (ANOVA) and ordinary least squares (OLS) regression analysis). In the context of multilevel data structures, the correlation between the errors (observations) is referred to as intra-class or cluster correlation. In most cases this correlation is positive, and this will cause the estimated standard errors to be too low and the test-statistics too high (just as in the case of positive auto-correlation in time series analysis).<sup>3</sup> As a result, Type I errors are more frequent, i.e., predictors appear to have a significant effect when in fact they do not.

Barcikowski's (1981) simulation study of the consequences of positive intra-class correlation for Type I errors in ANOVA illustrates this point. The nominal Type I error rate in this study was .05. But Barcikowski found that even a slight intra-class correlation can dramatically increase the real Type I error rate compared to the nominal rate, especially when the number of cases per contextual unit is large. For example, with 100 cases per contextual unit, an intra-class correlation of .01, which is generally considered very small, produced a Type I error rate of .17. With an intra-class correlation of .20, a sample size of 10 per contextual unit is sufficient to inflate the Type I error rate to .28.<sup>4</sup> Positive intra-class correlations thus have the potential of biasing statistical inference. This may happen even when one explicitly incorporates contextual covariates in the model because these covariates may fail to account for all of the dependence between the observations.<sup>5</sup>

<sup>3</sup>Negative intra-class correlations occur less frequently. They imply that observations in the same contextual unit are less similar to each other than they are to observations in other contextual units. One cause of this phenomenon is when there is competition between individuals within the same contextual unit.

<sup>4</sup>See Goldstein (1995) for similar results in the context of a simple regression model in which data clustering is due to contextual variation in the model intercepts.

<sup>5</sup>Further, it is sometimes useful to contemplate the existence of intra-class correlations even when there are no obvious contextual units that are of substantive interest. Often the sampling design of a study brings about intra-class correlations that may not (or should not) be ignored. This is true, for example, in cluster samples (see, e.g., Barnett 1991), which are widely used in political science (e.g., National Election Studies and General Social Survey).

## Conventional Approaches for Multilevel Data

Clustering in multilevel data structures poses a challenge to statistical analysis. In the past, political scientists have sometimes tried to meet this challenge by using two different approaches, both of which are implemented in an OLS regression analysis. Unfortunately, neither approach is satisfactory. The first approach is to absorb contextual or subgroup differences through a series of dummy variables. In models that do not contain a constant, one should include as many dummy variables as there are subgroups. When a constant is present in the model, then one of the subgroups is designated the baseline category and does not receive a dummy variable. In the context of experimental data, this model is often referred to as the analysis of covariance (ANCOVA) model (Kreft and De Leeuw 1998). In the context of pooled cross-sectional data it is sometimes referred to as the least squares dummy variable (LSDV) model (e.g., Dielman 1989; Sayrs 1989; Stimson 1985). Hereafter, we refer to this approach as the dummy variable model.<sup>6</sup>

Dummy variable models are popular in political science for two reasons. First, these models perfectly capture any clustering by subgroups that may exist in the data, since the dummy variables "absorb" the unique variation among the subgroups. Second, these models can be implemented easily within a standard OLS regression framework. Despite these strengths, the dummy variable approach has limitations for the analysis of multilevel data. Dummy variables are only indicators of subgroup differences; they do not explain why the regression regimes for the subgroups are different. As discussed earlier, an important goal of multilevel statistical analysis is to substantively account for causal heterogeneity. On this objective, dummy variable models provide little leverage.

The second approach to modeling multilevel data structures is the inclusion of subgroup level predictors in the analysis. By including such predictors as main effects in a regression model, it is possible to account for subgroup differences in the constant. By interacting subgroup characteristics with predictors measured at a lower level of analysis, it becomes possible to account for differences in the partial slopes of these predictors across subgroups. Because of these interaction terms, this approach is sometimes referred to as interactive modeling.<sup>7</sup>

<sup>6</sup>The dummy variable model can be extended by including interaction terms between the dummy variables and one or more predictors in the model. This makes it possible to model contextual heterogeneity in the partial slopes of the predictors.

<sup>7</sup>This type of modeling strategy has a long history in political science, in large part due to the work of Boyd and Iversen on contextual analysis (1979; see also Friedrich 1982 and Iversen 1991).

The main strength of interactive models is that they allow the analysis of substantively interesting predictors in order to account for causal heterogeneity. Thus, interactive models exploit the theoretical opportunities that multilevel data structures offer. This is a major improvement over dummy variable models. On another dimension, however, interactive models are inferior to dummy variable models. Implicitly, interactive models assume that the subgroup level predictors fully account for subgroup differences. This is so because interactive models incorporate random error only at the lowest level of analysis; at the higher levels of analysis (i.e., subgroups) the error components are assumed to be zero. This is a very strong assumption that will usually prove to be false. Unfortunately, when false, the interactive model does not avoid the statistical issues associated with clustering of the data.

It appears, then, that existing models make trade-offs between the opportunities and challenges of multilevel data structures. Dummy variable models meet the statistical challenges, but they fail to exploit the theoretical opportunities multilevel data offer. Interactive models take advantage of these theoretical opportunities, but they usually fall short on solving the statistical problems. Clearly, both of these conventional approaches have limitations. Unfortunately, combining these approaches to offset their weaknesses is not possible because there are insufficient degrees of freedom to estimate both contextual effects and dummy variables. Fortunately, there is an alternative modeling approach, which we now consider.

## The Multilevel Model

Recent developments in educational research (e.g., Bryk and Raudenbush 1992; De Leeuw and Kreft 1986; Goldstein 1995; Kreft and De Leeuw 1998; Longford 1993; Snijders and Bosker 1999), sociology (e.g., Mason, Wong, and Entwistle 1983) have resulted in the formulation of multilevel or hierarchical models. In the political methodology literature, the pioneering work of Jackson (1992) led to the consideration, development, and presentation of a random coefficients model, which provides the foundation for the models considered here.<sup>8</sup> These models are closely related to several older methodological traditions such as random coefficient models (see Rubin 1950; Swamy 1970; Swamy and Tavlas 1995)

<sup>8</sup>Other political methodologists who have contributed to multilevel analysis include Jones, Johnston, and Pattie (1992), Achen and Shively (1995), Beck and Katz (1995), and Western (1998). Zorn's (2001) work also follows in this tradition.

and variance components analysis (see Searle, Casella, and McCullough 1992).

To fix ideas, we first consider a multilevel model for the case involving two levels of analysis. We refer to these levels as "level-1" and "level-2" and assume the level-1 observations are nested within level-2 units. In applied research, level-1 observations may refer to individuals and level-2 units may refer to contextual units, for example, countries. We start by considering a model with one level-1 and one level-2 predictor, which are observable and have a linear relationship with the level-1 dependent variable (also observable). Later we generalize this presentation to encompass more complex model specifications, although the focus in this paper remains exclusively on linear models.<sup>9</sup>

## A Simple Linear Multilevel Model

We start by positing the following level-1 model:<sup>10</sup>

$$\gamma_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \varepsilon_{ij}. \quad (1)$$

Here  $\gamma_{ij}$  is the level-1 dependent variable for a level-1 unit  $i (=1, \dots, N_j)$  nested in level-2 unit  $j (=1, \dots, J)$ . Further,  $x_{ij}$  is the level-1 predictor and  $\varepsilon_{ij}$  is a level-1 disturbance term. This model is identical to the simple regression model with the important difference that the regression parameters are not fixed but, instead, vary across level-2 units (as indicated by the  $j$ -subscripts on the regression parameters  $\beta$ ). The introduction of these variable coefficients sets multilevel models apart from most other statistical models used in political science.

To see these differences more clearly, we can directly model the variation of the level-1 regression parameters in (1) as a function of level-2 predictors, such that

$$\beta_{0j} = \gamma_{00} + \gamma_{01}z_j + \delta_{0j}, \quad (2)$$

<sup>9</sup>There has been tremendous progress in the development of non-linear and latent variable multilevel models. An adequate discussion of these models would require another paper, however, so we leave the reader with some references to the relevant literature. Goldstein (1991, 1995), Hedeker and Gibbons (1994), and Gibbons and Hedeker (1997) develop multilevel models for categorical data. A discussion of multilevel event count and event history models can be found in Goldstein (1995). The idea of multilevel latent variable models was first probed in papers by Jöreskog (1971) and Sörbom (1974). Recently, Muthén (1989, 1991, 1994), McDonald (1994), and Chou, Bentler, and Pentz (2000) have expanded the statistical theory for such models, making it possible to conduct multilevel factor analyses (see Li et al. 1998, Muthén 1991) and multilevel simultaneous equation analyses.

<sup>10</sup>Our development and notation of the multilevel model draw liberally from the excellent discussion by Bryk and Raudenbush (1992).

and

$$\beta_{ij} = \gamma_{10} + \gamma_{11}z_j + \delta_{ij}. \quad (3)$$

Taken together, (2) and (3) comprise the level-2 model. Here the  $\gamma$ -parameters denote the *fixed* level-2 parameters and  $z_j$  denotes a level-2 predictor. The  $\delta$ -parameters in this model are disturbances. Thus, in (2) and (3), no assumption is made that the level-2 predictor accounts perfectly for the variation in the level-1 parameters.

The multilevel model is fully characterized by the level-1 model, (1), and the level-2 model shown in (2) and (3). A single-equation expression of the model is derived by substituting (2) and (3) into (1):

$$\begin{aligned} y_{ij} &= (\gamma_{00} + \gamma_{01}z_j + \delta_{0j}) + (\gamma_{10} + \gamma_{11}z_j + \delta_{1j})x_{ij} + \varepsilon_{ij} \\ &= \gamma_{00} + \gamma_{01}z_j + \gamma_{10}x_{ij} + \gamma_{11}z_jx_{ij} + \delta_{0j} + \delta_{1j}x_{ij} + \varepsilon_{ij}, \end{aligned} \quad (4)$$

where  $\gamma_{00}$  denotes the intercept or constant,  $\gamma_{01}$  is the effect of the level-2 predictor,  $\gamma_{11}$  is the effect of the level-1 predictor, and  $\gamma_{10}$  is the effect of the cross-level interaction between the level-1 and level-2 predictors. The disturbance term consists of  $\delta_{0j}$ ,  $\delta_{1j}$ , and  $\varepsilon_{ij}$ , which are the random parameters of the model. Specifically,  $\delta_{0j}$  gives the residual level-2 variation in the level-1 intercept that remains after controlling for  $z_{ji}$ ,  $\delta_{1j}$  gives the residual level-2 variation in the level-1 slope for  $x_{ij}$  after controlling for  $z_{ji}$ , and  $\varepsilon_{ij}$  is the level-1 disturbance capturing omitted level-1 predictors, measurement error in  $y_{ij}$ , and any idiosyncratic sources of variation in  $y_{ij}$  attributable to level-1 units. We can thus think of  $\delta_{0j}$  and  $\delta_{1j}$  as comprising parameter noise and of  $\varepsilon_{ij}$  as level-1 noise. Therefore, prediction error in the general multilevel model has two sources: imperfect level-1 modeling of the response variable ( $\varepsilon_{ij}$ ); and imperfect level-2 modeling of the level-1 parameters ( $\delta_{0j}$  and  $\delta_{1j}$ ).

The specification of the multilevel model in (4) is incomplete without specifying the assumptions concerning the disturbances. The following assumptions are common in multilevel analysis.

(1)  $E[\delta_{0j}] = E[\delta_{1j}] = E[\varepsilon_{ij}] = 0$ . This implies that there is no systematic parameter noise or level-1 noise.

(2)  $\text{Var}[\delta_{0j}] = \tau_{00}$ ,  $\text{Var}[\delta_{1j}] = \tau_{11}$ ,  $\text{Var}[\varepsilon_{ij}] = \sigma^2$ . This stipulates that the level-1 and level-2 disturbances have a constant variance.<sup>11</sup> An important objective of multilevel analysis is to estimate these variance components and to make inferences about them.

<sup>11</sup>This assumption may be relaxed for the level-1 disturbances (see Browne et al. 2000; Snijders and Bosker 1999). We are also aware of an application in which the level-2 units were characterized by different (co)variance structures (Thum 1997).

(3)  $\text{Cov}[\delta_{0j}, \delta_{1j}] = \tau_{01}$ . This means that the level-2 disturbances on the intercepts and slopes may be correlated. It is common to find that level-2 units with large slopes also have large intercepts or, conversely, have small intercepts. The covariance term  $\tau_{01}$  captures this relationship between the intercepts and slopes and, in general, one should always estimate this term (Snijders and Bosker 1999).

(4)  $\delta_{0j}$  and  $\delta_{1j}$  are normally distributed, as is  $\varepsilon_{ij}$ .<sup>12</sup> Taken together, assumptions (1)–(4) imply that the level-2 disturbances are drawn from a bivariate normal distribution with mean 0 and a variance-covariance matrix

$$\Sigma = \begin{pmatrix} \tau_{00}\tau_{01} \\ \tau_{01}\tau_{11} \end{pmatrix}. \quad (5)$$

The level-1 disturbances are drawn from a normal distribution with mean 0 and variance  $\sigma^2$ .

(5)  $\text{Cov}[\delta_{0j}, \varepsilon_{ij}] = \text{Cov}[\delta_{1j}, \varepsilon_{ij}] = 0$ . This implies that the errors in the location of slopes and intercepts are uncorrelated with errors that affect where level-1 units are located on the dependent variable. This assumption is usually necessary to obtain an identified model. It implies that the effects of omitted level-1 predictors is assumed to be fixed, which, in the absence of clear contradictory information, is a reasonable assumption.

With the major assumptions specified, we now consider the statistical features of the multilevel model. To begin, we denote the multilevel disturbance term as  $u_{ij} = \delta_{0j} + \delta_{1j}x_{ij} + \varepsilon_{ij}$ . This expression has several important characteristics. First, it can be shown that  $u_{ij}$  does not have constant variance:<sup>13</sup>

$$\begin{aligned} \text{Var}[u_{ij}] &= E[(\delta_{0j} + \delta_{1j}x_{ij} + \varepsilon_{ij})^2] \\ &= E[\delta_{0j}^2] + 2x_{ij}E[\delta_{0j}, \delta_{1j}] + x_{ij}^2E[\delta_{1j}^2] + E[\varepsilon_{ij}^2]. \end{aligned} \quad (6)$$

$$= \tau_{00} + 2x_{ij}\tau_{01} + x_{ij}^2\tau_{11} + \sigma^2$$

In (6), it is clear that  $\text{Var}[u_{ij}]$  (and therefore  $\text{Var}[y_{ij}]$ ) is in part a function of the level-1 predictor; hence  $u_{ij}$  has nonconstant variance (even though  $\delta_{ij}$  and  $\varepsilon_{ij}$  have constant variances by assumption (2)). Constant variance will only be achieved when  $\delta_{1j}$  is 0, which means that  $z_j$

<sup>12</sup>This last assumption is appropriate for the linear multilevel models on which this paper focuses. Of course, models for categorical, count, or duration data require a different specification of the distribution of the level-1 disturbances.

<sup>13</sup>For this proof, we make an additional assumption that  $\text{Cov}[\varepsilon_{ij}, \varepsilon_{kl}] = 0$  for  $i \neq j, k \neq l$ . This assumption is included merely for convenience. It can and should be relaxed in applications of the multilevel model to time series of pooled cross-sections (Goldstein 1995).

perfectly accounts for the differences in the slope of  $x_{ij}$  across level-2 units.

Second, it can be shown that the multilevel disturbances are also correlated for level-1 units nested in the same level-2 unit. Let  $u_{ij}$  and  $u_{kj}$  denote two such disturbances. Then

$$\begin{aligned}\text{Cov}[u_{ij}, u_{kj}] &= E[(\delta_{0j} + \delta_{1j}x_{ij} + \varepsilon_{ij})(\delta_{0j} + \delta_{1j}x_{kj} + \varepsilon_{kj})] \\ &= E[\delta_{0j}^2] + x_{ij}E[\delta_{0j}, \delta_{1j}] + x_{kj}E[\delta_{0j}, \delta_{1j}] \\ &\quad + x_{ij}x_{kj}E[\delta_{1j}^2] \\ &= \tau_{00} + x_{ij}\tau_{01} + x_{kj}\tau_{01} + x_{ij}x_{kj}\tau_{11}\end{aligned}\quad (7)$$

This covariance will go to 0 only if  $\delta_{0j} = \delta_{1j} = 0$ , which means that  $z_j$  perfectly accounts for the variation across level-2 units in the level-1 model intercepts and slopes. The covariance in (7) produces the so-called intra-class or cluster correlation that we discussed earlier. Specifically, the intra-class correlation is equal to

$$\frac{\text{Cov}[u_{ij}, u_{kj}]}{\sqrt{\text{Var}[u_{ij}]\text{Var}[u_{kj}]}}. \quad (8)$$

This correlation is an indicator of the amount of duplication among level-1 units that are nested in the same level-2 unit (Kreft and De Leeuw 1998). Alternatively, it can be interpreted as an indicator of group homogeneity: the more homogeneous level-2 units are, the higher the intra-class correlation is.

These results imply that multilevel models are tailored to situations in which the data are clustered and the variance is not constant. In contrast, standard regression analysis of such data will generally be inappropriate because OLS assumes there is constant variance and no clustering. As we have discussed, clustering alone can lead to false inferences about the impact of predictors. Further, it is useful to note that clustering and nonconstant variance are often more than statistical nuisances. In many applications, these data features are of substantive interest. For example, what does dependence among the observations imply substantively about the problem under study? Moreover, what factors help account for this dependence and for nonconstant variance? These may be important questions that prompt the researcher to explore the nature of variation and co-variation in the data. This in turn can help in the development of theoretical explanations that span multiple levels of analysis. It is when we look at clustering and nonconstant variance as substantively interesting phenomena that multilevel models have substantial value-added. We now elaborate on the multilevel model presented above.

## A General Two-Level Linear Multilevel Model and Submodels

We can generalize the models in (1), (2) and (3) by including multiple level-1 and level-2 predictors. To set notation, let there be  $P$  level-1 predictors,  $x_{pij}$  ( $p = 1, \dots, P$ ). Then the general level-1 model is given by:

$$y_{ij} = \beta_{0j} + \sum_{p=1}^P \beta_{pj}x_{pij} + \varepsilon_{ij}, \quad (9)$$

where  $\varepsilon_{ij} \sim N(0, \sigma^2)$  (cf. assumption (4)). Further, assume that there are  $Q$  level-2 predictors,  $z_{qj}$  ( $q = 1, \dots, Q$ ). Then the general level-2 model for the intercepts is given by:

$$\beta_{0j} = \gamma_{00} + \sum_{q=1}^Q \gamma_{0q}z_{qj} + \delta_{0j} \quad (10)$$

and the general level-2 model for the level-1 slopes is given by:

$$\beta_{pj} = \gamma_{p0} + \sum_{q=1}^Q \gamma_{pq}z_{qj} + \delta_{pj} \quad (11)$$

Here we assume (cf. assumption (4)) that the  $\delta$ -parameters follow a multivariate normal distribution with mean 0 and a covariance matrix consisting of elements  $\tau_{pr}$  (for  $p, r = 0, \dots, P$ ). Note that the parameters  $\gamma_{0q}$  and  $\gamma_{pq}$  can be restricted, so that not every equation for level-1 parameters needs to contain the same level-2 predictors. Substitution of (10) and (11) into (9) gives the single-equation general multilevel model:

$$\begin{aligned}y_{ij} &= \gamma_{00} + \sum_{q=1}^Q \gamma_{0q}z_{qj} + \sum_{p=1}^P \gamma_{p0}x_{pij} \\ &\quad + \sum_{q=1}^Q \sum_{p=1}^P \gamma_{pq}z_{qj}x_{pij} + \delta_{0j} + \sum_{p=1}^P \delta_{pj}x_{pij} + \varepsilon_{ij}.\end{aligned}\quad (12)$$

The meaning of the various components in this equation is similar to that shown in (4). Specifically, the first term gives the constant, the second term gives the effects of the level-2 predictors, the third term gives the effects of the level-1 predictors, the fourth term gives the cross-level interactions, and the remaining terms constitute the disturbance term.

The model shown in (12) is very general and contains a wide variety of submodels that are well known in political science. Table 1 lists these models with the restrictions that are needed to derive them from (12) (see Bryk and Raudenbush 1992). We consider these models in turn.

**TABLE 1** Submodels of the General Multilevel Model

Model	Parameter Noise	Level-1 Predictors	Level-2 Predictors
General Multilevel Model	Yes	Yes	Yes
Random Coefficient Model	Yes	Yes	No
Means-as-Outcomes Model	Yes	No	Yes
Random Effects ANOVA	Yes	No	No
Interactive Model	No	Yes	Yes

**The random coefficients model.** In the random coefficients model (see Swamy 1970; Swamy and Tavlas 1995), which is widely used in pooled cross-sections and time-series analysis (see Dielman 1989; Sayrs 1989; Stimson 1985) and also in other areas of political analysis (e.g., Jackson 1992), the level-2 predictors are dropped from (10) and (11). Thus, the level-2 model for the random coefficients model is

$$\beta_{0j} = \gamma_{00} + \delta_{0j} \quad (13)$$

and

$$\beta_{pj} = \gamma_{p0} + \delta_{pj}. \quad (14)$$

The single-equation multilevel model is then given by:

$$y_{ij} = \gamma_{00} + \sum_{p=1}^P \gamma_{p0} x_{pj} + \delta_{0j} + \sum_{p=1}^P \delta_{pj} x_{pj} + \varepsilon_{ij}. \quad (15)$$

This model is simpler than (12) in that there are no fixed effects from the level-2 predictors and no cross-level interactions. The disturbance term of this model is given by  $u_{ij} = \delta_{0j} + \sum_p \delta_{pj} x_{pj} + \varepsilon_{ij}$  and has a variance of

$$\text{Var}[u_{ij}] = \tau_{00} = 2 \sum_p \tau_{0p} x_{pj} + \sum_p \sum_r \tau_{pr} x_{pj} x_{rj} + \sigma^2, \quad (16)$$

(where  $x_{pj}$  and  $x_{rj}$  denote two level-1 predictors), and a covariance of

$$\text{Cov}[u_{ij}, u_{kj}] = \tau_{00} = \sum_p \tau_{0p} x_{pj} + \sum_p \tau_{0p} x_{pk} + \sum_p \sum_r \tau_{pr} x_{pj} x_{rkj}. \quad (17)$$

**The means-as-outcomes model.** Another submodel of interest from (12) is the so-called “means-as-outcomes” model. In this model, no level-1 predictors are included, so that the level-1 model is given by

$$y_{ij} = \beta_{0j} + \varepsilon_{ij}. \quad (18)$$

The level-2 model for this case is equivalent to (10). Thus, substitution of (10) into (18) gives rise to the following combined multilevel model:

$$y_{ij} = \gamma_{00} + \sum_{q=1}^Q \gamma_{0q} z_{qj} + \delta_{0j} + \varepsilon_{ij}. \quad (19)$$

Since  $E[y_{ij}] = \mu_j = \gamma_{00} + \sum_q \gamma_{0q} z_{qj}$ , the model treats the level-2 means of the dependent variable,  $\mu_j$ , as a function of the level-2 predictors (whence the name means-as-outcomes model). The disturbances of this model,  $u_{ij} = \delta_{0j} + \varepsilon_{ij}$ , possess compound symmetry (i.e., have an exchangeable covariance structure); their variances are given by  $\text{Var}[u_{ij}] = \tau_{00} + \sigma^2$  and their covariances by  $\text{Cov}[u_{ij}, u_{kj}] = \tau_{00}$ , resulting in an intra-class correlation of  $\tau_{00} / (\tau_{00} + \sigma^2)$ .

**The random effects ANOVA model.** Random effects ANOVAs play an important role in the analysis of within-subjects experimental designs and repeated measures (e.g., Winer 1971). This model can be derived by specifying the level-1 model as (18). Here  $\mu_j$  are the treatment means and  $\varepsilon_{ij}$  is within-group variance. It is assumed that the level-2 units are the products of sampling, so that

$$\beta_{0j} = \mu_j = \gamma_{00} + \delta_{0j}. \quad (20)$$

Substitution into (18) gives

$$y_{ij} = \gamma_{00} + \delta_{0j} + \varepsilon_{ij}. \quad (21)$$

The disturbances,  $u_{ij} = \delta_{0j} + \varepsilon_{ij}$  possess compound symmetry. The random effects ANOVA is often a good place to start in analyzing multilevel data structures. Because this model decomposes the variance in the dependent variable across different levels of analysis, the researcher is in a good position to assess the importance of each level and how much would be lost by ignoring a particular level.

**Interactive model.** Finally, it is instructive to note that if the disturbances are removed from (10) and (11) (or equivalently, if the estimates of their variances and covariances are 0), the interactive model is obtained. It is obvious that in the absence of (residual) level-2 heterogeneity,

ity, the disturbance term is simply  $u_{ij} = \varepsilon_{ij}$ . Given our assumptions, this implies that there is no clustering and that the variance is constant. Thus, the interactive model is, in fact, a special case of the general multilevel model in (12), which arises under stringent and usually unrealistic assumptions about the level-2 disturbances.

### Models with More Than Two Levels

The multilevel model can be extended to data structures that have more than two levels. From a conceptual perspective such an extension is straightforward: effects at the highest level are considered fixed, while effects at lower levels may vary over units at higher levels in the hierarchical data structure (although they do not have to). However, in terms of interpretation, these models can be very complex because of the rapid increase in possible interaction terms, including higher-level interactions. For instance, a three-level model may contain three sets of two-way cross-level interactions (between level-1 and level-2, level-1 and level-3, and level-2 and level-3), in addition to a set of three-way interactions (between level-1, level-2, and level-3) and three sets of main effects (one for each level). Not only can the interpretation of all of these terms be complicated, but the demands on the data may be quite steep if one is to achieve sufficient statistical power to unravel all of the interactions. Of course, not every model will contain all possible interactions because there may not be theoretical reasons to do this. An example of a relatively simple three-level model is presented in our application.

### Statistical Inference for the Multilevel Model

Our discussion so far has focused on the conceptual development of multilevel models and the purposes they serve. An unanswered question is how these models are estimated. We now turn to this question, drawing a distinction between estimation of the fixed effects and variance components, on the one hand, and of the level-1 coefficients, on the other.

**Fixed effects and variance components.** The dominant approach to estimating the fixed effects and variance components is maximum likelihood estimation (MLE).<sup>14</sup> MLE requires the specification of a density for the level-1 and level-2 disturbances. The most common choice of

<sup>14</sup>However, there is a wide range of alternative approaches. Among these alternatives are two-step OLS (see Chou, Bentler, and Pentz 2000; Kackar and Harville 1981; Van Den Eeden 1988), generalized least squares (De Leeuw and Kreft 1986), Bayesian estimation (Lindley and Smith 1972; Rubin 1981; see also Browne and Draper

density in the context of linear multilevel models is the normal density (see assumption (4)). This distributional assumption informs the likelihood function, which is maximized to obtain estimators of the fixed effects and the variance components (for detailed discussions see Bryk and Raudenbush 1992; Goldstein 1995; Longford 1993; Raudenbush 1988; Searle, Casella, and McCulloch 1992).

There are several algorithms for implementing MLE. In a pioneering paper on maximum-likelihood estimation of variance components models, Dempster, Rubin, and Tsutakawa (1981) proposed the EM algorithm (see also Bryk and Raudenbush 1992; Mason, Wong, and Entwistle 1983). Goldstein (1986) developed an iterative generalized least squares (IGLS) algorithm, while Longford (1987) proposed a Fisher scoring algorithm. Under the assumption of normality IGLS and Fisher scoring are equivalent (see Goldstein 1995). The advantage of EM is guaranteed convergence, while Fisher scoring (IGLS) has the advantage of being fast (see Kreft and De Leeuw 1998; Kreft, De Leeuw, and Van Der Leeden 1994).

MLE has certain desirable statistical properties. When the normality assumption holds, MLE is consistent and asymptotically efficient (Bryk and Raudenbush 1992; Goldstein 1995; Longford 1993; Raudenbush 1988; Searle, Casella, and McCulloch 1992). MLE remains consistent in the estimation of the fixed effects even when the normality assumption is violated (Goldstein 1995).<sup>15</sup> While the ideal circumstances for estimating multilevel models are balanced designs (i.e., equal numbers of level-1 units per level-2 unit), Raudenbush (1988) comments that MLE performs well even with unbalanced designs, which are common in political analysis. In terms of finite sample properties, a series of Monte Carlo simulation studies has demonstrated that MLE fixed effects estimators are consistent (Kreft 1996).

If there is a weakness to MLE it may be in the estimation of the variance components in small samples. MLE variance components tend to be biased downwardly in

2000; Seltzer, Wong, and Bryk 1996; Zeger and Karim 1991), and generalized estimating equations (see Goldstein 1995; Liang and Zeger 1986; Zorn 2001). Of interest are also the minimum variance (MINVAR) and minimum norm quadratic unbiased estimation (MINQUE) estimators of the variance components, which are often used in the context of random effects ANOVA (see Searle, Casella, and McCulloch 1992).

<sup>15</sup>However, the standard errors of the fixed effects will tend to be biased. In addition, the estimates of the variance components may also be severely biased. One solution to these problems is to use a Monte Carlo Markov Chain procedure (e.g., Gibbs sampling) or to rely on bootstrapping (see Goldstein 1995; Kuk 1995; Meijer, Van Der Leeden, and Busing 1995).

such cases because no correction is made for the degrees of freedom consumed by estimation of the fixed effects (Swallow and Monahan 1984). An alternative estimator of the variance components is restricted maximum-likelihood estimation (REML), which corrects this problem (see Bryk and Raudenbush 1992; Goldstein 1995; Kreft and De Leeuw 1998; Longford 1993; Searle, Casella, and McCulloch 1992).<sup>16</sup> Although REML and MLE are asymptotically equivalent (Richardson and Welsh 1994), the premise is that REML variance component estimates are less biased than MLE variance component estimates in small samples of level-2 units. Evidence from Monte Carlo simulation studies lends support to this premise. However, this evidence also points to a potential problem with REML. In the simulations, REML was sometimes less efficient than MLE. Consequently, the mean squared error (which is the sum of the squared bias and the variance of an estimator) did not always favor REML over MLE (Kreft 1996).

**Level-1 coefficients.** In multilevel modeling, attention is usually not restricted to estimating fixed effects and variance components. Random level-1 coefficients are also of considerable interest because these coefficients shed light on the behavior of the level-1 model in each of the level-2 units. The question is how one should estimate these coefficients.

One option is to estimate level-1 coefficients separately for the different level-2 units (e.g., using OLS). The advantage of this approach is that unbiased estimators of the level-1 coefficients can be obtained for each level-2 unit (assuming that the level-1 sample sizes are sufficient) to allow estimation. Unfortunately, these estimators often lack precision because they are based on limited data, namely, the data for a particular level-2 unit. Especially when the sample sizes for the level-2 units are small, the variance of the separate regression estimators can be large.

An alternative estimation approach is pooled regression, i.e., regression analysis of the level-1 units combined across all level-2 units. The pooled regression estimator of the effect of a level-1 predictor may be biased for a particular level-2 unit, in that the true effect of the predictor in this unit is different from the pooled estimate. On the other hand, there are efficiency gains in relying on pooled estimators because all of the data is used and fewer parameters are estimated (see Bartels 1996).

<sup>16</sup>More specifically, REML partitions the likelihood function into two parts, one of which contains the fixed effects. Variance component estimates are based on the part of the likelihood function that does not include the fixed effects.

Neither a separate regression approach nor a pooled approach, then, is ideal. The question is whether there is a way in which the approaches can be combined that results in a “compromise estimator” (Morris 1983, 47) with more desirable properties. This question has been addressed in the literature on empirical Bayes (EB) inference, which shows that combining estimators is indeed useful (Carlin and Louis 1996; Efron and Morris 1975; Morris 1983).

The EB estimator of a level-1 coefficient  $\beta_{pj}$  is a weighted combination of the separate regression estimator of that coefficient and  $\hat{\gamma}_{p0} + \sum_q \hat{\gamma}_{pq} z_{qj}$ . The weights correspond inversely to the precision of these two estimators. That is, the more precise an estimator is, the more important its contribution is to the EB estimator. Thus, there is “shrinkage” in the direction of the most precise estimator. The resulting EB estimator has the desirable property that its MSE is typically smaller than the MSE of the estimators on which it is based (Carlin and Louis 1996; Efron and Morris 1975; Morris 1983).

One of the biggest advantages of EB estimation is that the estimators of the level-1 parameters are no longer rooted only in the data for a particular level-2 unit. Because the EB estimator is a function of the fixed effects, which are estimated from the pooled data, the level-1 parameter estimates, in effect, borrow information from the data for other level-2 units. Borrowing strength is a major advantage of multilevel modeling (see Bryk and Raudenbush 1992; Kreft and De Leeuw 1998; Raudenbush 1988; for a political science example see Western 1998). It makes it possible to draw inferences about level-2 units even in the light of sparse data for those units. Of course, the need to borrow strength is inversely related to the sample sizes for level-2 units. And the ability to borrow strength depends on how much the level-2 units have in common.

## Discussion

Now that the logic and estimation of multilevel models have been discussed, it is natural to ask what these models have to offer for political analysis compared to alternative approaches. First, the notion of borrowing strength that is embedded in EB estimation offers a major improvement over separate regression estimation. When faced with causal heterogeneity, political scientists often resort to separate estimation of level-1 units for each level-2 unit. However, this is not an efficient estimation strategy and we expect that in many cases EB estimation is a much better choice. In this regard, a multilevel modeling approach to causal heterogeneity can be very beneficial.

A second major benefit of multilevel models is that they permit the analysis of substantive contextual effects while still allowing for heterogeneity between contextual units. This is an important improvement over the alternatives of dummy variables models and contextual models. Dummy variables models can account for level-2 heterogeneity, but they contain no substantive explanation of this heterogeneity. Contextual models provide substantive explanation, but make the unrealistic assumption that this explanation eliminates any remaining level-2 heterogeneity. Multilevel models combine substance with (more) reasonable assumptions about level-2 heterogeneity.

A third major benefit of multilevel models concerns statistical inference. The multilevel character of much of political science data is often ignored. Political scientists often treat multilevel data structures as if no hierarchy between units of analysis existed. Consequently, observations are treated as independent, whereas in fact they are to some extent dependent because of the hierarchical nesting structure. This can easily lead to incorrect inferences, such as rejecting the null hypothesis of no effect too frequently. Multilevel models improve inferences because the hierarchical data structure is explicitly taken into consideration.

To illustrate the application of a multilevel model, we now provide an empirical illustration of multilevel analysis.

## Application

We illustrate multilevel models through an analysis of public support for the European Union (EU). Studies of EU support typically rely on multilevel data. As such, the topic of EU support is an excellent test case for illustrating multilevel models. Our goals in this section are two-fold. First, we show how to develop multilevel models from substantive research questions and how to interpret the results. Second, we show the implications of ignoring the multilevel data structure.

The literature on EU support discusses at least three different levels of analysis. First, a large number of studies concentrate on the *individual* level; these studies assess the nature and determinants of inter-individual differences in attitudes toward the EU (e.g., Inglehart, Rabier, and Reif 1991). Second, a sizable literature documents the nature of *cross-national* differences in EU support (e.g., Inglehart, Rabier, and Reif 1991). More recently, scholars have become interested in national *political parties* as a third, intermediate level of analysis.

The argument here is that political parties are an important context that shapes the opinions of party supporters. As a result supporters from the same party tend to share similar views about the EU (e.g., Franklin, Marsh, and McLaren 1994), suggesting clustering.<sup>17</sup>

Considering these three levels of analysis together, what emerges is a multilevel data structure: individuals (to the extent that they are party supporters) are nested within political parties, which in turn are nested within EU member states. We can model this data structure through a three-level multilevel model. The dependent variable in this model can be written as  $Support_{ijk}$ , which denotes the level of support for the EU for an individual  $i$  who supports party  $j$  in country  $k$ .

## Theory

Three substantive questions arise when we think of EU support as a three-level phenomenon. First, what is the importance of each of the three levels for understanding EU support? Second, how do we account for EU support at the different levels? That is, what predictors can explain EU support? Finally, is there causal heterogeneity in the effects of predictors? These questions follow a logical progression—later questions presuppose an answer to earlier questions. Let us now outline how one would translate these questions into appropriate multilevel models.

First, does support for the EU vary across the three levels of analysis that we have identified? We can answer this question by way of an ANOVA that decomposes the variance in EU support:<sup>18</sup>

$$Support_{ijk} = \gamma_{000} + v_{00k} + \delta_{0jk} + \varepsilon_{ijk}. \quad (22)$$

<sup>17</sup>A fourth level of analysis is time: how do opinions about the EU vary over time (e.g., Eichenberg and Dalton 1993)? For the sake of simplicity, we will not consider this level in this paper, focusing instead on one particular point in time, namely 1996.

<sup>18</sup>To derive this model, consider the individual-level model

$$Support_{ijk} = \alpha_{0jk} + \varepsilon_{ijk},$$

where  $\alpha_{0jk}$  is the mean level of EU support in political party  $j$  in country  $k$ , and  $\varepsilon_{ijk}$  is individual-level variation around this mean. We model the mean by way of the party-level model

$$\alpha_{0jk} = \beta_{00k} + \delta_{0jk}.$$

Here  $\beta_{00k}$  is a national mean for EU support and  $\delta_{0jk}$  is party-level variation around this mean. Finally,

$$\beta_{00k} = \gamma_{000} + v_{00k},$$

where  $\gamma_{000}$  is the overall mean of EU support and  $v_{00k}$  is cross-national variation around this mean. Back-substitution of this formula yields

$$\alpha_{0jk} = \gamma_{000} + v_{00k} + \delta_{0jk}.$$

Further back-substitution yields (22).

In this model,  $\gamma_{000}$  is the grand mean of EU support (i.e., the mean across individuals, parties, and countries). The sources of cross-national variation, which cause particular EU member states to deviate from the grand mean, are contained in  $v_{00k}$ . Similarly,  $\delta_{0jk}$  contains sources of cross-party variation. Finally,  $\varepsilon_{ijk}$  captures inter-individual differences. The variances of these different sources of variation are given by  $\omega_{00}$ ,  $\tau_{00}$  and  $\sigma^2$ , respectively. In order to argue that all three levels of analysis are important for EU support, we should find that all three of these variance components are statistically significant.

If EU support varies significantly at all three levels, the next question to ask is how we should account for this variance? The EU literature suggests several important covariates that help predict EU support. At the individual level, scholars have discussed three major predictors of EU support (see Gabel 1998). First, an extensive literature describes the impact of economic considerations on EU support. A key finding in this literature is that bad economic conditions for an individual tend to reduce support for the EU, while good economic conditions tend to increase such support (e.g., Gabel 1998). In this article, we use a person's income as an indicator of economic circumstances. Specifically, we distinguish between people in the bottom income quartile, the top income quartile, and the two middle income quartiles (our baseline category) in the income distribution of an EU member state. We expect people in the bottom income quartile to be the least supportive of the EU.

Second, political orientations appear to be associated with EU support. There is considerable debate about the relative importance of these orientations, but it appears that citizens with a leftist ideological orientation are more opposed to the EU than those with a rightist orientation (Inglehart, Rabier, and Reif 1991). Thus, we include ideology as an individual-level predictor.

Third, Inglehart (1970) has argued that EU support is higher among opinion leaders than among the general public. The reason is that opinion leaders have a better understanding of the EU, which makes them feel less threatened by it. The impact of opinion leadership is assumed to be *message independent*: opinion leadership enhances EU support regardless of the information environment (Inglehart, Rabier, and Reif 1991). We put this assumption to test later in this application.

Moving to the party level, recently a literature has emerged about partisan cue-taking effects. Such effects arise when parties take a position on an issue such as European integration, which is then used by party supporters to inform their own opinions about the issue. As a result, the supporters assimilate their opinions to those of the party. There is now good empirical evidence that this

indeed has happened in the case of public opinion toward the EU (Franklin, Marsh, and McLaren 1994; Franklin, Marsh, and Wlezien 1994; Franklin, Van Der Eijk, and Marsh 1995; cf. Siune and Svenson 1993). As suggested before, cue-taking effects induce clustering at the party-level because the supporters of a particular party will tend to assimilate to the same cue.

Finally, consider the national level of analysis. Here we focus on two predictors. First, we consider the date at which a country acceded to the EU. Accession to the EU has been largely an elite-driven process. It is not uncommon to find that public opinion toward the EU is very negative in countries that have recently acceded to the union. However, often public opinion toward the EU becomes more favorable as time passes. Hence, we expect EU support to be greater in older EU member states than in more recent member states.

Second, individuals may look at the dependence of the national economy on the EU in determining whether to support or oppose the union. As the national economy becomes more dependent on EU membership, EU support may increase because people realize how important the EU is for their country. One indicator of this dependence is the percentage of the trade-flow (imports and exports) that is intra-EU (e.g., Eichenberg and Dalton 1993). We expect that higher percentages of intra-EU trade are associated with greater EU support.

Having defined the relevant theoretical factors predicting EU support, we now consider how to bring these different predictors together in a multilevel model. We start by defining an individual-level model:

$$\begin{aligned} Support_{ijk} = & \alpha_{0jk} + \alpha_{1jk} LI_{ijk} + \alpha_{2jk} HI_{ijk} \\ & + \alpha_{3jk} ID_{ijk} + \alpha_{4jk} OL_{ijk} \\ & + \alpha_{5jk} Male_{ijk} + \alpha_{6jk} Age_{ijk} + \varepsilon_{ijk}. \end{aligned} \quad (23)$$

Here  $LI_{ijk}$  is a dummy variable for the lowest income quartile,  $HI_{ijk}$  is a dummy variable for the highest income quartile,  $ID_{ijk}$  is a person's ideology, and  $OL_{ijk}$  is opinion leadership. We include two demographic control variables in the model:  $Male_{ijk}$  is a dummy variable for gender (1 is male) and  $Age_{ijk}$  is an individual's age.

By modeling the individual-level constant,  $\alpha_{0jk}$ , we can introduce the party-level predictor that we identified:

$$\alpha_{0jk} = \beta_{00k} + \beta_{01k} Cue_{jk} + \delta_{0jk}, \quad (24)$$

where  $Cue_{jk}$  stands for party cue. Further, by modeling the party-level constant,  $\beta_{00k}$ , we can introduce the country-level predictors:

$$\beta_{00k} = \gamma_{000} + \gamma_{001} Tenure_k + \gamma_{002} Trade_k + v_{00k}, \quad (25)$$

where  $Tenure_k$  denotes the length of a country's EU membership and  $Trade_k$  denotes the percentage of the country's trade that is intra-EU.

Substitution of (25) into (24) gives

$$\alpha_{0jk} = \gamma_{000} + \gamma_{001}Tenure_k + \gamma_{002}Trade_k \\ + \beta_{01k}Cue_{jk} + v_{00k} + \delta_{0jk}.$$

If we make the assumption that the effect of party cues is fixed (i.e.,  $\beta_{01k} = \gamma_{010}$ ) and that the effect of the individual-level predictors is fixed as well (i.e.,  $\alpha_{pjk} = \gamma_{p00}$  for  $p \neq 0$ ), then substitution of this result into (23) yields:

$$Support_{ijk} = \gamma_{000} + \gamma_{001}Tenure_k + \gamma_{002}Trade_k + \gamma_{010}Cue_{jk} \\ + \gamma_{100}LI_{ijk} + \gamma_{200}HI_{ijk} + \gamma_{300}ID_{ijk} + \gamma_{400}OL_{ijk} \\ + \gamma_{500}Male_{ijk} + \gamma_{600}Age_{ijk} + v_{00k} + \delta_{0jk} + \varepsilon_{ijk}. \quad (26)$$

This model has several notable features. First, it is comprehensive in that it brings together the predictors at different levels. Second, (26) makes no assumption that these predictors fully account for the variation in EU support at the different levels. Thus, the model implies variance components  $\sigma^2$  for  $\varepsilon_{ijk}$ ,  $\tau_{00}$  for  $\delta_{0jk}$ , and  $\omega_{00}$  for  $v_{00k}$ . These features make the model appropriate for answering the second question that the EU data raise, namely, how do we account for EU support at different levels of analysis?

An important limitation of (26) is the assumption that the individual-level predictors have fixed effects. For one of the predictors, namely opinion leadership, there is an a priori reason to question this assumption.<sup>19</sup> Although Inglehart has argued that the effect of opinion leadership on EU support is uniformly positive (Inglehart, Rabier, and Reif 1991), there is theoretical reason to expect the effect of this predictor varies as a consequence of party cues.<sup>20</sup> In other words, there is causal heterogeneity in the effect of opinion leadership.

The theoretical rationale for this hypothesis comes from Zaller's (1992) theory of mass opinion. This theory claims that individuals are persuaded by a message to the extent that they receive the message and accept it. Na-

<sup>19</sup>We also tested the assumption for the other individual level predictors and for party cues. For these predictors we had no theoretical expectation that their effects vary. The tests do not suggest such variation either, so that we retain the assumption of fixed effects for these predictors.

<sup>20</sup>Inglehart (1970) alluded to the possibility that the effect of opinion leadership is related to the content of elite messages, but did not pursue it.

tional political parties are an important source of messages—or cues—about the EU. Opinion leaders are likely to receive these cues because they are tuned into the communication flow about the EU (which is usually not true of the general public). Moreover, opinion leaders are also likely to accept these cues if they come from the party they support. After all, this party is a trusted source in the eyes of the opinion leaders. Thus, we expect that opinion leaders are likely to be persuaded by the party cues.

But if this is true, opinion leaders' support for the EU depends on the nature of the party cues to which they are exposed. If the cue that the party sends is pro-EU, then opinion leaders should also be pro-EU (and more so than the general public). On the other hand, if the party cue is anti-EU, then opinion leaders should oppose the EU. Thus, instead of predicting a uniformly positive effect of opinion leadership on EU support, we hypothesize an effect that varies with party cues.

How do we model this effect? We do this by dropping the assumption in (23), that the effect of opinion leadership is fixed. Instead, we model this effect as follows:

$$\alpha_{4jk} = \beta_{40k} + \beta_{41k}Cue_{jk} + \delta_{4jk}. \quad (27)$$

This model stipulates that the effect of opinion leadership, which is given by  $\alpha_{4jk}$  in (23), varies as a function of party cues. We do not expect these cues to explain all of the cross-party variation in  $\alpha_{4jk}$ . Our theory does not suggest that the right-hand side coefficients in (27) vary across countries, so that we assume these coefficients to be fixed (i.e.,  $\beta_{40k} = \gamma_{400}$  and  $\beta_{41k} = \gamma_{410}$ ). If we retain our earlier assumption that the remaining individual-level and party-level predictors have fixed effects, we obtain:

$$Support_{ijk} = \gamma_{000} + \gamma_{001}Tenure_k + \gamma_{002}Trade_k + \gamma_{010}Cue_{jk} \\ + \gamma_{100}LI_{ijk} + \gamma_{200}HI_{ijk} + \gamma_{300}ID_{ijk} + \gamma_{400}OL_{ijk} \\ + \gamma_{500}Male_{ijk} + \gamma_{600}Age_{ijk} + \gamma_{410}Cue_{jk} \\ \times OL_{ijk}v_{00k} + \delta_{0jk} + \delta_{4jk}OL_{ijk} + \varepsilon_{ijk}. \quad (28)$$

This model contains the same variance components as (26). In addition, it contains a variance component  $\tau_{44}$  (for  $\delta_{4jk}$ ) and a covariance component  $\tau_{04}$  (which captures the relationship between  $\delta_{0jk}$  and  $\delta_{4jk}$ ). In terms of the fixed effects, the model contains a cross-level interaction,  $Cue_{jk} \times OL_{ijk}$ . We expect this interaction to have a positive effect. Because of the presence of this cross-level interaction, (28) allows us to answer the third question that we posed, namely whether the effect of the predictors is uniform or heterogeneous.

## Data

Data for the dependent variable and for the individual level predictors come from Eurobarometer survey 46.0, which was fielded in the fifteen EU member states in October and November of 1996.<sup>21</sup> We consider only the subset of respondents who declared support for one of the parties in our set (see below) and who were of voting age (at least 18 years old) at the time of the survey. As a party supporter we count anyone who indicated in the survey that they would be a likely voter for that party in the next general elections (the respondents could give only one choice). A total of 6354 respondents meet these requirements. Our measures of the dependent variable and the predictors are described in Table 2.<sup>22</sup>

At the party level, we consider only those parties for which we have data about their EU position. These data were collected via an expert survey (Ray 1999). The survey instrument consists of a seven-point rating scale of the position of a party's leadership vis-a-vis the EU in 1996 (the scale runs from "strongly opposed to European integration" to "strongly in favor of European integration"). The survey was sent to party experts for each of the EU member states, who provided ratings for the most important parties in that country. We use the average expert rating for a party as a measure of the EU cue that the party was sending at that time.<sup>23</sup> We have information on party cues for 100 different parties.<sup>24</sup>

The data for intra-EU trade are for 1995, the year prior to the Eurobarometer survey. We introduce this one-year lag because we believe that individuals base

<sup>21</sup> Mellich, Anna. 1996. *Eurobarometer 46.0: Personal Health, Energy, Development Aid, and the Common European Currency* (computer file), 2nd release, Brussels: INRA Europe (producer), 1997; Köln: Zentral Archiv für Empirische Sozialforschung (distributor), 1999; Ann Arbor, MI: Inter-University Consortium for Political and Social Research (distributor), 1999.

<sup>22</sup>We have centered several of the predictors, including those that form the cross-level interaction in (28). We centered about the grand mean of these predictors, i.e., their mean across individuals, parties, and countries. On the topic of centering see Kreft, De Leeuw, and Aiken (1995) and Hofmann and Gavin (1998).

<sup>23</sup>Here we make two assumptions. First, the average expert rating is a reasonable approximation of a party's stance vis-a-vis the EU. Second, a party's stance is public information so that it can serve as a cue. Both assumptions seem reasonable. First, with very few exceptions, there was a remarkable agreement among experts in positioning parties on the issue of the EU. Moreover, there also was agreement between the expert judgments and party manifesto data (Ray 1999). Second, we are not aware of any cases in which parties tried to hide their EU position from the public.

<sup>24</sup>The number of parties per EU member state ranges between 4 and 10. The number of supporters per party ranges between 1 and 383.

**TABLE 2 Variable Descriptions**

Variable	Description
Support	A composite of two items: (1) "Generally speaking, do you think that [our country's] membership of the European Union is (a bad thing, neither good nor bad, a good thing)? and (2) the desired speed of European integration (1=integration should be brought to a "standstill;" 7=integration should run "as fast as possible"), inter-item $r = .474$ ; standardized item $\alpha = .643$ . The composite ranges between 0 and 8, with higher scores indicating greater support for the EU. Source: Eurobarometer 46.0.
LI	A dummy variable indicating that a respondent falls in the bottom quartile of the income distribution of his/her country. Source: Eurobarometer 46.0.
HI	A dummy variable indicating that a respondent falls in the top quartile of the income distribution of his/her country. Source: Eurobarometer 46.0.
ID	Respondent's ideological self-placement on a 10-point scale (0=left; 9=right). Source: Eurobarometer 46.0.
OL	Respondent's level of opinion leadership on a 4-point scale. This measure captures the respondents' potential for active political involvement (for measurement details see Inglehart 1977). The original scale ranges from 1 (high opinion leadership) to 4 (low opinion leadership). We reversed this scale and centered it around the sample mean. Source: Eurobarometer 46.0.
Male	Respondent's gender (1=male). Source: Eurobarometer 46.0.
Age	Respondent's age (in years). Source: Eurobarometer 46.0.
Cue	Party cue as measured by the following item: "[What is] the overall orientation of the party leadership toward European integration?" (1 = strongly opposed to integration; 7 = strongly in favor of integration). The responses to this item came from party experts for the different EU member states. We averaged the responses of all experts who evaluated a particular party. For purposes of the analysis this average was centered around its mean. Source: Ray (1999).
Tenure	The number of years a country has been an EU member state. This variable was centered around the mean.
Trade	The ratio of a country's 1995 intra-EU trade balance (in 1000 ECUs) over the country's total trade balance (in ECUs). This variable was centered around the mean. Source: 1997 <i>International Statistical Yearbook</i> .

their evaluations of the EU on *past* trade information since this information is most accessible. These trade data come from Eurostat, the Statistical Office of the EU.

## Results

Is there significant variation in EU support at the individual, party and national levels? To answer this question we estimated the ANOVA model described in equation (22). Table 3 shows the ML estimates of the grand mean and the variance components.<sup>25</sup> All of the variance components are statistically significant, suggesting that there is significant variance in EU support at all three levels of analysis.<sup>26</sup> This is evidence that the multilevel character of the EU support data should not be ignored.

In order to obtain a better sense of the importance of the various levels of analysis we consider the ratio of each variance component to the total variance in EU support (see Bryk and Raudenbush 1992; Snijders and Bosker 1999), which is equal to  $(\omega_{00} + \tau_{00} + \sigma^2)$ . For example, the ratio of  $\sigma^2$  over the total variance indicates the importance of the individual level of analysis. With data measured at the individual level, it should not come as a surprise that the variance component at this level accounts for a major portion of the variance in EU support, namely about 79 percent (i.e., 100 times  $3.844 / (.745 + .283 + 3.844)$ ). Importantly, however, 21 percent of the variance in EU support is due to higher levels of analysis. Specifically, the party level accounts for about 5.8 percent of the variance in EU support. The contribution of the national level is even greater, at 15.3 percent. Clearly, to ignore these sources of variance is to miss out on important aspects of support for the EU. This could result in erroneous substantive conclusions about EU support (Opdenakker and Van Damme 2000).

Ignoring the multilevel character of the data has another adverse consequence as well. It leads the researcher to neglect the clustering of observations at the party and national levels. As we have seen, this could produce erro-

<sup>25</sup>These estimates, as all other estimates in this paper, were obtained using MLwiN V1.1 (Rasbash et al. 2000).

<sup>26</sup>There are several ways in which the statistical significance of variance components can be established. Here we rely on a likelihood-ratio test in which we compare the fit of two models (see Bryk and Raudenbush 1992; Snijders and Bosker 1999). The first model does not restrict a particular variance component and produces a likelihood of  $L_1$ . The second model restricts the variance component to zero, producing a likelihood of  $L_0$ . Now  $2[\ln(L_1) - \ln(L_0)]$  is asymptotically distributed as a  $\chi^2$ -variate with one degree of freedom. Thus, it is possible to compute a  $p$ -value for the test of the restriction. This test approach can be extended to tests of multiple parameters, including fixed effects.

TABLE 3 ANOVA

Parameter	Estimate
Fixed Effects	
Constant	4.990** (.233)
Variance Components	
Country-Level ( $\omega_{00}$ )	0.745** (.297)
Party-Level ( $\tau_{00}$ )	0.283** (.062)
Individual-Level ( $\sigma^2$ )	3.844** (.069)
-2 x Log Likelihood	26765.530

Note: Table entries are maximum likelihood (IGLS) estimates with estimated standard errors in parentheses.

\*\*=p <.01.

neous statistical inferences. The estimates of the intra-cluster correlations at the party and national levels are equal to the proportion of the total variance that the variance components at these levels account for (Bryk and Raudenbush 1992; Snijders and Bosker 1999). Thus, at the party level we observe an intra-cluster correlation of .058, while the intra-cluster correlation at the national level is .153. These are sizable positive correlations that suggest a fair amount of clustering—or overlap between observations—in particular at the national level. One can ignore this aspect of the data only at the peril of drawing incorrect statistical inferences.

The results of the ANOVA model indicate very clearly that there is significant variation in EU support at all three levels of analysis. Now we turn to the question of whether the model specified in equation (26) can account for this variance. The second column in Table 4 gives the ML estimates of the fixed effects and the variance components of this multilevel model. This model is a significant improvement over the ANOVA model:  $\chi^2 = 175.56$ ,  $df = 9$ ,  $p < .01$ , suggesting that at least a subset of the predictors has effects that are different from 0.

At the individual level we observe a powerful effect of opinion leadership on EU support. This effect is in the direction that Inglehart (1970; Inglehart, Rabier and Reif 1991; see also Janssen 1991) predicted: opinion leaders are more supportive of the EU than the general public. Despite this strong effect of opinion leadership, the overall performance of the individual-level predictors is disappointing. The effect of income is weak, while ideology does not have a statistically significant effect. Of the control variables, age has the strongest effect; the effect of gender is only marginally significant. As a set,

**TABLE 4 Determinants of EU Support**

Parameter	Multilevel Estimate	Regression Estimate
<i>Fixed Effects</i>		
Constant	5.504** (.220)	5.016** (.124)
Tenure	0.014 (.014)	0.011** (.002)
Trade	0.032 (.025)	0.039** (.003)
Party Cue	0.233** (.028)	0.275** (.018)
Lowest Income Quartile	-.106+ (.064)	-.181** (.068)
Highest Income Quartile	0.048 (.059)	-.001 (.062)
Ideology	0.019 (.015)	0.023+ (.013)
Opinion Leadership	0.152** (.028)	0.166** (.030)
Male	0.088+ (.050)	0.093+ (.053)
Age	-.013** (.002)	-.014** (.002)
<i>Variance Components</i>		
Country-Level ( $\omega_{00}$ )	0.551** (.213)	
Party-Level ( $\tau_{00}$ )	0.100** (.029)	
Individual-Level ( $\sigma^2$ )	3.771** (.067)	
Regression Disturbance		4.335** (.003)
-2 × Log Likelihood	26589.970	
F		71.502

Note: Multilevel table entries are maximum likelihood (IGLS) estimates with estimated standard errors in parentheses. Regression table entries are ordinary least squares estimates with estimated standard errors in parentheses.

+ =  $p < .10$ , \*\* =  $p < .01$ .

the individual-level predictors explain very little of the individual level variance in EU support. We can assess this explanatory power by comparing the individual-level variance component in Table (4) to the same component in Table (3) (see Bryk and Raudenbush 1992). The difference in these variance components is  $3.844 - 3.771 = .073$ . Relative to the size of the ANOVA variance component, this is a reduction of  $.073/3.844 = .019$ . Thus, the individual-level variance components explain less than 2 percent of the individual-level variance in EU support.

The story is quite different for our party-level predictor. Party cues exert a powerful effect on EU support. These cues by themselves account for about 65 percent of

the cross-party variance in EU support in the data. We determine this level of explained variance by comparing the party-level variance components of two models, one including the level-1 and level-3 predictors but not party cues, and another model that adds party cues as a predictor (see Bryk and Raudenbush 1992). The relative change in the party-level variance component is  $(.282 - .100)/.282 = .645$ . This is impressive and demonstrates the importance of party cues for understanding public opinion toward the EU.

At the national level, we find no statistically significant effects from the predictors. While the signs on the coefficients for tenure and trade are in the correct direction, neither predictor approaches statistical significance. Clearly, our explanation for cross-national variation in EU support is inadequate.

Considering the random effects, we observe that the variance components at all three levels of analysis remain significant after controlling for the predictors at these levels. This is even true at the party level, where the variance component remains sizable despite the powerful effect of party cues. Clearly, the predictors at the party and country levels do not account for all of the variance in EU support at these levels.

In sum, the results suggest that our explanation of EU support at the different levels of analysis is only partially successful. The explanation is best at the party level, where we find strong support for the position taken by Franklin and his colleagues that party cues affect support for the EU (Franklin, Marsh, and McLaren 1994; Franklin, Marsh, and Wlezien 1994; Franklin, Van Der Eijk, and Marsh 1995). The explanation is considerably less successful at the individual level, where much of the variance in EU support remains unaccounted for. In terms of explaining cross-national variation in EU support, our explanation is the least successful. Clearly, other predictors should be considered at this level (perhaps in a design that includes multiple time points).

It is useful at this point to consider the implications of ignoring the multilevel data structure. Specifically, what inferences would we draw if we were to ignore the multilevel character of the EU data? To answer this question we estimated an OLS regression model on the data, treating them as a "flat" data structure instead of a hierarchy. The last column in Table (4) presents the results from this analysis. The noteworthy feature of these results is the large number of statistically significant predictors. Indeed, with the exception of the highest income quartile, all predictors are significant at the .10 level and most are significant at the .01 level.

These results stand in stark contrast with the multilevel results discussed earlier. The differences are particularly clear at the national level of analysis. Neither of the

national level predictors achieved statistical significance in the multilevel analysis. However, the OLS results suggest that both of these predictors are significant at the .01 level. Clearly, the inferences drawn from the regression analysis are different than those drawn from the multilevel analysis.

These differences arise precisely because the OLS standard errors are *too small*. This attenuation is caused by ignoring the clustering of the data. The OLS analysis assumes that we have 6354 independent observations in our data. With such a large  $N$ , it is not surprising that most of the predictors in Table (4) attain statistical significance. The problem, of course, is that we do not have 6354 independent observations. As the variance components in Table (4) show, considerable clustering remains even after controlling for the party- and country-level predictors—the observations are hence *not* (conditionally) independent. To pretend that they are independent is to assume that one has more information than really exists. Thus, the OLS analysis presents too optimistic a view about the significance of the predictors.

To continue with our exposition of the multilevel model, we now consider the effect of opinion leadership in greater detail by accounting for possible causal heterogeneity. In Table 4, opinion leadership had a fixed effect, in keeping with the theory developed by Inglehart (1970; Inglehart, Rabier, and Reif 1991). However, our own theory suggests that the effect of opinion leadership varies across parties as a function of party cues. Which of these theories are supported by the data?

Table 5 gives the ML estimates of the multilevel model in (28). Consistent with our predictions, the cross-level interaction between party cues and opinion leadership is positive and statistically significant at the .05 level. It seems that the effect of opinion leadership is moderated by party cues.

To obtain a better sense of the cue-dependence of opinion leadership we make some comparisons. The minimal value of party cues in the sample is -4.581 (on a mean-centered scale), indicating an anti-EU party position. For this position the expected effect of opinion leadership is -.068. Although small, this effect is opposite to the hypothesis that opinion leadership always enhances EU support. The maximum value of party cues in the sample is 1.419 (again on a mean-centered scale). For this pro-EU party position the expected effect of opinion leadership is .196. This is the expected positive effect, but its existence clearly depends on the nature of party cues.

Thus, it appears that the argument that the effect of opinion leadership is uniformly positive is incorrect. The multilevel statistical results clearly reveal considerable variation in the effect of opinion leadership across political parties. This suggests a different rationale for opinion

**TABLE 5** Model with Cross-Level Interaction

Parameter	Estimate
<i>Fixed Effects</i>	
Constant	5.507** (.220)
Tenure	0.015 (.014)
Trade	0.031 (.025)
Partisan Cue	0.225** (.028)
Lowest Income Quartile	-.104+ (.064)
Highest Income Quartile	0.046 (.058)
Ideology	0.019 (.015)
Opinion Leadership	0.148* (.034)
Male	0.088+ (.050)
Age	-.013** (.002)
Opinion Leadership $\times$ Partisan Cue	0.044* (.021)
<i>Variance Components</i>	
Country-Level ( $\omega_{00}$ )	0.548** (.212)
Party-Level Constant ( $\tau_{00}$ )	0.107** (.031)
Opinion Leadership ( $\tau_{44}$ )	0.019 (.012)
Constant, Opinion Leadership ( $\tau_{04}$ )	-.011 (.015)
Individual-Level ( $\sigma^2$ )	3.751** (.067)
-2 $\times$ Log Likelihood	26578.970

Note: Table entries are maximum likelihood (IGLS) estimates with estimated standard errors in parentheses.

+ =  $p < .10$ , \* =  $p < .05$ , \*\* =  $p < .01$

leadership effects; such effects are at least partially produced by cue-taking rather than an inherent tendency of opinion leaders to support the EU.

## Discussion

In this application, we have shown the use of a three-level multilevel model for understanding public opinion toward the EU. We asked three questions about this topic. First, are the individual, party and national levels all relevant for EU support? The answer to this question is affirmative. Second, can we account for the variation in EU support at these levels? Here we found that our predictors

provide an imperfect explanation at best. Finally, is the effect of these predictors heterogeneous? The answer to this question was affirmative for one of our predictors, opinion leadership, and resulted in an important modification of existing theory about the effect of this predictor.

While this particular example is interesting because of the obvious multilevel character of EU opinion data, the lessons that this application teaches are much more general. This application demonstrates all of the advantages of multilevel modeling that we discussed before. First, it demonstrates the ease with which causal heterogeneity is introduced in multilevel models. Our exploration of the varying effect of opinion leadership across political parties required only a few changes from the basic model and the interpretation was as easy as interpreting any interaction term.

Second, our application shows the advantages of multilevel modeling in comparison to dummy variable models and interactive models. Instead of including dummy variables at the party and national levels of analysis, we were able to include substantively interesting predictors. Unlike interactive models, however, our multilevel models did not assume that these predictors account perfectly for cross-party and cross-national variation in EU support, which would have been incorrect.

Finally, this application shows the statistical benefits of multilevel models. Too often political scientists ignore the multilevel character of the data they frequently work with. As we have demonstrated here, this can have pernicious effects on the statistical inferences that are drawn from those data. Multilevel models offer a statistical tool that can capture the data structure and thereby produce correct inferences.

## Conclusion

Researchers who wish to analyze multilevel data face the task of selecting a methodology that addresses the challenges such data pose and that allows exploitation of the opportunities offered by those data. In this paper, we have discussed one such methodology—multilevel modeling. Widely applied in other social sciences, we think multilevel analysis is a powerful tool that can greatly benefit political science—if used appropriately. Appropriate use is based on an awareness of the requirements and limitations of multilevel analysis, and on a sense of when this modeling strategy can be used most beneficially.

First, researchers should be aware that multilevel models are data intensive. A growing body of statistical literature suggests that sufficient power to test hypoth-

eses about cross-level interactions and variance components hinges on the availability of sizable numbers of contextual units (see Kreft 1996; Kreft and De Leeuw 1998; Raudenbush 1998; Snijders and Bosker 1994, 1999). This is a steep requirement that is not always met in political science data.<sup>27</sup> Multilevel models also place a hefty premium on valid and reliable measurements. Bad measures in multilevel models “get worse” because such a heavy demand is placed on the data in terms of estimating coefficients and variance components. In the absence of adequate measurement, it may be impossible to exploit the benefits of multilevel analysis (see Bryk and Raudenbush 1992).

Second, we should be equally aware that multilevel models are theory intensive. The specification and interpretation of multilevel models hinge on a theoretical understanding of the relevant levels of analysis (see Hox, Van Den Eeden, and Hauer 1990; Lazarsfeld and Menzel 1969; Opdenakker and Van Damme 2000; Van Den Eeden and Huttner 1982) and of the processes at work at each level of analysis and between (or among) levels of analysis. What complicates matters is that theory bridging the gap between micro and macro levels is still relatively scarce in political science. Microscopic and macroscopic explanations in political science, say of voting behavior, often have developed next to each other with few points of contact. Fortunately, this situation is changing, and we expect that cross-level theories will become more common.

Third, multilevel models increase the number of assumptions that one has to make about the data. Not only do we have to assume a distribution for the dependent variable, we also have to assume distributions for one or more of the parameters that link predictors to this variable. When these assumptions are incorrect, this could adversely affect statistical inference.

Multilevel models, then, make heavy demands on theory and on data. Thus, we caution researchers against “blindly” using these models in data analysis. Instead, we urge them to consider the full range of methods for handling clustered data that is now available. These methods include replicated sampling techniques (Lee, Forthofer, and Lorimor 1989), sandwich estimation of the standard errors (Huber 1967), generalized estimating equations (GEE; Liang and Zeger 1986; Zorn 2001), and multilevel modeling.

An important consideration in choosing between these methods is whether data clustering is merely a sta-

<sup>27</sup>Of course, if an interaction or variance component is large it will probably be detected even when the number of contextual units is small.

tistical nuisance or if it is of substantive interest. If clustering is just a nuisance, our first choice would not be multilevel modeling. Instead, we would use any of the other methods that we described (or a combination thereof), in large part because they make fewer assumptions about the data. However, if clustering is of substantive interest—if the dependencies between observations reveal a theoretically interesting aspect of reality—then multilevel is an ideal statistical tool.

*Manuscript submitted September 18, 2000.*

*Final manuscript received July 1, 2001.*

## References

- Achen, Christopher H., and W. Philips Shively. 1995. *Cross-Level Inference*. Chicago: The University of Chicago Press.
- Barcikowski, Robert S. 1981. "Statistical Power with Group Mean as the Unit of Analysis." *Journal of Educational Statistics* 6:267–285.
- Barnett, Vic. 1991. *Sample Survey Principles and Methods*. New York: Oxford University Press.
- Bartels, Larry. 1996. "Pooling Disparate Observations." *American Journal of Political Science* 40:905–942.
- Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to Do (and Not to Do) with Time-Series Cross-Section Data." *American Political Science Review* 89:634–647.
- Boyd, L.H., Jr., and Gudmund R. Iversen. 1979. *Contextual Analysis: Concepts and Statistical Techniques*. Belmont, CA: Wadsworth.
- Browne, William J., and David Draper. 2000. "A Comparison of Bayesian and Likelihood Methods for Fitting Multilevel Models." Unpublished Manuscript.
- Browne, William J., David Draper, Harvey Goldstein, and Jon Rasbash. 2000. "Bayesian and Likelihood Methods for Fitting Multilevel Models with Complex Level-1 Variation." Unpublished Manuscript.
- Bryk, Anthony S., and Stephen W. Raudenbush. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage Publications.
- Burton, Paul, Lyle Gurrin, and Peter Sly. 1998. "Extending the Simple Linear Regression Model to Account for Correlated Responses: An Introduction to Generalized Estimating Equations and Multi-Level Mixed Modelling." *Statistics in Medicine* 17:1261–1291.
- Caldeira, Gregory A., and James L. Gibson. 1995. "The Legitimacy of the Court of Justice in the European Union: Models of Institutional Support." *American Political Science Review* 89(2):356–76.
- Charnock, David. 1997. "Class and Voting in the 1996 Australian Federal Election." *Electoral Studies* 16(3):281–300.
- Chou, Chih-Ping, Peter M. Bentler, and Mary Ann Pentz. 2000. "A Two-Stage Approach to Multilevel Structural Equation Models: Applications to Longitudinal Data." In *Modeling Longitudinal and Multilevel Data: Practical Issues, Applied Approaches and Specific Examples*, ed. Todd D. Little, Kai U. Schnabel, and Jürgen Baumert. Mahwah, NJ: Lawrence Erlbaum Associates.
- De Leeuw, Jan, and Ita Kreft. 1986. "Random Coefficient Models for Multilevel Analysis." *Journal of Educational Statistics* 11:57–85.
- Dempster, Arthur P., Donald B. Rubin, and Robert K. Tsutakawa. 1981. "Estimation in Covariance Components Models." *Journal of the American Statistical Association* 76:341–353.
- Dielman, Terry E. 1989. *Pooled Cross-Sectional and Time Series Data Analysis*. New York: Marcel Dekker.
- Effron, Bradley, and Carl Morris. 1975. "Data Analysis Using Stein's Estimator and Its Generalizations." *Journal of the American Statistical Association* 70:311–319.
- Eichenberg, Richard, and Russell Dalton. 1993. "Europeans and the European Community: The Dynamics of Public Support for European Integration." *International Organization*, 47:507–534.
- Fienberg, Stephen E. 1980. *The Analysis of Cross-Classified Categorical Data*. Cambridge: MIT Press.
- Franklin, Mark N., Michael Marsh, and Lauren McLaren. 1994. "The European Question: Opposition to Unification in the Wake of Maastricht." *Journal of Common Market Studies* 32:455–472.
- Franklin, Mark N., Michael Marsh, and Christopher Wlezien. 1994. "Attitudes toward Europe and Referendum Votes: A Response to Siune and Svensson." *Electoral Studies* 13:117–121.
- Franklin, Mark N., Michael Marsh, and Cees Van Der Eijk. 1995. "Referendum Outcomes and Trust in Government: Public Support for Europe in the Wake of Maastricht." *Western European Politics* 18:101–107.
- Friedrich, Robert J. 1982. "In Defense of Multiplicative Terms in Multiple Regression Equations." *American Journal of Political Science* 26: 797–833.
- Gabel, Matthew. 1998. "Public Support for European Integration: An Empirical Test of Five Theories." *Journal of Politics* 60:333–354.
- Geddes, Barbara. 1990. "How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics." *Political Analysis* 2:131–150.
- Gibbons, Robert D., and Don Hedeker. 1997. "Random-Effects Probit and Logistic Regression Models for Three-Level Data." *Biometrics* 53:1527–1537.
- Goldstein, Harvey. 1986. "Multilevel Mixed Linear Model Analysis Using Iterative Generalized Least Squares." *Biometrika* 74:43–56.
- Goldstein, Harvey. 1991. "Nonlinear Multilevel Models, with an Application to Discrete Response Data." *Biometrika* 78:45–51.
- Goldstein, Harvey. 1995. *Multilevel Statistical Models*. London: Edward Arnold.
- Hedeker, Don, and Robert D. Gibbons. 1994. "A Random-Effects Ordinal Regression Model for Multilevel Analysis." *Biometrics* 50:933–944.
- Hofmann, David A., and Mark B. Gavin. 1998. "Centering Decisions in Hierarchical Linear Models: Implications for Research in Organizations." *Journal of Management* 24:623–641.

- Hox, Joop, Pieter Van Den Eeden, and Joost Hauer. 1990. "Preliminary Notes." In *Theory and Model in Multilevel Research: Convergence or Divergence?* ed. Pieter Van Den Eeden, Joop Hox and Joost Hauer. Amsterdam: SISWO.
- Huber, Paul J. 1967. "The Behaviour of Maximum Likelihood Estimates Under Non-Standard Conditions." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1:221–233.
- Huckfeldt, Robert and John Sprague. 1987. "Networks in Context: The Social Flow of Political Information." *American Political Science Review* 81:1197–1216.
- Huckfeldt, Robert, and John Sprague. 1993. "Citizens, Contexts, and Politics." In *Political Science: The State of the Discipline II*, ed. Ada Finifter. Washington, D.C.: American Political Science Association.
- Inglehart, Ronald. 1970. "Cognitive Mobilization and European Identity." *Comparative Politics* 3:45–70.
- Inglehart, Ronald. 1977. *The Silent Revolution: Changing Values and Political Styles among Western Publics*. Princeton: Princeton University Press.
- Inglehart, Ronald, Jacques-Rene Rabier, and Karlheinz Reif. 1991. "The Evolution of Public Attitudes toward European Integration: 1970-86." In *Eurobarometer: The Dynamics of European Public Opinion*, ed. Karlheinz Reif and Ronald Inglehart. London: Macmillan.
- Iversen, Gudmund R. 1991. *Contextual Analysis*. Newbury Park, CA: Sage.
- Jackson, John E. 1992. "Estimation of Models with Variable Coefficients." *Political Analysis* 3:27–49.
- Janssen, Joseph I. H. 1991. "Postmaterialism, Cognitive Mobilization, and Public Support for European Integration." *British Journal of Political Science* 21:443–468.
- Jones, Kelvin, Ronald John Johnston, and Charles J. Pattie. 1992. "People, Places and Regions: Exploring the Use of Multi-Level Modelling in the Analysis of Electoral Data." *British Journal of Political Science* 22:343–380.
- Jöreskog, Karl G. 1971. "Simultaneous Factor Analysis in Several Populations." *Psychometrika* 36:409–426.
- Kackar, Raghu, and David A. Harville. 1981. "Unbiasedness of Two-Stage Estimation and Prediction Procedures for Mixed Linear Models." *Communications in Statistics—Theory and Methods* 10:1249–1261.
- King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
- Kreft, Ita G.G. 1996. "Are Multilevel Techniques Necessary? An Overview, Including Simulation Studies." Multilevel Models Project at the Institute of Education, University of London.
- Kreft, Ita G.G., and Jan De Leeuw. 1998. *Introducing Multilevel Modeling*. London: Sage.
- Kreft, Ita G.G., Jan De Leeuw, and Leona S. Aiken. 1995. "The Effect of Different Forms of Centering in Hierarchical Linear Models." *Multivariate Behavioral Research* 30:1–21.
- Kreft, Ita G.G., Jan De Leeuw, and Rien Van Der Leeden. 1994. "Review of Five Multilevel Analysis Programs: BMDP-5V, GENMOD, HLM, ML3, and VARCL." *The American Statistician* 48:324–335.
- Kuk, Anthony Y.C. 1995. "Asymptotically Unbiased Estimation in Generalized Linear Models With Random Effects." *Journal of the Royal Statistical Society, Series B* 57:395–407.
- Lau, Richard R. 1989. "Individual and Contextual Influences on Group Identification." *Social Psychology Quarterly* 52:220–231.
- Lazarsfeld, Paul F., and Herbert Menzel. 1969. "On the Relation Between Individual and Collective Properties." In *A Sociological Reader on Complex Organizations*, ed. Amitai Etzioni. New York: Holt, Rinehart & Winston.
- Lee, Eun Sul, Ronald N. Forthofer, and Ronald J. Lorimor. 1989. *Analyzing Complex Survey Data*. Newbury Park, CA: Sage.
- Li, Fuzhong, Terry E. Duncan, Peter Harmer, Alan Acock, and Mike Stoolmiller. 1998. "Analyzing Measurement Models of Latent Variables through Multilevel Confirmatory Factor Analysis and Hierarchical Linear Modeling Approaches." *Structural Equation Modeling* 5:294–306.
- Liang, Kung-Yee, and Scott L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73:13–22.
- Lindley, Dennis V., and A.F.M. Smith. 1972. "Bayes Estimates for the Linear Model (with Discussion)." *Journal of the Royal Statistical Society, Series B* 34:1–41.
- Lijphart, Arend. 1971. "Comparative Politics and the Comparative Method." *American Political Science Review* 65:682–693.
- Longford, Nicholas T. 1987. "A Fast Scoring Algorithm for Maximum Likelihood Estimation in Unbalanced Mixed Models with Nested Random Effects." *Biometrika* 74: 817–827.
- Longford, Nicholas T. 1993. *Random Coefficient Models*. Oxford: Clarendon Press.
- MacKuen, Michael, and Courtney Brown. 1987. "Political Context and Attitude Change." *American Political Science Review* 91:471–490.
- Mason, William M., George Y. Wong, and Barbara Entwistle. 1983. "Contextual Analysis Through the Multilevel Linear Model." *Sociological Methodology 1983-1984*:72–103.
- McDonald, Roderick P. 1994. "The Bilevel Reticular Action Model for Path Analysis with Latent Variables." *Sociological Methods & Research* 22:399–413.
- Meijer, Erik, Rien Van Der Leeden, and Frank M.T.A. Busing. 1995. "Implementing the Bootstrap for Multilevel Models." *Multilevel Modelling Newsletter* 7:7–11.
- Morris, Carl. 1983. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association* 78:47–65.
- Muthén, Bengt O. 1989. "Latent Variable Modeling in Heterogeneous Populations." *Psychometrika* 54:557–585.
- Muthén, Bengt O. 1991. "Multilevel Factor Analysis of Class and Student Achievement Components." *Journal of Educational Measurement* 28:338–354.
- Muthén, Bengt O. 1994. "Multilevel Covariance Structure Analysis." *Sociological Methods & Research* 22:376–398.
- Opdenakker, Marie-Christine, and Jan Van Damme. 2000. "The Importance of Identifying Levels in Multilevel Analysis: An Illustration of the Effects of Ignoring the Top or Intermediate Levels in School Effectiveness Research." *School Effectiveness and School Improvement* 11:103–130.
- Przeworski, Adam, and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley-Interscience.

- Quillian, Lincoln. 1995. "Prejudice as a Response to Perceived Group Threat: Population Composition and Anti-Immigrant and Racial Prejudice in Europe." *American Sociological Review* 60: 586–611.
- Rasbash, Jon, William Browne, Harvey Goldstein, Min Yang, Ian Plewis, Michael Healy, Geoff Woodhouse, David Draper, Ian Langford, and Toby Lewis. 2000. "A User's Guide to MLwiN." London: Multilevel Models Project at the Institute of Education, University of London.
- Raudenbush, Stephen W. 1988. "Educational Applications of Hierarchical Linear Models: A Review." *Journal of Educational Statistics* 13:85–116.
- Ray, Leonard. 1999. "Measuring Party Orientations towards European Integration: Results from an Expert Survey." *European Journal of Political Research* 36:283–306.
- Richardson, A.M., and Alan H. Welsh. 1994. "Asymptotic Properties of Restricted Maximum Likelihood (REML) Estimates for Hierarchical Linear Models." *Australian Journal of Statistics* 36: 31–43.
- Rokkan, Stein. 1966. "Comparative Cross-National Research: The Context of Current Efforts." In *Comparing Nations: The Use of Quantitative Data in Cross-National Research*, ed. Richard L. Merritt and Stein Rokkan. New Haven: Yale University Press.
- Rubin, Donald B. 1981. "Estimation in Parallel Randomized Experiments." *Journal of Educational Statistics* 6(4): 377–400.
- Sayrs, Lois W. 1989. *Pooled Time Series Analysis*. Newbury Park, CA: Sage.
- Searle, Shayle R., George Casella, and Charles E. McCulloch. 1992. *Variance Components*. New York: John Wiley & Sons.
- Seltzer, Michael H., Wing Hung Wong, and Anthony S. Bryk. 1996. "Bayesian Analysis in Applications of Hierarchical Models: Issues and Methods." *Journal of Educational and Behavioral Statistics* 21:131–167.
- Siune, Karen, and Palle Svensson. 1993. "The Danes and the Maastricht Treaty: The Danish EC Referendum of June 1992." *Electoral Studies* 12:99–111.
- Snijders, Tom A.B., and Roel J. Bosker. 1994. "Standard Errors and Sampling Sizes for Two-Level Research." *Journal of Educational Statistics* 18:237–261.
- Snijders, Tom A.B., and Roel J. Bosker. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.
- Sörbom, Dag. 1974. "A General Method for Studying Differences in Factor Means and Factor Structure between Groups." *British Journal of Mathematical and Statistical Psychology* 27: 229–239.
- Steenbergen, Marco R., Whitt Killburn, and Jenny Wolak. 2001. "Affective Moderators of On-Line and Memory-Based Processing in Candidate Evaluation." Presented at the annual meeting of the Midwest Political Science Association.
- Stimson, James A. 1985. "Regression in Space and Time: A Statistical Essay." *American Journal of Political Science* 29: 914–947.
- Swallow, William H., and John F. Monahan. 1984. "Monte Carlo Comparisons of ANOVA, MINQUE, REML, and ML Estimators of Variance Components." *Technometrics* 26: 47–57.
- Swamy, P.A.V.B. 1970. "Efficient Inference in a Random Coefficient Regression Model." *Econometrica* 38:311–323.
- Swamy, P.A.V.B., and George S. Tavlas. 1995. "Random Coefficient Models: Theory and Applications." *Journal of Economic Surveys* 9:165–196.
- Thum, Yeow Meng. 1997. "Hierarchical Linear Models for Multivariate Outcomes." *Journal of Educational and Behavioral Statistics* 22:77–108.
- Van Den Eeden, Pieter. 1988. "A Two-Step Procedure for Analysing Multi-Level Structured Data." In *Sociometric Research 2: Data Analysis*, ed. Willem E. Saris and Irmtraud N. Gallhofer. New York: St. Martin's Press.
- Van Den Eeden, Pieter, and Harry J.M. Huttner. 1982. "Multilevel Research." *Current Sociology* 20:1–178.
- Western, Bruce. 1998. "Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modeling Approach." *American Journal of Political Science* 42:1233–1259.
- Winer, Benjamin J. 1971. *Statistical Principles in Experimental Design*. New York: McGraw-Hill.
- Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.
- Zeger, Scott L., and Rezaul Karim. 1991. "Generalized Linear Models with Random Effects: A Gibbs Sampling Approach." *Journal of the American Statistical Association* 86:79–95.
- Zorn, Christopher J.W. 2001. "Generalized Equation Models for Correlated Data: A Review with Applications." *American Journal of Political Science* 45:470–90.