MULTILEVEL MODELS:
Methods and Substance

Thomas A. DiPrete and Jerry D. Forristal
Department of Sociology, Duke University, Box 90088, Durham, North Carolina
27708-0088

KEYWORDS: contextual-effects models, methodology, micro-macro, hierarchical models,
theory specification

Abstract
This paper reviews recent developments in the application of multilevel models to substantive problems in sociology. There is no single multilevel model in sociology, but rather a set of more or less closely related approaches for exploring the link between the macro and micro levels of social phenomena. Methodological developments of the last ten years are discussed and contrasted with older methods. Illustrative examples of how multilevel analysis has contributed to sociological knowledge are provided for several areas of the discipline, including demography, education, stratification, and criminology. Cautions in the use of these models for empirical research are discussed, along with possible further developments.

INTRODUCTION

Multilevel models are used in sociology to specify the effect of social context on individual-level outcomes. The idea that individuals respond to their social context is a defining claim of the sociological discipline, which is found in Marx’s work on political economy (1846), in Durkheim’s studies of the impact of community on anomia and suicide (1897), in Weber’s research on how religious communities shape economic behavior (1905), in Merton’s work on communities, relative deprivation, and social comparison theory (1968), and in Berelson et al’s (1954) research into the effect of social context on voting. Blau (1960) and Davis (1961) published important papers devoted to contextual effects, and a thriving research agenda into these issues existed in the 1960s and 1970s (Boyd & Iversen 1979, Blalock 1984, Iversen 1991). More
recently, Coleman (1986) classified sociological theories into three groups according to their multilevel content. In the first group, variation in a dependent variable is explained through independent variables that apply to the same social level (e.g. country, organization, local community, individual). In the second group, variation in a dependent variable at one level is explained by processes that operate at a higher level. In a third type, variation in outcomes at one level is explained by variation in outcomes at a lower level. Theories in the second and third groups are multilevel theories, though nearly all existing multilevel research falls into group two.

Blalock (1984) reviewed the theoretical and methodological literature on multilevel models about 10 years ago. Most of his observations—particularly about the problems of interpreting multilevel effects and about consistency of estimates—are still valid. However, much has changed in the past decade, too. The most useful metatheoretical formulation of micro and macro levels of social reality and the relationship between those levels continue to be active topics for debate in theoretical sociology (Alexander et al 1987, Huber 1991). Multilevel theories for particular social phenomena have become more refined. Moreover, new statistical techniques have been applied in several areas of sociology, with useful results. These new methods are not a panacea for the problems raised by Blalock in his 1984 review, as Mason (1991) notes, but they do represent a significant advance over earlier methods, and they have led to new types of substantive analyses as well.

Because so much work is ongoing in so many different areas, we cannot attempt an exhaustive review of all sociological research that is arguably multilevel. Instead, we selectively review research in order to describe the recent methodological developments, to show how the newer methodological work is being used in substantive areas, to discuss how the new models relate to each other, and to identify areas where further developments are needed or in process.

The term *multilevel* can be used in different ways. There is no single multilevel model; there is rather a variety of models that have been used to analyze social processes postulated to operate at more than one level of analysis. One can characterize theories, methods, and data collection as either unilevel or multilevel; while they tend to be associated, the association is not perfect. Researchers sometimes theorize at multiple levels but use data at only one level. They sometimes theorize at one level, but use data at a different level. The issues raised by these efforts (aggregation, the "ecological fallacy" etc) are beyond the scope of this review (see Blalock 1984 for a summary). In this review, we use the term "multilevel model" to designate the specification and testing of multilevel theories with multilevel data.

Blalock (1984) defined contextual analysis as follows: "The essential feature of all contextual-effects models is an allowance for macro processes that are
presumed to have an impact on the individual actor over and above the effects of any individual-level variables that may be operating" (p. 354). If we generalize his use of the term *individual* to apply to any unit that is micro relative to some other macro level in the analysis, his definition is still quite serviceable. The notion of context is quite general and can include spatial contexts, (e.g. countries, states, communities), temporal contexts (i.e. history), organizational contexts (classrooms, schools, firms), and social/cultural/economic contexts (ethnic groups, social classes, economic sectors). Analogous models are employed when the "context" is the method of data collection or some other common property for a set of data points, as in meta-analysis—the use of statistical models to reconcile results from different studies on the same topic—or the analysis of factorial surveys (Hedges & Olkin 1985, Braun 1989, Hox et al 1991, Bryk & Raudenbush 1992). The theoretical specification of context can be simple, or rather involved (Blalock 1984, Iversen 1991; or at a more metatheoretical level, see Alexander et al 1987). Multiple contexts can apply to a given unit. Contexts can be overlapping or nested. They can have fuzzy boundaries or clear ones. Not surprisingly, more complex specifications can introduce formidable methodological difficulties.

Generally speaking, multilevel models explain microlevel outcomes in two ways: (i) by showing that parameters of models specified at the micro level—where microlevel covariates are used to explain microlevel outcomes—are a function of context, and (ii) by showing that this micro-macro relationship can be expressed in terms of characteristics of the context, which take the form of macrolevel variables. Two types of macrolevel variables are commonly found in multilevel models: (a) context-specific means or higher moments of microlevel variables, and (b) global variables (Lazarsfeld & Menzel 1969) that are not expressible as functions of microlevel variables, or at least not the microlevel variables found in the microlevel equation.

We have already noted implicitly that multilevel models are also called contextual models. Readers of the recent literature will have noted many other names for these models, including, but not limited to, hierarchical linear model, hierarchical linear regression, random coefficients model, hierarchical mixed linear model, or bayesian linear model (Kreft et al 1990). To a certain extent, the proliferation of names comes from the statistical properties of the various modeling strategies used to analyze multilevel data. These names apply primarily to the extensions of contextual regression analysis that have been used extensively in sociology since the 1960s (as reviewed in Boyd & Iversen 1979, or Blalock 1984).

However, several other types of multilevel models are also in use or being developed in sociology, including (i) extensions of contingency table analysis to include multilevel effects; (ii) extensions of event-history analysis to include multilevel effects; (iii) endogenous switching regressions; (iv) extensions of
latent-variable models to include multilevel effects; (v) the development of micro-macro (as opposed to macro-micro) models; and (vi) nonstatistical methods of multilevel analysis. We discuss these alternative methods later in this paper.

The familiar model for contextual effects and more recent extensions are the models most widely associated with the term *multilevel model*. These are regression models that are linear in their coefficients. Generally speaking, they are used to analyze data that consist of multiple macro units (contexts) and multiple micro units within each macro unit. These regression models can be expressed in two algebraically equivalent forms: (i) as an equation relating a microlevel outcome to a set of microlevel variables along with a set of equations in which the coefficients of this microlevel model are expressed as functions of macrolevel variables, or (ii) as a single equation where the microlevel dependent variable is expressed as a function of both micro and macro variables. This second form generally includes interactions between the micro and macro variables.

**The Basic Model**

The major advances in the past ten years concern not the functional form relating micro and macro variables, but rather a more sophisticated treatment of the error structure for these models. Whereas the older models can be characterized as fixed effects regression models, the new models specify the regression coefficients as random effects. In a fixed effects multilevel regression model, the microlevel coefficient is expressed as an exact function of macrolevel variables. Random effects multilevel models, in contrast, contain error terms in the macro equations. The inclusion of these error terms at the macro level implies a more complex error structure in the single-equation version of the multilevel model. The use of random coefficient models allows the data analyst to decompose the variance in the dependent variable into the within-context variance and the between-context variance, and to study these two sources of variation for the microlevel outcome. Thus, random coefficient multilevel models are a type of variance components model.

The distinction between the fixed effects multilevel model and the random effects multilevel model can be seen in a simple example, adopted from Bryk & Raudenbush (1992). Suppose that math achievement in elementary school ($Y_j$) for individual "i" is a function of socioeconomic status (SES), and suppose that the relationship follows the following simple equation:

$$Y_i = \alpha_0 + \beta_1 X_i + \varepsilon_i$$

where $X_i$ is the SES of student $i$. As written, the intercept $\alpha_0$ gives an estimate of the expected math achievement for a student whose socioeconomic status score is zero. Since zero is not a meaningful value for socioeconomic status,
we can “center” $X$ (e.g. Willms 1986, Iversen 1991, Bryk & Raudenbush 1992) as a deviation from the mean SES in the sample.

$$Y_i = \beta_0 + \beta_1(X_i - \bar{X}) + \epsilon_i$$

Here $\beta_0$ provides an estimate of the expected math achievement for a student whose SES is at the mean, and $\beta_1$ provides an estimate of the effect of a unit of SES on math achievement. This equation is a typical microlevel equation, which specifies the achievement of all students with mean SES to have the same expected math achievement $\beta_0$ and specifies the effect of SES ($\beta_1$) to be the same for all students. In other words, the effects in this model are equal for all students. In an extreme alternative to this fixed effects model, the analyst might hypothesize that each student has his/her own unique $\beta_0$, which would predict his/her math achievement if this student were to come from a mean SES family. Furthermore, the analyst might argue that each student has his/her own unique response to the academic advantages or disadvantages related to family SES. Under the assumption that each student has his/her own coefficients, the model now becomes:

$$Y_i = \beta_{0i} + \beta_{1i}(X_i - \bar{X}) + \epsilon_i$$

where $\beta_{0i}$ is the intercept for student “i,” and $\beta_{1i}$ is the effect of SES for student “i.” Because the parameters of Equation 1 vary by individual, this model is called a varying parameters model. If it is assumed that these parameters are fixed but unknown, then the model is a fixed effects varying parameter model (Judge et al 1985, Greene 1990). If the students in the sample constitute the entire population of students, the conceptualization of these parameters as varying but fixed is plausible on theoretical grounds. These parameters could not be estimated without further constraints, since the number of parameters is twice as great as the population size. Usually, however, the students in the study would only be a sample from some larger population, and it would make more conceptual sense to view the individual-level coefficients in Equation 1 as random draws from some larger population of coefficients, just as the students in the sample are random draws from some larger population of students.

Under this assumption, the coefficients can be reexpressed to indicate that they are varying and random.

$$\beta_{0i} = \bar{\beta}_0 + \nu_i$$
$$\beta_{1i} = \bar{\beta}_1 + \omega_i$$

Inserting these values back into the microlevel equation gives:

$^1$See Bryk and Raudenbush (1992) or Iversen (1991) for an extended discussion of the various types of centering employed in multilevel work.
This model, which is sometimes called the Swamy regression model (1971), is similar to Equation 1 except that it has a more complex, heteroskedastic error structure that includes the random components from Equation 2 multiplied by microlevel covariates. The statistical consequences of this change are fairly straightforward. Under the standard (strong) assumptions found in elementary discussions of regression, ordinary least squares (OLS) estimates of $\beta_0$ and $\beta_1$ are consistent, but OLS estimates of their standard errors are inconsistent, and OLS is less efficient than alternative estimation procedures.

Model 3 is clearly more general than Model 1, but it is obviously unsatisfying from a substantive point of view; it provides no explanation for how these coefficients vary across individuals. The random effects multilevel model can be viewed as a way of imposing further structure on Equation 3 and, in the process, providing an explanation for the varying coefficients (Kreft et al 1990). It does so in the specific case where the analyst can argue that these microlevel units are located in different and distinguishable social contexts, and also that the properties of these social contexts provide an explanation for variation in the microlevel coefficients.

In the student example above, a natural context is the school (other possibilities are the classroom or the school district). If we use the letter “j” to index the school of each student, we can replace Equations 2 by a more substantively meaningful set of equations.

\[
\beta_{0ij} = \gamma_0 + \nu_j \\
\beta_{1ij} = \gamma_1 + \omega_j
\]

In Equations 4, all the individual-level variation in microlevel coefficients is attributed to the school, and every student in a given school is presumed to have the same microlevel coefficients. In other words, the intercepts and slopes vary systematically by school, but there is no further variation within schools. One can go further and model the expected value for each school’s coefficients as functions of that school’s characteristics. For example, letting $Z_{ij}$ represent the mean SES in school “j”, and letting $Z_{2j}$ be a dummy variable that indicates whether school j is a Catholic school or not, we could reexpress Equation 4 as:

\[
\beta_{0ij} = \gamma_{00} + \gamma_{01}Z_{1j} + \gamma_{02}Z_{2j} + \nu_j
\]

The seminal discussion of this approach to the linear model is found in Lindley & Smith’s (1972) discussion of bayesian approaches to regression. Bayesian approaches are closely related to—and in some respects equivalent to—Stein-like estimators (Efron & Morris 1972, or Judge et al 1985 for a recent extended discussion).
\[ \beta_{ij} = \gamma_{10} + \gamma_{11}Z_{ij} + \gamma_{12}Z_{ij} + \omega_{ij} \]

or in a single equation form as:

\[
Y_{ij} = \gamma_{00} + \gamma_{01}Z_{ij} + \gamma_{02}Z_{ij} + \gamma_{10}(X_{ij} - \bar{X}) + \gamma_{11}Z_{ij}(X_{ij} - \bar{X})
+ \gamma_{12}Z_{ij}(X_{ij} - \bar{X}) + \{\varepsilon_i + \psi_j + (X_{ij} - \bar{X})\omega_{i}j\} \]

In Equation 6, the macrolevel variables interact with the microlevel variables, and the error structure in brackets contains both microlevel terms and macrolevel terms. If the data contain multiple microlevel observations within the same context, Equation 6 implies that the errors will be heteroskedastic and correlated across microlevel units. Under the usual assumptions, OLS estimation of Equation 6 will produce consistent estimates of the coefficients, but inconsistent estimates of their standard errors.

If the analyst knew the variance-covariance matrix for the error structure of Equation 6, it would be straightforward to estimate Equation 6 with generalized least squares (GLS) (see e.g. Goldstein 1987). Because this matrix is generally not known, however, it must be estimated along with the coefficients in Equation 6. The best methods for doing this are iterative methods such as the EM algorithm, which alternately estimates the parameters of Equation 6, then the variance-covariance matrix of the error, and so forth until convergence is reached (Mason et al 1983, Goldstein 1987, Bryk & Raudenbush 1992; see Kreft et al 1990 for a discussion of alternative iterative methods). These methods have been incorporated into a collection of standalone statistical programs such as GENMOD, HLM, ML3, VARCL, and routines that run under SAS, GAUSS, and BMDP (see Kreft et al 1990 for a discussion of many of these routines).

The basic difference between the random effects multilevel model shown in Equation 6 and the older fixed-effects multilevel model found in most of the literature reviewed in Boyd & Iversen (1979) or Blalock (1984) concerns the error structure shown in Equations 5. The fixed effects model assumes these errors are zero. This assumption converts Equation 6 into a standard regression that contains interaction terms, and the coefficients of this regression can be estimated using OLS without any of the statistical difficulties described above. As Bryk & Raudenbush (1992) note, the choice of fixed effects or random effects is not arbitrary, because the hypothesis of zero errors in the macro equations can be tested. It is always possible to convert a random-effects formulation into a fixed-effects formulation, in which the error variances in Equation 5 become zero, by including dummy variables for each context (with appropriate constraints to ensure identification) in Equations 4 or 5. This approach, however, has two difficulties: it introduces a large number of additional coefficients, and it blurs the distinction between sample and population...
that is preserved by the random-effects model. The random effects model is superior on both grounds.

Equation 6 incorporates the assumption that both the intercept and slope coefficients are functions of context. One could restrict the model so that only some slopes vary while others are fixed, or one could specify that only the intercept coefficients vary, while all slope coefficients are fixed. One could also specify that some but not all of the error variances in the macro equations are zero. These hypotheses are all testable (Bryk & Raudenbush 1992). Finally, it is not necessary that all of the within-context regressions be identically specified; all that is required is that one or more of the microlevel coefficients be comparable across contexts (Wong & Mason 1989).

The use of newer statistical techniques to study a particular multilevel problem can alter one’s substantive conclusions. Furthermore, the newer techniques can lead to interesting new substantive analyses. The importance of context in shaping outcomes is illuminated by an analysis of the variance components of the multilevel model. For example, if most of the variation in stratification outcomes occurs between families, or if most of the variation in achievement occurs between schools, one might reasonably conclude that the family is a major determinant of stratification outcomes, or that the school is a major determinant of achievement. In contrast, a finding that almost all the variation in stratification outcomes was within families, or that almost all the variation in achievement outcomes was within schools leads to different substantive conclusions. These types of analyses are natural products of a random-effects multilevel analysis (Goldstein 1987, Bryk & Raudenbush 1992, see also Hauser & Mossel 1985).

Random-effects multilevel models also allow more sophisticated attention to the task of estimating microlevel models for specific contexts. Within-context models are of particular interest to policy analysts when the contexts (e.g. schools) can be manipulated if their performance is wanting (Willms 1992). Multilevel data provide three distinct estimators for within-context coefficients. The analyst could use within-context data to estimate a within-context regression, but the estimates will be imprecise if the within-context sample is small. Alternatively, the analyst could use the macro model shown in Equations 4 or 5 to obtain estimates of the within-context coefficients, either by using the estimated mean of Equations 4, or by plugging in the appropriate values for the macro variables in Equations 5. This second estimator uses data for all contexts, not just the context for which coefficient values are desired, because the regression coefficients of the macro equation are estimated with data for all

3See, for example, Kreft & de Leeuw’s 1988 demonstration how conclusions about the effects of a school’s sex-ratio on individual reading ability changed when one used a random effects instead of a fixed effects multilevel model.
contexts. As a third possibility, one could weight the alternative estimators and average them. The random-effects multilevel model follows this third course; it provides an empirical-Bayes (EB) estimator for individual-level coefficients that is an "optimally" weighted average of the within-context estimate and the between-context estimate (Mason et al. 1983, Bryk & Raudenbush 1992). This weighted average "shrinks" the within-context estimate toward the between-context estimate, with the level of shrinkage determined by the precision of the within-context estimator (e.g. Rubin 1989).4 Because information is "borrowed" from other contexts, this technique provides more stable within-context estimates, though the new estimates may have greater bias if the macrolevel model is misspecified (Braun 1989, McCullagh 1989, Lewis 1989).

ELABORATIONS OF THE MULTILEVEL LINEAR MODEL

As Blalock (1984) and Iversen (1991) noted in their earlier reviews, the theoretical specification of context can become quite complicated. While the basic multilevel model discussed above incorporates a single context within which microlevel units are nested, multilevel effects can operate through multiple contexts at once, which can be either nested or overlapping. When multiple contexts are nested within each other (e.g. classrooms within schools within districts), the two-level multilevel model discussed above can be extended by specifying each of the coefficients in the second-level equation as a function of the attributes of a higher level of context. Goldstein (1987), Kreft et al. (1990) and Bryk & Raudenbush (1988; see also Bryk & Raudenbush 1992, Lee & Bryk 1989) have made extended presentations of this model.

A three-level model is also useful in the analysis of multilevel longitudinal data, i.e. multiple observations of micro-units within a set of macro-contexts. Bryk & Raudenbush (1987) proposed one such model to explain human development in an educational setting. In their model, the variables at the first level are time-varying individual-level variables, most notably, time itself. The second level is also specified at the micro level: it contains time-invariant individual effects, which are used to explain the first-level coefficients of the time-varying variables. The third level contains macro variables, which in Bryk & Raudenbush’s model were attributes of schools. Willms & Raudenbush (1989) proposed a different three-level model for longitudinal analysis. Their first level model contains student-level variables, measured at a particular time. The second level equation models the coefficients of the first-level equation

4See also Clogg et al (1991) and Stasny (1991) for related uses of shrinkage estimators to stabilize within-group estimates when sample sizes are small. Also see the related discussion of empirical Bayes meta-analysis (Braun 1989, Bryk & Raudenbush 1992). Here the various studies function as alternative "contexts."
as functions of school characteristics, which are measured contemporaneously with the individual-level measures in the first-level equation. These school characteristics at the second level are expressed as deviations from the average effects for the school over time. The third level model then specifies the coefficients of these time-varying school variables as functions of school characteristics averaged over time. This model is designed to distinguish the effects of short-term and long-term characteristics of schools on student achievement.

Nonnested models have also been developed for sociological problems. One type of nonnested model arises when two macrolevel contexts are presumed to affect the same microlevel unit at the same time. By crossing the two dimensions along which these multiple contexts are defined, one obtains a larger number of multidimensional contexts, such that each micro unit is nested within a single multidimensional context. With this modification, the process can be studied with the usual two-level multilevel model (Braun 1989).

A second type of nonnested model can be used to allow contextual effects to be heterogeneous within contexts. The usual multilevel model assumes that the effects of context are homogeneous for all units located in the same context. As Blalock (1984) noted, this assumption of homogeneity may often be unwarranted. As a remedy, Blalock suggested that actors be divided into groups that have homogeneous “attachments, dependencies, and vulnerabilities.” DiPrete & Grusky (1990a,b) employed a similar strategy when they developed a three-level nonnested model to overcome the homogeneity problem in their analysis of temporal variation in the coefficients of the basic status attainment equation. Their three-level model used data from repeated cross-sectional samples of the American population. The first-level equation was a basic microlevel status attainment equation, with individual-level variables measured at the different time points available in the repeated cross-sections. The second level expressed the coefficients of these individual-level variables as functions of lagged macro factors. The third level permitted the strength of the macro factors on individual processes to vary across individuals. DiPrete & Grusky used an individual-level variable—a measure of labor force experience—to capture the differential vulnerabilities of workers to macro processes, but in principle these vulnerabilities could have been specified as a function of other macro variables such as firm size, labor market sector, etc.

An alternative elaboration of the basic multilevel model involves relaxing the assumption that the dependent variable is a continuous, normally distributed, random variable. Wong & Mason (1985) developed a hierarchical logistic regression model for multilevel analysis. Goldstein (1987) developed a hier-

5In a school context, for example, Willms (1986) speculated that the impact of school composition on various outcomes might be a function of the initial ability of the pupil.
architectural loglinear model. Other such work is ongoing (e.g. Kreft et al 1990, Goldstein 1991, Breslow & Clayton 1993).

OTHER FORMS OF MULTILEVEL MODELS

Most of the discussion to this point has concerned the analysis of data for multiple contexts, where several microlevel data points are available for each context. This type of data allows a variance components analysis that can separate the within-context and between-context contributions to total variance. In some cases, however, researchers may develop a multilevel theory and collect multilevel data, but cannot use variance components multilevel models because the researchers do not have measurements for multiple micro units within each macro unit. A notable case in point is found in contemporary research in stratification and work. A central theme in the recent literature is the influence of work organization and labor market segment on stratification outcomes. Several scholars (Villemez & Bridges 1988, Parcel et al 1991, Tomoskovic-Devey et al 1993) have collected the names of employers from individual survey respondents to generate a sample of organizations from which additional information was collected. The combination of the individual and the employer surveys is a multilevel database on workers and their work context. The most recent data of this form comes from the National Organization Survey (NOS), which combines microlevel data from the General Social Survey with macro data about the employing organizations of GSS sample members and their employed spouses. Unfortunately, because these datasets typically contain information on only one microlevel unit within each macrolevel unit, it is not possible to estimate the variance components multilevel models described above, and corrections for the heteroskedasticity in random-effects multilevel models fit to these data must follow the modeling strategies discussed, for example, by Maddala (1977) or Muthén & Satorra (1989).

Analyses with data such as the NOS can specify microlevel coefficients as functions of macrolevel variables, but they cannot estimate the "total effects" of context because only one micro observation is available per organization. Another important modeling strategy in stratification research—sibling models—in effect does the reverse. Hauser & Mossel (1985, see also Hauser & Sewell 1986, Hauser & Wong 1989) note that their sibling model is a specification for the contextual effects of the family of origin on stratification outcomes—in fact, it can be conceptualized as a random effects model for the intercept of a regression of status on schooling. Unlike the variance components models discussed earlier or the models for the effects of work organization just discussed, Hauser et al's sibling model was not designed to provide a substantive explanation for varying coefficients of individual-level variables (e.g. schooling, motivation, mental ability) as a function of family-level vari-
ables (e.g. father’s occupational status, mother’s education). Rather, it removes the error component in the status attainment equation that arises from unmeasured family effects, which—because they are correlated with schooling—normally create an upward bias in estimates of the effects of schooling on outcomes. Models of sibling resemblance have not yet been extended to consider family-based variation in other coefficients of the basic status attainment equation. A different multilevel modeling strategy has been employed in the analysis of mobility data arranged in contingency tables. As Iversen (1991) notes, some of the earliest contextual models were formulated as cross-classifications of microlevel and macrolevel variables (Kendall & Lazarsfeld 1950, Berelson et al 1954, Davis et al 1961). The more recent work goes beyond these early efforts and follows the same basic strategy used in multilevel linear regression models. In a single context (usually country), the mobility structure is reflected in the parameters of a loglinear or logmultiplicative model. Grusky & Hauser (1984) combined data for several countries and modeled the variation in parameters for each country’s model as a function of country characteristics such as political structure and level of economic development (see also Hauser & Grusky 1988, Erikson & Goldthorpe 1992). These models are fixed-effects models. Random-effects multilevel loglinear or logmultiplicative models have not been applied to sociological research to our knowledge. Another related multilevel model is found in event-history analysis. Event-history models are related to the life-course analyses mentioned earlier (e.g. Willms & Raudenbush 1989, Raudenbush & Bryk 1989), though time is measured continuously instead of discretely, and the outcome variable is discrete instead of continuous in event history models. Scholars have used multilevel event-history models to specify community effects in outcomes such as the timing of marriage or fertility (e.g. Billy & Moore 1992, Lichter et al 1992). Except in the case of models for unmeasured heterogeneity (discussed below), almost all the work to date has employed fixed-effects versions of these models, where interactions between macro and micro variables are added, but no adjustment is made for a possibly more complex error structure. Little work has been done with event-history models on the analysis of variance components or in the use of EB methods to obtain more stable estimates of context-specific event-history models (though see Braun 1989 for some discussion of these issues). Endogenous switching regression models can also be viewed as a form of multilevel model (Greene 1990, for applications see, e.g. Dickens & Lang 1988, Gamoran

6However, the downward bias on the schooling coefficient from measurement error appears to be roughly the same size as the upward bias on the schooling coefficient from omitted family variables (Hauser & Mossell 1985).

7Searches for interactions between family characteristics and the schooling coefficient with conventional regression techniques have found little of interest to date (Jencks et al 1971, Hauser 1973).
Researchers using this strategy specify separate regression models for each of a small number of distinct contexts (e.g., the primary or secondary sector of an economy, or the different educational tracks in high school), and they also specify a model that explains the observed distribution of micro units across the contexts. Because the number of contexts is small, macro variables cannot be introduced to provide quantitative explanations of between-context differences in regression coefficients.

Latent structure models are related in some ways to multilevel models. A latent structure model can be thought of as a type of multilevel model in which associations between microlevel variables are "explained" by the uncovering of latent classes which arguably generate the microlevel associations (Goodman 1974, Clogg 1981, McCutcheon 1987). A latent structure model is analogous to a regression model in which the measured covariates are conceptualized as indicators of latent contexts, without "true" direct effects of their own. The intercepts of the microlevel models are specified to vary by latent context. A computer search is done in order to identify latent contexts that minimize the measured direct effects of the covariates, and maximize the variation explained by the varying intercepts. The latent classes cannot be interpreted as pure contextual effects, however, because they will generally include individual-level common factors as well.

Unmeasured heterogeneity models in event history analysis are also related to multilevel models (Heckman & Singer 1984, Manton et al 1992). The presence of duration dependence in event-history models is often interpreted to mean that units in the analysis differ in ways not captured by the measured covariates. These unmeasured factors will bias estimates of duration dependence, and correlations between these unmeasured factors and measured covariates will bias the estimated coefficients of the measured covariates. The solution increasingly used in the analysis of event-history models is in effect to specify the intercept of event-history models as a random function of unmeasured person effects, then to specify a parametric or mixing distribution for these unmeasured effects, and then to incorporate this mixing distribution into the likelihood function. However, the mixing distribution is usually not given a substantive interpretation, and it will generally be a function of micro as well as macro factors.

Perhaps the use of latent variables that is closest to multilevel analysis is...
when scholars specify parallel covariance structure models for several distinct contexts (e.g. Kohn et al. 1990). In the work by Kohn and associates, cross-national variation in model coefficients is given a multilevel interpretation, but the variations are not quantitatively modeled as a function of macro variables. Of course, when the number of macro contexts is small—as is the case when a small number of countries are compared—it is not possible to model variation in micro parameters as a function of macro variables, other than to test for coefficient differences. Significant variation in the microlevel coefficients can of course be attributed to various attributes of the macro contexts, but confidence in the validity of such assertions must be gained by identifying other factors (which could be macro variables or coefficients of other micro equations) that vary with these attributes in ways predicted by theory.

Related issues exist in the growing literature in comparative-historical sociology. Comparative-historical scholars have paid a great deal of attention to the distinction between "small-N" and "large-N" studies (Ragin 1987), but they have not yet studied in detail the benefits from combining "medium-size N" macro and "large-N" micro studies. One explanation for this omission is that comparative-historical researchers as a rule prefer "rich" complex comparisons that require them to put a great deal of effort in a small number of cases. Time and money constraints limit their ability to include the number of macro cases required for statistical analysis. In addition, they have not yet pursued in sufficient depth the possibilities that even "small-N-macro"/"large-N-micro" studies can afford, of the form used by Kohn and associates. The number of such studies is beginning to grow, though they differ from the statistical models discussed above in that cross-context comparisons of within-context results are often done in a qualitative rather than a quantitative way (Janoski 1991; see also Hage et al. 1989, Flora et al. 1988a,b).

Lastly, there have been some nascent efforts to generate models that explain macro processes or outcomes as a function of micro-level processes and outcomes. While the importance of using microlevel processes to explain macrolevel outcomes has long been recognized in the social sciences (Merton 1936, Olson 1965, Schelling 1978), relatively little research has attempted to specify models that characterize or explain processes at the macro level by using microlevel data. Coleman & Hao (1989) propose one such method by looking at individual-level exchanges to characterize the exchange relationships in a system. This model differs from more conventional multilevel regression models, however, in that it employs no macro characteristics—either as independent or as dependent variables—other than those revealed through data analysis of micro-level relationships among actors (e.g. the

For an example of a network analysis that looks instead at how macro network level properties affect micro behavior, see Markovsky et al. 1988).
“price” of items being exchanged). Models that use microlevel relationships to explain characteristics of the system other than those that emerge from the model itself are still underdeveloped.

SUBSTANTIVE RESEARCH WITH MULTILEVEL MODELS

The methodological developments discussed above have been used to study a variety of substantive problems, many of which have been alluded to. In the space available, we can only briefly illustrate the utility of multilevel models for substantive research.

In the area of demography, random-effects multilevel models have, to date, been used primarily to analyze the additive and interactive effects of community-level characteristics on the reproductive behavior of women in developing countries (Casterline 1985). Entwisle & Mason (1985) used multilevel models to study the connection between socioeconomic status and fertility. They found that, while socioeconomic factors affect the number of children ever born, the magnitude of these effects depends on a country’s socioeconomic development and on the strength of its family planning programs. Countries at a low level of socioeconomic development and with no family planning programs have positive micro socioeconomic fertility differentials. Countries at a high level of socioeconomic development have strong negative individual-level socioeconomic differentials when family planning programs are weak, but weaker individual-level socioeconomic differentials when family planning programs are stronger. In a study of contraceptive use, Entwisle et al (1986) find that while individual socioeconomic effects on contraceptive use are strong, country-level effects are weak, in contrast to some earlier studies (see also Entwisle et al 1989).

Some recent efforts have also been made to examine the main effects of various community-level factors on the reproductive and marital behavior of young adults in the United States. Hogan & Kitagawa (1985) and Hogan et al (1985) found that growing up in a poor, highly segregated community decreases a black female’s chances of using contraception at first intercourse and significantly increases her risk of having a premarital pregnancy. Crane (1991) recently found that the likelihood of having a birth among young black and white females is inversely related to the quality of local neighborhoods, as measured by the proportion of all workers in the community who were employed as professionals and managers. Finally, Billy & Moore (1992) found that the proportion of women working full-time, the proportion of white collar workers, and the proportion of females who are currently separated or divorced in the community are all inversely related to the risk of experiencing a birth among married nonblack women, while the female unemployment rate, the
sex ratio of the never married population, and the child/woman ratio for women aged 15 to 24 are positively associated with the risk of experiencing a birth for unmarried nonblack women.

The recent study by Lichter et al (1992) is the first to use multilevel models to examine the main effects of marriage market factors on first marital transitions of young black and white women in the United States. These authors found that while differences between blacks and whites in marriage market opportunities are quite large, such differences are unable to account for racial differences in the timing of first marriage. Three additional studies, which examine the impact of factors such as the proportion of female headed families, the proportion of adults with a college degree, or the divorce rate, generally support Lichter et al's conclusion (Hoffman et al 1991, Li 1992, Forristal unpublished). Studies of community factors on rates of marriage and fertility have so far largely focused on the main effects of context; studies that examine whether macro variables affect the values of microlevel coefficients other than the microlevel intercept remain to be done.

The literature on school effects is voluminous. Coleman et al (1982) found that SES effects on test scores are flattened in Catholic schools compared with public schools, and Lee & Bryk (1989) extended their research by using random effects multilevel models. Their research also suggests that the effects of race are reduced in orderly schools, while the effects of class and academic background are reduced in smaller schools, in schools where the math curriculum is more homogeneous, and in schools where discipline procedures are fair and effective. Willms & Raudenbush (1989), in their analysis of the stability of the effects of school composition and policy on student achievement over time, found that variation in school effects for a given school tends to be small relative to the variation in school effects across schools. Other studies have looked at the strength of local labor market effects on decisions to remain in school (Raffe & Willms 1989), and sex differences and school effects in the growth in mathematics skills in lower and middle schools (Willms & Jacobsen 1990). Still other studies have examined the effects of placement within educational tracks, or attendance at schools that use educational tracks on achievement outcomes (e.g. Gamoran & Mare 1989, Gamoran 1992, Kerckhoff 1993). A complete review of the multilevel educational research would require an entire review paper (see Raudenbush & Bryk 1986, Goldstein 1987, Bryk & Raudenbush 1988, Bock 1989, Willms & Jacobsen 1990, Rumberger & Willms 1991, or Crane 1991, for examples of related research).

In the area of stratification, Grusky & Hauser (1984) examined the impact of societal attributes in explaining cross-national variation in social mobility. They found reasonably strong though complex effects of economic development on social mobility, and they found that the effects of political organization were equally strong. Treiman & Yip (1989) studied how macro factors might
explain cross-national variation in the coefficients of the basic status attainment equation. They found that industrialized societies tend to be more open than developing societies, and that societies which are more industrialized and which have more status equality show stronger effects of education, and weaker effects of family background on status attainment. Erikson & Goldthorpe (1992), in contrast, concluded that cross-national variation in social mobility is largely unsystematic once differences in occupational distributions are controlled. At a more micro contextual level, research into neighborhood effects on poverty especially in the United States found significant negative effects of coming from a welfare-dependent community or from a highly segregated community, but many effects of neighborhood context on economic outcomes appear weak (Datcher 1982, Mayer & Jencks 1989, Tienda 1991, Garner & Raudenbush 1991, Massey et al 1992, Corcoran et al 1993).

Meanwhile, other scholars have examined how social change within a given society can change the process of stratification. DiPrete & Grusky (1990b) found that the gradual growth of bureaucratic personnel policies in the United States may have played a role in equalizing opportunity, but that direct effects of political intervention may have been more important sources of change. At the third level of their multilevel model, they showed that—consistent with prevalent theories about how labor markets operate—the magnitude of these changes was stronger for new entrants into the labor market than for experienced workers. In addition, many studies have analyzed trends in the coefficients of stratification models without necessarily introducing macro variables to explain these trends (Hauser & Featherman 1978, Mare 1981, Hout 1988, Grusky & DiPrete 1990, Blossfeld & Shavit 1993).

A huge literature examines the impact of organizational structure and labor market segment on career outcomes (Baron 1984, Rosenfeld 1992), usually with data from specific firms or with data about firms and labor markets collected from individuals. But, as noted above, some recent research uses multilevel data that was collected at both the individual and the firm level. For example, Villemez & Bridges (1988) have explored the influence of firm size and other organizational characteristics on earnings, and how these effects vary by gender and occupation. Published research based on the new National Organization Study is expected in the near future.

In the area of criminology, several scholars have examined the impact of contextual factors on individual risks of victimization (Sampson et al 1987, Smith & Jarjoura 1989, Kennedy & Forde 1990, Miethe & McDowall 1993; PW Rountree, KC Land, TD Miethe, unpublished ms). For example, Miethe & McDowall (1993) showed that community measures such as the extent of family disruption, density of ownership of various consumer goods, racial heterogeneity, median income, proportion living alone, unemployment rates, and divorce rates influence risks of victimization. Aside from their finding
that individual safety precautions against burglary had a greater impact in more affluent neighborhoods than in poorer or commercial neighborhoods, community characteristics appeared to have only weak effects on the relationship between microlevel variables and victimization, though Rountree et al found evidence that interactions between community and individual-level variables are larger than would appear from estimation of a fixed effects model.\(^\text{10}\)

Multilevel research is found in many other areas of sociology and related fields aside from what is covered in our illustrative discussion above. The main purpose of the above discussion is to present evidence that, for some sociologically important outcomes, context matters. It is probably a fair generalization to state that most such evidence involves the main effects of context on behavior, though important research has obviously also been done in studying the interaction structure, too. The evidence certainly does not justify the conclusion that context always matters when theory suggests it might, nor that its effects, net of other factors, will always be substantively important (Mason 1991). Furthermore, the above discussion does not address the underlying theoretical questions concerning the correct interpretation of measured contextual effects (assuming these measured effects are, in fact, "real" and not the product of faulty estimation).

With sociology still limited to "middle range" theorizing (Merton 1968), there is no general theory of multilevel interrelationships. Theoretical research in different substantive areas, however, is proceeding toward the goal of constructing plausible explanations for what context means and how it might affect individual behavior, while empirical research tests these ideas with multilevel data. The amount of multilevel research in sociology should grow rapidly as theories are refined, as already-researched problems are revisited with newer methods, and as new data allow the testing of new hypotheses on the relationship between macro and micro.

**THE NEED FOR CAUTION**

Despite the advantages of the new techniques for multilevel analysis, one must be careful in applying these techniques to substantive problems. The new techniques, while superior to the older ones, rely on assumptions that often will not hold in specific substantive contexts.\(^\text{11}\) Some of the potential problems are:

\(^{10}\)Note that the interaction effects are not necessarily increased in size when a random-effects model is used. Cf Kreft & de Leeuw (1988).

\(^{11}\)Partially overlapping lists of the potential difficulties raised by multilevel modeling can be found in Blalock (1984), Kreft et al (1990), and Mason (1991).
[1] Model complexity. The estimation of the coefficients of random-effects multilevel models is not trivial, and the sociological imagination can easily outrun the capacity of the data, the computer, and current optimization techniques to provide robust estimates. Kreft et al.'s observation in this regard is trenchant: "Investigators (if the past is any indication) will tend to choose models that are too complicated (five levels, with 10 variables on each level). This leads to impossibly difficult search problems over the space of models and to impossibly difficult likelihood maximization problems" (Kreft et al 1990, p. 99).

[2] The assumption of fixed regressors. The statistical theory underlying random-effects multilevel models assumes that the regressors are fixed and estimates the models conditional on their values. But in most practical applications, the regressors are random, and unconditional estimation is desired (Kreft et al 1990).

[3] The problem of missing data. While increasingly sophisticated procedures are now available for handling missing data in statistical analysis (Little & Rubin 1987), existing software for random-effects multilevel regression models does not provide for internal (to the software package) treatment of the missing data problem. Given the growing evidence that naive treatment of the missing data problem produces biased estimates (Wang et al 1992), and given the amount of computation required to generate random-effects multilevel regression estimates, it is quite possible that an analyst who is faced both with missing data problems and with constraints on computing would do better by combining sophisticated methods for handling missing data with simple fixed effects multilevel regression methods than by combining naive methods for handling missing data with sophisticated random-effects multilevel regression methods. The former strategy would yield consistent estimates of the regression parameters, though biased estimates of standard errors, but the latter strategy might produce inconsistent estimates of both the regression parameters and their standard errors.¹²

[4] Correlation between micro variables and errors in the macro equation. As Equation 6 shows, the error in the macro equation is a component of the error in the micro equation, too. If this error is correlated with microlevel regressors, the estimates of the microlevel coefficients will be biased. In this situation, the use of dummy variables to convert a random-effects multilevel regression model to a fixed effects multilevel regression model may be advised (Judge et al 1985, Greene 1990).

[5] Normally distributed errors. Existing statistical packages generally assume normally distributed errors. Little is known about the robustness of this

¹²See Muthén (1993) for a discussion of recent progress in the treatment of missing data in the context of a random effects multilevel model.
assumption. If the microlevel errors are normal or can be made normal through a suitable transformation, nonnormality in the macro equation can be addressed as in [4], through the use of dummy variables to convert from a random-effects to a fixed-effects multilevel regression problem.

[6] Correlation between macro variables and errors in the macro equation, or between either macro or micro variables and errors in the micro equation. This problem bedevils all research that employs nonexperimental data to make causal inferences (e.g. Holland 1986, Berk 1988; see also Cook & Campbell 1979, Blalock 1984, Bryk & Raudenbush 1992). If the errors in the macro equation are correlated with the measured macro-level variables, the macro-level coefficients are biased, and a causal interpretation may be inaccurate. Similarly, if the error in the microequation is correlated either with microlevel or macrolevel variables, the estimates will be inconsistent.

There are various reasons why such correlations may arise. Hauser (1970, 1974) discussed the case where individual-level variables are poorly measured, such that macrolevel variables function as proxies for microlevel variables in a regression. For example, average SES in a school might be a proxy for individual SES if the latter variable is measured with error. Another problem occurs when characteristics of measurement error vary by context. For example, if measurement error for SES was greater in Catholic than in public school samples, or if the true variance of SES in Catholic schools was restricted relative to public schools while the error variance was the same in the two contexts, the measured effect of SES on achievement in Catholic schools might be smaller than in public schools when the true effects are the same size (RM Hauser, personal communication). In all of these examples, what appears to be a multilevel effect may be an artifact of the failure to correct for measurement error in the microlevel model.

Another situation where such correlations could arise is when the macro-level variables are in some sense endogenous to the microlevel process (Blalock 1984). To take an example from the published controversy over Coleman et al.'s report on Public and Non Public Schools (1982), if parents of children who—for unmeasured reasons—are high achievers seek out Catholic schools, the correlation between the individual-level errors and the school type variable would lead to bias in the effect of school type in the macroequations (e.g. Goldberger & Cain 1982). As with the problem of measurement error, selectivity bias is not eliminated by the use of more sophisticated multilevel procedures. Combining a sophisticated treatment of measurement error, selectivity bias, and related problems with the use of simpler fixed effects multilevel regression models may often produce more accurate results than a strategy which combines naive treatment of these issues with random-effects multilevel modeling.

[7] The assumption of exchangability: An assumption underlying the use of empirical Bayes procedures to increase the reliability of within-context esti-
mates is that the different contexts be "exchangeable" (Lindley & Smith 1972). In the absence of explicit covariates in the macro equation, this amounts to assuming that the microlevel coefficients differ across contexts only by chance variation. In the presence of explicit covariates in the macro equation, this amounts to assuming that the residual variation in the macro equations is unsystematic. If the contexts are exchangeable, then the researcher can improve the precision of within-context coefficient estimates by "borrowing" information from other contexts. However, if the contexts are not exchangeable—in other words, if the distinctiveness of a given context is not fully captured by the macrolevel variables—then the "shrinkage" estimator for the context-specific coefficients may be biased (McCullagh 1989, Bryk & Raudenbush 1992). Since we do not know how bad the bias is, we do not know (e.g. in a mean-square error sense) whether the tradeoff is worthwhile for any particular set of within-context estimates. Furthermore, the bias-variance tradeoff may differ for each context in the model.

The failure of exchangeability may be viewed as a problem of data collection at the macro level, or it may be viewed as a problem of sample size at the macro level. If the number of factors that make contexts distinctive is large relative to the number of contexts, then the ability of the analyst to model these variations is obviously limited. Arguably, this situation is more of a problem for the analysis of complex, unique entities such as countries than it is for more standardized organizations such as schools. When analyzing a set of unique contexts, the justification for performing random-effects regression models may be weak, and a fixed-effects dummy-variable model may be more appropriate (Greene 1990). Unfortunately, there is no easy answer to the question of when a fixed or a random effects model is more appropriate.

DIRECTIONS FOR FUTURE RESEARCH

The rate of development in the methodology of multilevel models and in their substantive application to a variety of sociological problems has been enormous in the past decade. This rate is likely to continue in the near future (e.g. Hox & Kref 1994). A variety of methodological problems await solutions. At the mundane level, there is the prospect of faster computational algorithms and more capable hardware that will allow users to analyze more complex problems. We can also hope for a closer linkage between the literature on multilevel estimation and multilevel study design. Often a tradeoff exists between the number of microlevel observations that can be collected within each context and the number of contexts that can be sampled. Users await useful guidance about how they should allocate their data collection resources.

We should also expect the development of more general multilevel models and software with which to estimate them. One such development would be the extension of the random-effects multilevel regression model to the case of
event-history analysis, while another would be a similar extension for loglinear and related models for contingency tables. Because many common event history models and models for the analysis of contingency tables are generalized linear models (McCullagh & Nelder 1989), developments of multilevel statistical theory for this more general case (along with appropriate computer software) would have wide applicability in sociology. Generalizations of the generalized linear model to include random terms in the linear predictor are called generalized mixed linear models (GLMM), and recent work by e.g. Goldstein (1991) or Breslow & Clayton (1993) on approximate methods of inference in GLMMs may lead to improved practical approaches to the estimation of multilevel event history models, multilevel loglinear and logmultiplicative models, and other related multilevel models in the near future.

A second important development would be the development of multilevel models for systems of microlevel and macrolevel endogenous variables. The current techniques can handle only a single microlevel dependent variable. Research is currently underway to create a "path analysis" for multilevel research (Kreft et al 1990, p. 23).

A third hoped for development would be the extension of latent variable methods so that they could be applied to multilevel problems. One welcome advance would involve the case where the microlevel variables are latent, while the contexts and macrolevel variables are measured exactly. As Muthén & Satorra noted (1989), estimation of such a model requires both further statistical development and new statistical software (see also Muthén 1994).

A more ambitious latent variable model would allow the macro contexts to be latent as well. The use of latent contexts would obviously require that the macro variables for these contexts also be latent, since the boundaries of observable "contexts" would not exactly correspond to the boundaries of latent contexts. Such a model might be viewed as an extension of latent class analysis. Instead of searching for latent classes that explain variation in some outcome variable by reducing the partial association (conditional on the latent class) between measured covariates and this outcome variable to a minimum, an extended latent class model might be employed to explain remaining variation in the effects of covariates as a function of properties of the latent classes, modeled as a function of macro indicator variables.

For example, several scholars in the 1970s and 1980s employed notions of a dual economy or a segmented labor market in attempts to explain variation in stratification outcomes (Hodson & Kaufman 1982). However, these structures were not precisely measured; often it was not obvious which segment any given worker belonged to. This fuzziness did not prevent scholars from proposing variables that were indicators of market segment (e.g. measures of technology, firm size, organizational and industry attributes, or occupational characteristics). In a latent multilevel model, market segments might be latent contexts that are
explained by and indicated by firm, occupational and industrial characteristics. The latent properties of these latent contexts would explain coefficient variation in microlevel equations for stratification outcomes. Given the formidable computational difficulties that already exist for current versions of multilevel models, it is hard to say what the prospects are for the development of statistical theory and useful algorithms for such multilevel latent contextual models. Clearly, however, there is room for further advancement.

The existence of latent multilevel models might also create a unified framework for the interpretation of the coefficients of ordinary microlevel models. Suppose an analyst has constructed a model for attitudes toward particular government policies that includes measures of education and income, and that is estimated with OLS regression. Suppose we then consider the interpretive and statistical difficulties raised by the coefficient for income. What does it mean? From a substantive perspective, interpretations of the income effect are often contextual. Income has an effect not only because one's interests are related to the amount of money one has in the bank, but also because income is associated with patterns of interpersonal association, of upbringing, and of exposure to mass media. Ideally, one would like to identify the characteristics of this socio-economic-cultural context, and use its attributes to explain the effects of individual-level variables on outcomes. From this perspective, virtually all sociological models become contextual, with micro and macro variables defining and explaining contexts that provide the explanations for varying parameters at the individual level. Needless to say, identification issues for such latent class models would be quite formidable. The likelihood functions would be complex, and the robustness of results would be uncertain.

In this context, it is helpful to recall the still-germane observations made by Blalock in his 1984 review of contextual analysis. The ultimate goal of multilevel analysis is to determine how social context affects and is affected by social behavior. Improved answers to this question require parallel advances in theory and data collection as well as in statistical methodology. While there is plenty of room for improvement in all of these domains, the rate of progress is encouraging.

ACKNOWLEDGMENTS

We wish to acknowledge the helpful comments of Robert M. Hauser, Thomas F. Janoski, and Allan Parnell on various aspects of this manuscript. This work was supported by National Science Foundation grants SES-9012619 and SES 92-09159, and by a grant from the Alexander Humboldt Foundation.

13See Dickens & Lang (1988) for an application of an endogenous switching regressions model with latent regimes to the problem of employing dual labor markets as a context without an a priori definition of labor market boundaries.
Literature Cited


Entwistle B, Mason WM. 1985. Multilevel ef-
feets of socioeconomic development and family planning programs on children ever born. *Am. J. Sociol.* 91:616–49


Flora P. 1988b. *Growth to Limits: The Western European Welfare States Since World War II.* Vol. II: Germany, United Kingdom, Ireland, Italy. New York: de Gruyter


Lewis C. 1989. Difficulties with Bayesian inferences for random effects. See Bock 1989, pp. 75–86
Mason WM. 1991. Problems in quantitative comparative analysis: Ugly ducklings are to swans as ugly scatter plots are to...? See Huber 1991, pp. 231–43
Parcel TL, Kaufman RL, Jolly L. 1991. Going up the ladder: multiplicity sampling to create
Rubin DB. 1989. Some applications of multilevel models to educational data. See Bock 1989, pp. 1–17