



California Center for Population Research
University of California - Los Angeles

Multilevel Methods of Statistical Analysis

William M. Mason

CCPR-006-01

April 2001

California Center for Population Research
On-Line Working Paper Series

MULTILEVEL METHODS OF STATISTICAL ANALYSIS

By

William M. Mason

Departments of Sociology and Statistics,
University of California-Los Angeles
masonwm@ucla.edu

To Appear In:

Neil J. Smelser and Paul B. Baltes, editors,
International Encyclopedia of the Social and Behavioral Sciences.
Amsterdam, Elsevier Science, 2001.

The assistance of Vincent K. Fu is gratefully acknowledged.

ABSTRACT

MULTILEVEL METHODS OF STATISTICAL ANALYSIS

In hierarchical data structures, observational units at one level are nested within units at other levels. For longitudinal data on individuals, within-person observations are nested within individuals. Within-person (micro) observations are generally more like each other than observations sampled randomly across individuals (contexts). In generalized regression models, several approaches can accommodate this lack of independence: (i) random effects; (ii) fixed effects; (iii) marginal models in the non-Gaussian case, and (iv) regression coefficient covariance matrix adjustment. Each approach has specific advantages and drawbacks. Fixed effects are often used in situations in which a key assumption of the random effects approach is thought to be implausible. In those instances, the use of marginal models or regression coefficient covariance matrix adjustment would also merit reconsideration. Where more than one approach can justifiably be applied—and there are many such cases—estimation results can be method-dependent, leading to different conclusions about the effects of specific covariates. In addition, particularly for the random effects approach, alternative methods of estimation can yield somewhat different results, as can different algorithmic implementations of the same method across software packages. The impact of small within-context sample sizes in unbalanced designs merits further exploration. Where possible, researchers should consider whether their conclusions are method- or algorithm-dependent.

Multilevel Methods of Statistical Analysis

Multilevel, or hierarchical, data are blocked or clustered. Children within families; individuals within organizational units; longitudinal data on individuals; repeated measures on individuals; organizational units within larger organizations; multistage sample surveys; and time series of cross-sections exemplify multilevel or hierarchical data structures. When data are clustered, each observation counts for less than it would if the data points were independent. If clustering is ignored, statistical inference can be faulty. With clustering, regression methods that assume independent observations provide inefficient and possibly biased coefficient estimates, and associated confidence intervals that are too narrow. In the context of regression modeling, any method that attempts to take account of clustering is multilevel.

Several generalized regression methods are commonly used in the analysis of hierarchical data. Hierarchical or random coefficient models are multilevel. So too are fixed effects models, marginal models, classical single-level estimation followed by regression coefficient covariance matrix adjustment, and hybrid approaches. The choice of an approach for a particular research problem must depend on the goals of the research and assessment of the design and measurement deficiencies presented by a particular body of data, and is often consequential. Multilevel modeling approaches emphasize different aspects of the data, vary in the nature and stringency of their assumptions, and support different kinds of descriptions and interpretations. Poor choice of method can lead to the application of an incorrectly specified model, and thereby to good estimates of the wrong quantities.

1. Random Coefficient Models

The random coefficient model offers a compelling formulation that is consistent with the social scientific goal of understanding how units at one level affect, and are affected by, units at other levels. Consider a design with two levels, and suppose that there is a single level-1 regressor (X), a single level-2 regressor (G) that is by definition *constant* over all level-1 observations within a context, and that the dependent variable (Y) is Gaussian distributed conditional on X and G . The random coefficient model can be expressed as

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij} \quad (1.1)$$

with

$$\begin{aligned} \beta_{0j} &= \eta_{00} + \eta_{01}G_j + \alpha_{0j} \\ \beta_{1j} &= \eta_{10} + \eta_{11}G_j + \alpha_{1j} \end{aligned} \quad (1.2)$$

where $j = 1, 2, \dots, J$ denotes contexts (level-2 units), and $i = 1, 2, \dots, n_j$ denotes individuals within contexts (level-1 units). The level-2 equations (Eqn. (1.2)) assert that the level-1 intercepts and slopes vary over contexts as linear functions of G . There is often justification for supposing that the within-context coefficients depend on contextual characteristics. It is appealing that the dependence includes random error components (α_{0j} and α_{1j}).

Assumptions about the error structure of this two-level model are that, for each context, the level-1 errors follow the Gaussian distribution, are centered on zero, have constant variance, and are independent. The error variances may vary over contexts. The level-2 errors are also Gaussian distributed, centered on zero, independent within

equations, and have constant variances. The level-2 errors may be correlated across equations. It is further assumed that the level-1 errors are orthogonal to the level-2 errors, and that all error terms are orthogonal to X and G . The orthogonality assumptions are needed to secure consistent estimates of the η 's.

Many generalizations of this model are possible—coefficients can be constrained to be nonrandom; the dependent variable may be discrete, counted, truncated, ordinal, or time to event; there may be more regressors at any level as well as more than two levels; the nesting of levels may be cross-classified; different assumptions about the functional forms of the distributions of the errors can be made; and the assumption of independence of errors within level-2 (or higher) units can be relaxed. Nevertheless, the assumptions of orthogonality between regressors and errors, and of cross-level orthogonality between error components, continue to remain important. Until that is no longer so, the credibility of an application of a random coefficient model rests on the plausibility of these assumptions, which should be thought of as inextricably linked to the systematic components of the model, and not as the fine print of a contract that is never read, much less enforced.

However compelling it may be on substantive grounds to specify the random coefficient multilevel model as Eqn's. (1.1)–(1.2), substitution of (1.2) into (1.1) results in the equivalent mixed-model representation

$$Y_{ij} = \eta_{00} + \eta_{01}G_j + \eta_{10}X_{ij} + \eta_{11}X_{ij}G_j + (\alpha_{0j} + \alpha_{1j}X_{ij} + \varepsilon_{ij}), \quad (1.3)$$

which demonstrates that treating intercepts and slopes as dependent on G is the same as supposing that Y depends on X , G , and their cross-level (product) interaction.

Random coefficient models with Gaussian errors are commonly estimated by maximum likelihood or restricted maximum likelihood. For binary response the prevalent methods are marginal and penalized quasi-likelihood, although Bayesian methods based on Markov Chain Monte Carlo (MCMC) estimation are also beginning to be used. A consensus on preferred methods of estimation of multilevel binary response and other nonlinear (non-Gaussian) models has yet to emerge. Estimation of multilevel models is iterative and computationally intensive, more so for non-Gaussian cases. And especially for these cases, researchers are well advised to check numerical estimates across algorithms, methods, and software. Other problems, such as substantial design imbalance in combination with context size, need also to be evaluated for their potential impact on numerical estimates. (See *Estimation, Point and Interval; Hierarchical Models: Random and Fixed Effects; Meta Analysis: Overview; Monte Carlo Methods and Bayesian Computation: MCMC, Including Gibbs.*)

2. Fixed Effects Models

The fixed effects model does not require the assumption of orthogonality between X and α_0 inherent to the random coefficient model. It can be written as

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 X_{ij} G_j + \delta_j + u_{ij}, \quad (2.1)$$

where the δ_j are between-context contrasts, and the u_{ij} are assumed to be classical Gaussian errors. Relative to the random coefficient model, the additive term in G has been dropped, and the intercept and slope are no longer allowed to vary randomly over contexts, although each context is allowed its own intercept and the cross-level

interaction is retained. The δ_j absorb the additive component of *all possible* G -variables, not just the finite number that can be included in Eqn. (1.3). Additive components for any level-1 regressors that do not vary within contexts are also absorbed by the δ_j .

In the fixed effects approach, “fixing” the between-context contrasts means that the contexts are themselves assumed to be fixed, in the sense that statistical inference is conditional on the particular selection of contexts present in the data. Because the contrasts can be defined as the coefficients of a set of indicator variables that assign individuals to contexts, there is no need to assume that context membership is orthogonal to X . This is a critical difference between the fixed effect and random coefficient models. If the assumption of orthogonality between X and α_0 is implausible in the random coefficient model, and it can be when the list of G -variables is incomplete, then the fixed effects model provides an alternative that solves the problem.

In the Gaussian case, the fixed effects model is a conventional regression model. When J is large—and often when it is not—there may be little interest in describing estimated values of the δ_j , in which case estimation of the other covariate coefficients can be performed after within-context centering of the variables. For a binary response variable, conditional logistic regression can be used to estimate fixed effect models (Chamberlain 1980). In this case, the context contrasts are not estimated, although additive context differences are controlled. Software for fixed effects estimation is widely available.

If it is desired to obtain estimates of the additive component of the contextual variables, then the fixed effects approach is not the method of choice. Also, in the binary response case, conditional logistic regression does not necessarily use all of the data, and in this sense may be inefficient. (See *Economic Panel Data; Linear Regression: Basics; Multivariate Analysis: Discrete Variables, Logistic Regression.*)

3. Coefficient Variance Estimates Adjusted for Clustering

The adjusted coefficient variance estimator in common use for multilevel analysis is a robust estimator. In actual applications, care must be taken to ensure that the robust estimator selected includes an extension of the algorithm for the general robust estimator that accommodates clustering. With the robust estimator, it is possible to specify a model whose systematic component is identical to that of Eqn. (1.3), and whose error component need not be independently and identically distributed, Gaussian distributed, or orthogonal to the regressors. The robust equivalent of Eqn. (1.3) can be expressed as:

$$Y_{ij} = b_0 + b_1 G_j + b_2 X_{ij} + b_3 X_{ij} G_j + e_{ij}. \quad (3.1)$$

Least squares computation can be used to produce coefficients whose estimated variances and covariances are subsequently adjusted for context membership. The coefficients will be unbiased estimates of the corresponding terms in Eqn. (1.3) if the orthogonality assumption is satisfied. The standard errors will be adjusted for clustering and other sources of heteroscedasticity regardless of whether this is so. When Eqn. (1.3) is true, confidence intervals for its coefficients obtained by prevailing methods will tend to be smaller than confidence intervals based on the robust clustering adjustment associated with Eqn. (3.1).

The robust approach is also an alternative to the fixed effects approach. The latter achieves orthogonality between the regressors and the error component by including context membership in the systematic component of the model, but is unable to include main effects of contextual variables even though it controls for all possible contextual variables. In the robust approach coefficients of contextual regressors are estimated, but all possible contextual variables are not controlled. Coefficient standard errors are adjusted for clustering, but the coefficient estimates are not efficient.

The robust, variance adjustment approach to multilevel analysis has been extended to models of discrete and counted outcomes, and to survival problems, and can be used where a fixed effects approach is not available. Nevertheless, where it can be justified, explicit incorporation of clustering into model specification is preferable to post-hoc adjustments of estimators that ignore clustering. (See *Sample Surveys: The Field*; *Sample Surveys: Survey Design Issues and Strategies*; *Sample Surveys, Model Based Approaches*.)

4. Marginal Models

When Y is a discrete or count response variable, it is possible to specify a multilevel model in which the within-context correlatedness of observations is incorporated as a side condition on the model, rather than as a consequence of a random intercept. For Y binary, and ω_{ij} the logit of “success” for the ij -th observation, the multilevel dependence of Y on regressors can be written as

$$\omega_{ij} = \beta_0 + \beta_1 G_j + \beta_2 X_{ij} + \beta_3 X_{ij} G_j, \quad (4.1)$$

with an additional specification indicating the pattern of correlation between observations within a context ($\text{corr}(Y_{ij}, Y_{i'j}), i \neq i'$). Specifying that all pairs of observations within a context are equally correlated corresponds to the assumption of correlatedness induced by a random coefficient model, but virtually any pattern can be specified.

Models that characterize the marginal distribution of Y as a function of regressors are said to be marginal models. The model of Eqn. (4.1) and its side condition on correlated observations is a member of this class. The generalized linear model (McCullough and Nelder 1989), a unifying theoretical framework for the linear model and discrete and count response variables, has been extended to allow for the analysis of multilevel data (Liang and Zeger 1986; Zeger and Liang 1986). This extension, known as the generalized estimating equations approach (GEE), provides a marginal model (population-average) alternative to random coefficient models when Y is discrete or a count. Equation (4.1) can be estimated using the GEE method.

Gaussian random effects models sustain a marginal model interpretation of regression coefficients. Random effects models for discrete and count response variables do not. Marginal models are appropriate when the analytic goal is to infer how Y depends on covariates for a population of level-1 units averaged over contexts. Random effects models are appropriate when within-context estimates are desired. For the logistic response case, the estimated coefficients of marginal models tend to be smaller in absolute value than the corresponding coefficients in random effects models (Diggle et al. 1994:141).

Marginal models for Y discrete or a count can be estimated for a broad range of structures of potential within-context correlation; random coefficient models for discrete and count response variables currently can not. Computation is fast (iterative generalized least squares), and estimation is relatively straightforward compared to binary response random coefficient models, although issues surrounding optimal characterization of within-context correlation and estimation of coefficient standard errors continue to attract attention.

5. Omitted Variables

Omitted variables in random coefficient models are of particular concern when they are endogenous, that is, correlated with Y and the included regressors, because they induce coefficient bias. The fixed effect approach provides one solution to this problem when the omitted variables are contextual. There are others: measurement of omitted variables, random assignment to contexts, and instrumental variables estimation.

Measuring the omitted variables is not an option in secondary analysis or in the midst of primary analysis—by then the data have been collected and returning to the field is generally infeasible. This should not, however, preclude future attempts to extend the range of measurement.

Random assignment to contexts presumes that an experimental design is feasible in the study of a particular phenomenon. Often it is not. Even when randomization is possible, it is not necessarily well suited to answering certain kinds of questions. (See *Experimental Design: Randomization and Social Experiments*; *Experimental Design: Quasi-Experimental Designs*; *Experimental Design: Large Scale Experimentation*.)

Instrumental variables estimation is often used in an effort to solve an endogeneity problem. When an instrumental variables approach is used, aspects of the clustering problem may be sacrificed to solve what is perceived as a much more important one—selection into contexts. In Eqn. (1.3) replace X by T , suppose that the underlying data structure is quasi-experimental, and let $T_{ij} = 1$ if an individual is treated (0 otherwise). If Y is a health outcome and contexts are hospitals, and if assignment to treatment is not random, then ε potentially contains unmeasured variables correlated with T and also with Y . In quasi- and non-experimental data, whether a patient receives a particular kind of treatment can be nonrandom with some causes known, others not. An instrumental variables approach to this problem attempts to find one or more covariates W correlated with T and Y , where Y and W controlling T are uncorrelated, and W and ε are uncorrelated. It is often difficult to find plausible instruments (W), although contextual variables are sometimes used. Thus, instrumental variables estimation cannot be employed with the automaticity of any of the other approaches to clustered data. (See *Instrumental Variables, in Statistics and Economics*.)

6. Examples

An example of a Gaussian application, and another of a binary response application, illustrate the kinds of differences in findings that can occur using alternative multilevel approaches. The data are from a one percent clustered sample of the 1990 Census of China, and are specific to Hainan province. Clustering is at the enumeration district level, and within clusters enumeration is 100 percent.

6.1 Gaussian Example

In the Gaussian example, there are 17 662 men in the data, distributed over 332 clusters, the smallest having 114 men and the largest having 3735. The occupational socioeconomic status (SEI) of men ages 20–65 is a function of age and years of schooling at the individual level, and of type of place of residence (rural vs. urban) and the square of a contextualized measure of education. Type of place of residence is a characterization of place, and constant within contexts. Substantive theory and exploratory data analysis lead to this random effects model:

$$\begin{aligned}
 SEI &= \beta_{0j} + \beta_{1j}Age + \beta_{2j}Ed + \varepsilon \\
 \beta_{0j} &= \eta_{00} + \eta_{01}Rural + \alpha_{0j} \\
 \beta_{1j} &= \eta_{10} + \eta_{11}CED^2 + \alpha_{1j} \\
 \beta_{2j} &= \eta_{20} + \eta_{21}CED^2 + \alpha_{2j}
 \end{aligned} \tag{6.1}$$

Table 1 reports coefficient estimates and ratios of coefficients to standard errors (“*t*-ratios”) for several approaches. Regressions 1-3 are the same ordinary least squares (OLS) regression, differentiated to keep track of the type of standard error. When clustering is ignored, the *t*-ratios are far larger (Regressions 1 and 2). Unstructured adjustment of standard errors makes little difference. Adjusting OLS standard errors for clustering radically reduces the *t*-ratios (Regression 3). The fixed effects coefficients (Regression 4) are roughly the same as the OLS coefficients, with larger *t*-ratios than obtained with the clustering adjustment. Only one term is “significant” when the clustering adjustment is used, whereas three of the four estimable coefficients are highly “significant” in the fixed effects model. The coefficients of the random effects model (Regression 5) are comparable to those for the fixed effects model (where present), but

the t -ratios are smaller. For those terms present in both the fixed effects and random coefficient models, coefficients are either significant in both models, or not significant in both. In computations not displayed here, the random coefficient specification was computed twice within two different programs, once using maximum likelihood and once using restricted maximum likelihood. There was agreement across programs and methods. Although the basic conclusion is the same for the fixed effects and random coefficient models and differs somewhat from that provided by the clustering adjustment, the variation in coefficient magnitudes and t -ratios across the models is cautionary. Even in the Gaussian case, it can be misleading to estimate a model using only one approach.

6.2 Binary Response Example

In the binary response example, data from the 1990 Hainan sample are again used. Infants and toddlers born in the 18 months prior to the starting date of the Census enumeration are the level-1 units of observation, and enumeration districts are level-2 observations. The available data do not record child mortality, which has been inferred from maternal information elicited in the Census questionnaire. The response variable is whether an imputed death to a child born after 1988 occurred prior to July 1, 1990. There are 43 imputed deaths out of 2572 mother-child pairs. A majority of enumeration districts have no imputed deaths. For this reason, only the within-context intercept is treated as random. For ω the logit of a death within the first 18 months of life, the specification treats the probability of death as a function of maternal years of schooling, the sex of the child, and the interaction between maternal education and sex of child. The

intercept is treated as a function of type of place of residence (the same for all residents of an enumeration district) and a random error:

$$\begin{aligned}\omega_{ij} &= \beta_{0j} + \beta_1 Ed + \beta_2 Male + \beta_3 Ed \cdot Male \\ \beta_{0j} &= \eta_{00} + \eta_{01} Rural + \alpha_{0j}\end{aligned}\tag{6.2}$$

Table 2 presents a selection of regressions based on alternative multilevel approaches and algorithms. Regression 1 is the result of classical maximum likelihood estimation followed by an adjustment for clustering. Using this approach, neither the sex of the child nor the sex-education interaction is “significant.” Using the GEE approach, the coefficient for sex is at the cusp of conventional significance and the sex-education interaction is significant (Regression 2). The fixed effects results, using conditional logistic regression, indicate significance for all included terms (Regression 3). Penalized quasi-likelihood estimates are provided by Regression 4. These are consistent with the fixed effects results for those terms present across the fixed and random effects specifications. Regression 5 presents Bayesian results based on MCMC. The type of place of residence coefficient is smallest and not significant in the Bayesian regression. Application of a clustering adjustment to the GEE results (not shown) also substantially affects conclusions about statistical significance.

It is not unusual to encounter sparse data. This binary response example is perhaps even more cautionary than the Gaussian example. Researchers are not yet routinely able to determine whether convergence is to a local or a global maximum, whether sufficient iterations have been calculated using MCMC, whether all of the assumptions required for a particular method are valid in a given instance, or whether

software is performing as intended.

7. Conclusions

Analysis of multilevel data structures using methods that ignore the clustering of observations in the data is unsatisfactory because it can lead to confidence intervals that are too narrow, and to coefficient bias. Random coefficient models—and their specializations and generalizations—are attractive tools for the analysis of multilevel data. When there is selection into contexts, the random coefficient approach is subject to coefficient bias, as it is more generally when there are omitted regressors at any level. The fixed effects approach provides one kind of solution to the omitted variables problem. Instrumental variable estimation offers another. A posterior adjustment to accommodate clustering, applied to the estimated variances of coefficients, is robust but inefficient when the random effects model is correct. Although this adjustment does not require orthogonality between errors and covariates, robust standard errors for biased coefficients are far from universally desirable. For discrete and count responses, marginal models estimated using GEE provide an alternative to random effects models. Marginal models are attractive when there is little interest in within-context estimates, or when the pattern of within-context correlation is not simple. The choice of a multilevel modeling approach ideally is governed by research goals, the study design, and the nature of the data. There is also value in cross-checking results across methods and software.

8. Further Reading

For an introduction to random coefficient models (terminology often used interchangeably with random effects models and hierarchical models), see Snijders and

Bosker (1999). The level of exposition in Bryk and Raudenbush (1992) ranges from introductory to advanced, as is also true for Goldstein (1995). The exposition in Longford (1993) is technically demanding. Littell et al. (1996) survey mixed models in general, and include material on hierarchical models. Gilks et al. (1996) and Gelman et al. (1995) are useful for Bayesian random coefficient models. Draper (2002) provides a detailed introduction to random coefficient models from the Bayesian perspective.

Fixed effects modeling is treated in econometrics texts (e.g., Greene 2000), and is often discussed as one approach to the analysis of panel data (Baltagi (1995)). For the binary response case, see Pendergast et al. (1996), who provide a review with broad scope. Robust adjustments of standard errors for clustering (e.g., Rogers 1993) are an extension of the methodology of Huber (1967) and White (1980, 1982). The GEE approach is well explicated in Diggle et al. (1994), who provide detailed comparisons with random coefficient and other models. In this same vein, Allison's (1999) introduction should not be missed. Instrumental variables estimation is a standard subject in econometrics texts (e.g., Greene (2000)).

For an illuminating use of random coefficient models in the study of neighborhoods see Sampson et al. (1997). Geronimus and Korenman (1992) use a fixed effects approach in the study of teen pregnancy. Hamilton and Hamilton (1997) use a fixed effects approach in the analysis of surgery outcomes. For the same kind of problem, but using a different data structure, McClellan et al. (1994) employ an instrumental variables approach. Applications of marginal models are largely found in the

epidemiological and biomedical literature. Multilevel analysis has long been mainstream in economics, and will soon become so in the rest of the social sciences.

Software

Dedicated random coefficient packages include HLM5 (Raudenbush et al. 2000) and MLwiN (Goldstein et al. 1998). HLM5 additionally offers GEE estimation; MLwiN includes Bayesian MCMC estimation. Most major statistical packages include facilities for all of the approaches discussed here, as integral components or contributed macros. Software for multilevel analysis is in continuous development. The multilevel models mail list is an essential portal for current information (multilevel@jiscmail.ac.uk), as is the web site hosted by the Multilevel Models Project (<http://www.ioe.ac.uk/multilevel/>).

Bibliography

- Allison P D 1999 *Logistic Regression Using the SAS System*.
- Baltagi B H 1995 *Economic Analysis of Panel Data*. Wiley, New York
- Bryk A S, Raudenbush S W 1992 *Hierarchical Linear Models*. Sage, Newbury Park, CA
- Chamberlain G 1980 Analysis of covariance with qualitative data. *Rev. Economic Studies* 47:225-38
- Diggle P J, Liang K-Y, Zeger S L 1994 *Analysis of Longitudinal Data*. Oxford University
- Draper D 2002 *Bayesian Hierarchical Modeling*. Springer-Verlag, New York
- Gelman A, Carlin J B, Stern H S, Rubin D B 1995 *Bayesian Data Analysis*. Chapman and Hall, London
- Geronimus A T, Korenman S 1992 The socioeconomic consequences of teen childbearing reconsidered. *Q. J. Econ.* 107:1187-1214
- Gilks W R, Richardson S, Spiegelhalter D J 1996 *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London
- Goldstein H 1995 *Multilevel Statistical Models*. Wiley, New York
- Goldstein H, Rasbash J, Plewis I, Draper D, Browne W, Yang M, Woodhouse G, Healy M 1998 *A User's Guide to MLwiN*. Multilevel Models Project, Institute of Education, University of London, London
- Greene W H 2000 *Econometric Analysis*. 4th ed. Prentice-Hall, Upper Saddle River, NJ
- Hamilton B H, Hamilton V H 1997 Estimating surgical volume—outcome relationships applying survival models: accounting for frailty and hospital fixed effects. *Econometrics and Health Econ.* 6:383-95
- Huber P J 1967 The behavior of maximum likelihood estimators under nonstandard conditions. In: LeCam L M, Neyman J (eds.) *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley, CA 1, 221-33
- Liang, K-Y, Zeger S L 1986 Longitudinal data analysis using generalized linear models. *Biometrika* 73:13-22

Littell R C, Milliken G A, Stroup W W, Wolfinger R D 1996 *SAS System for Mixed Models*. SAS Institute, Cary, NC

Longford N T 1993 *Random Coefficient Models*. Oxford University Press, New York

McClellan M, McNeil B J, Newhouse J P 1994 Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? *J Am. Med. Assoc.* 272:859-66

McCullough P, Nelder J A 1989 *Generalized Linear Models*. Chapman and Hall, New York

Pendergast J F, Gange S J, Newton M A, Lindstrom M J, Palta M, Fisher M 1996 A survey of methods for analyzing clustered binary response data. *Int. Statist. Rev.* 64:89-118

Raudenbush S, Bryk A, Cheong Y F, Congdon R 2000 *HLM 5: Hierarchical Linear and Nonlinear Modeling*. Scientific Software International, Lincolnwood, IL

Rogers W H 1993 Regression standard errors in clustered samples. *Stata Technical Bulletin* 13:19-23

Sampson, R J, Raudenbush, S W, Earls, F 1997 Neighborhoods and violent crime: a multilevel study of collective efficacy. *Science* 277:918-24

Snijders T, Bosker R 1999 *Multilevel Analysis*. Sage Publications, London

White H 1980 A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48:817-30

White H 1982 Maximum likelihood estimation of misspecified models. *Econometrica* 50:1-25

Zeger, S L, Liang, K-Y 1986 Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 42:121-30

Table 1
Alternative Multilevel Estimates, Occupational Socioeconomic Status of Men

Covariates	Regressions		
	Ordinary Least Squares Unadjusted/Robust/Cluster (1)/(2)/(3)	Fixed Effect (4)	Random Coefficient (5)
Age	0.09 (7.78)/(7.92)/(1.32)	-0.01 (-0.91)	0.02 (0.50)
Education	0.34 (7.29)/(6.00)/(1.00)	0.36 (6.96)	0.43 (2.36)
Rural	-4.92 (-18.47)/(-13.50)/(-1.01)		-0.72 (-0.35)
Age•CED ²	0.31 (13.16)/(13.03)/(1.72)	0.66 (17.83)	0.55 (5.75)
Ed•CED ²	2.67 (25.51)/(21.78)/(3.06)	2.75 (20.81)	2.38 (4.72)
Intercept	15.50 (33.08)/(32.15)/(4.63)	11.48 (29.42)	12.89 (6.71)
R ²	0.45	0.57	

Source: One percent sample, males ages 20-65, Hainan Province, 1990 Census of China.

Note: $N = 17\ 662$. Quantities in parentheses are ratios of coefficients to standard errors. The coefficients of Regressions (1)-(3), estimated by ordinary least squares, are identical. The standard errors of (1) are unadjusted; those of (2) are adjusted, but without taking explicit account of clustering; those of (3) are adjusted for cluster membership. The context contrasts in the fixed effects regression (4) are not shown. The level-2 error variances in the random coefficient regression (5), not shown, are significant.

Table 2
Binary Response Example, Imputed Infant Mortality

Variables	Regression				
	Classical	GEE	Fixed	Random Coefficients	
	Cluster Adj.			Effects	PQL
(1)	(2)	(3)	(4)	(5)	
Maternal Ed	-0.25 (-2.62)	-0.284 (-2.93)	-0.27 (-2.77)	-0.29 (-3.19)	-0.32 (-3.56)
Male	-0.96 (-1.26)	-1.18 (-1.95)	-1.25 (-2.13)	-1.20 (-2.07)	-1.43 (-2.55)
Ed•Male	0.24 (1.73)	0.29 (2.48)	0.30 (2.63)	0.29 (2.60)	0.32 (2.84)
Rural	2.07 (2.17)	1.99 (1.48)		1.81 (2.03)	1.38 (1.38)
Intercept	-4.74 (-4.71)	-4.74 (-3.50)		-4.75 (-5.07)	-4.74 (-4.24)
Level-2 Var.				0.85 (1.78)	2.05 (1.53)

Source: One percent sample, infants to age 18 months paired with mothers, Hainan Province, 1990 Census of China.

Note: $N = 2572$. Quantities in parentheses are ratios of coefficients to standard errors.