CONTRIBUTORS' PROOFCHECKING INSTRUCTIONS FOR – ENCYCLOPEDIA OF SOCIAL MEASUREMENT

PROOFREADING

The text content and layout of your article is not in final form when you receive proofs. Read proofs for accuracy and clarity, as well as for typographical errors, but DO NOT REWRITE.

Check titles and headings for spelling and capitalization. Ensure that the right typeface and size have been used to indicate the proper level of each heading. Review numbered items such as figures, tables, and lists for proper order. Proofread the captions and credit lines for illustrations and tables. Ensure that any permissioned material has the required credit line. Artwork appears in its final size and placement, but may not match the image quality of the final printed publication and is for content checking only.

Any questions from the copy editor will be listed following the article proof. Please address these questions as necessary. While it is appreciated that some articles may require updating/revising, please try to keep alterations to a minimum. Excessive alterations may be charged to the contributors.

Please keep a copy of any corrections you make.

RETURNING PROOFS

Proofs should be returned to Aimee Bishop at Elsevier by **May 26** by one of the following methods:

1. If corrections are minor, they should be sent in an e-mail to **SOME_proofs@elsevier.com.** This e-mail should state the encyclopedia title and article title and ID number. Each correction listed should clearly identify the page number, paragraph number, line number within that paragraph, and the correction itself.

2. If corrections are more substantial, the amended hard copy should be mailed by express courier or faxed (only if marked very clearly) to:

Aimee Bishop
Production Project Manager
Elsevier
The Boulevard
Langford Lane
Kidlington
Oxford OX 5 1GB
UK

Fax: **+44 (0) 1865 843974**

3. If the proofs are OK as is, please inform Aimee Bishop via either email (**SOME_proofs@elsevier.com**) or fax (**+44 (0) 1865 843974**); be sure to list the article ID number and title.

Please note that your corrections should be returned as quickly as possible. Scheduling is crucial at this stage of production, and a delay in the return of the proofs could mean missing the press date, resulting in delayed publication.

CHECKLIST

Author name(s) and affiliation(s) checked? ☐
Figures and tables checked? ☐
Further Reading section checked and completed? ☐
Outstanding permissions letters returned to Elsevier? ☐
Author's mailing address checked and updated? ☐
[Note: this information will be used only to update our records; this section will not be printed in the final article.]
Manuscript queries addressed/answered? ☐

If you have any questions regarding your proofs, contact Aimee Bishop at **SOME_proofs@elsevier.com.**

a0005 # Hierarchical Linear Models

Jeff Gill
University of Florida, Gainesville, Florida, USA

Glossary

g0005 **analysis of covariance model (ANCOVA)** A varying intercept hierarchical linear model with the second-level effect fixed across groups.

g0010 **between-unit model** The component of a hierarchical linear model that describes the variability across the groups.

g0015 **context-level variables** Variables defined at the second or higher level of the hierarchical linear model.

g0020 **empirical Bayes** Using the observed data to estimate terminal-level hierarchical model parameters.

g0025 **exchangeability** The property of a hierarchical linear model that the joint probability distribution is not changed by re-ordering the data values.

g0030 **expectation-maximization (EM) algorithm** An iterative procedure for computing modal quantities when the data are incomplete.

g0035 **fixed effects coefficients** Model coefficients that are assumed to pertain to the entire population and therefore do not need to be distinguished by subgroups.

g0040 **hierarchy** The structure of data that identifies units and subunits in the form of nesting.

g0045 **interaction term** A model specification term that applies to some mathematical composite of explanatory variables, usually a product.

g0050 **random coefficients regression model** A hierarchical linear model in which the only specified effect from the second level is seen through error terms.

g0055 **random effects coefficients** Model coefficients that are specified to differ by subgroups and are treated probabilistically at the next highest level of the model.

g0060 **two-level model** A hierarchical linear model that specifies a group level and a single contextual level.

g0065 **varying intercept model** A hierarchical linear model with only one (noninteractive) effect from the second level of the model.

g0070 **within-unit model** The component of a hierarchical linear model that describes variability confined to individual groups.

Hierarchical linear models (HLMs) are statistical specifications that explicitly recognize multiple levels in data. Because explanatory variables can be measured at different points of aggregation, it is often important to structure inferences that specifically identify multilevel relationships. In the classic example, student achievement can be measured at multiple levels: individually, by class, by school, by district, by state, or nationally. This is not just an issue of clarity and organization. If there exist differing effects by level, then the substantive interpretation of the coefficients will be wrong if levels are ignored. HLMs take the standard linear model specification and remove the restriction that the estimated coefficients be constant across individual cases by specifying levels of additional effects to be estimated. This approach is also called random effects modeling because the regression coefficients are now presumed to be random quantities according to additionally specified distributions. p0005

Essential Description of Hierarchical Linear Models s0005

The development hierarchical linear model (HLM) starts with a simple bivariate linear regression specification for individual i: p0010

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (1)$$

which relates the outcome variable to the systematic component and the error term. The standard conditions for this model include the Gauss-Markov assumptions (linear functional form, independent errors with mean zero and constant variance, and no relationship between

regressor and errors), and normality of the errors (provided reasonable sample size): $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.

p0015 Suppose, for example, that we are interested in measuring university student evaluations of faculty teaching through the standard end-of-semester survey. Then the outcome variable, Y_i, is considered the mean score for instructor i for a given class, recorded along with the explanatory variable, X_i, indicating years of teaching experience. In this example setup, the intercept, β_0, is the expected score for a new instructor. Now, consider that this analysis is temporarily taking place only in department j. This means that Eq. (1) becomes:

$$Y_{ij} = \beta_{j0} + \beta_{j1} X_{ij} + \varepsilon_{ij} \tag{2}$$

There is no substantive change here; for the moment, the j coefficient is just a placeholder to remind us that we are studying only the jth department so far.

p0020 Now consider broadening the analysis to evaluate student evaluations of teaching across the entire university by looking at multiple departments. Although we expect some differences, it would be rare to find that there was no underlying commonality among the instructors. A more realistic idea is that, although each instructor has idiosyncratic characteristics, because he or she is teaching at the same university at the same point in time and being evaluated by the same student body there is a common distribution from which β_0 and β_1 are drawn. Actually, it would be unfair and inaccurate to calculate these means across all undergraduate majors at the university. It is well known, for instance, that sociology departments enjoy higher mean student evaluations than chemistry departments. So, now add a second level to the model that explicitly nests instructors within departments and index these departments by $j = 1$ to J:

$$\begin{aligned} \beta_{j0} &= \gamma_{00} + \gamma_{10} Z_{j0} + u_{j0} \\ \beta_{j1} &= \gamma_{01} + \gamma_{11} Z_{j1} + u_{j1} \end{aligned} \tag{3}$$

where all individual level variation is assigned to departments producing department-level residuals: u_{j0} and u_{j1}. The variables at this level are called context-level variables, and contextual specificity is the existence of legitimately comparable groups. Here, the example explanatory variables Z_{j0} and Z_{j1} are the average class size for department j and the average annual research output per faculty member in department j, respectively. Note that if we were interested in performing two nonhierarchical department-level analyses, this would be straightforward using these two equations, provided that the data exist. Of course, our interest here is not in developing separate single-level models; the two-level model is produced by inserting the department-level specifications, Eq. (3), into the original expression for instructor evaluations, Eq. (2). Performing this substitution and rearranging produces:

$$\begin{aligned} Y_{ij} &= \left(\gamma_{00} + \gamma_{10} Z_{j0} + u_{j0}\right) + \left(\gamma_{01} + \gamma_{11} Z_{j1} + u_{j1}\right) X_{ij} + \varepsilon_{ij} \\ &= \gamma_{00} + \gamma_{01} X_{ij} + \gamma_{10} Z_{j0} + \gamma_{11} X_{ij} Z_{j1} + u_{j1} X_{ij} + u_{j0} + \varepsilon_{ij} \end{aligned} \tag{4}$$

This equation also shows that the composite error structure, $u_{j1} X_{ij} + u_{j0} + \varepsilon_{ij}$, is now clearly heteroscedastic because it is conditioned on levels of the explanatory variable. Unless the corresponding variance-covariance matrix is known, and therefore incorporated as weights in the general linear model, it must also be estimated. Ignoring this effect and calculating with ordinary least squares (OLS) produces consistent estimators but incorrect standard errors because it is equivalent to assuming zero intraclass correlation.

Often the first-level expression describing the performance of instructor i in a given department, as specified by Eq. (2), is labeled the within-unit model because its effects are confined to the single department; the second-level expressions describing the performance of department j as a whole, as specified by Eq. (3), are labeled the between-unit model because they describe the variability across the departments. Looking closely at Eq. (4) reveals that there are three distinct implications of the effects of the coefficients for the explanatory variables:

- γ_{01} gives the slope coefficient for a one-unit effect of teacher experience in department j. This slope varies by department.
- γ_{10} gives the slope coefficient for a one-unit change in department class size in department j, completely independent of individual teacher effects in that department.
- γ_{11} gives the slope coefficient for the product of individual teacher experience by department and mean annual research output by department.

Because this set of variables contains both fixed and random effects, Eq. (4) is called a mixed model.

The fundamental characteristic of the multilevel data discussed here is that some variables are measured at an individual level and others are measured at differing levels of aggregation. This drives the need for a model such as Eq. (4) that classify variables and coefficients by the level of hierarchy they affect. Interestingly, a large proportion of HLMs in published work come from education policy studies. This is due to natural nesting of education data through the bureaucratic structure of these institutions. Other applications include studies of voting, bureaucracy, medical trials, and crime rates.

Special Cases of the Hierarchical Linear Model

p0035 There are several interesting ramifications that come from fixing various quantities in the basic HLM. The most basic is produced by setting the full β_{j1} component and the γ_{10} term in Eq. (4) equal to zero. The result is the standard ANOVA model with random effects:

$$Y_{ij} = \gamma_{00} + u_{j0} + \varepsilon_{ij} \tag{5}$$

In another basic case, if the second-level model defines a fixed-effect rather than random-effect model, (u_{j1}, $u_{j0} = \mathbf{0}$), then the resulting specification is just simple linear regression model with an interaction term between the instructor level explanatory variable and the department level explanatory variable:

$$Y_{ij} = \gamma_{00} + \gamma_{01}X_{ij} + \gamma_{10}Z_{j0} + \gamma_{11}X_{ij}Z_{j1} + \varepsilon_{ij} \tag{6}$$

This is one of the most studied enhancements of the basic linear form in the social sciences.

p0040 Another very basic model comes from assuming that the second-level model introduces no new error terms and there is also no interaction effect. Specifically, this means that we can treat the intercept term as a composite of a constant across the sample and a constant across only the j groupings:

$$Y_{ij} = \left(\gamma_{00} + \gamma_{10}Z_{j0}\right) + \gamma_{01}X_{ij} + \varepsilon_{ij} \tag{7}$$

This is routinely called a varying intercept model because the parenthetical expression is now a group-specific intercept term. If we add the second assumption that there is no articulated structure within the first term, that is, $(\gamma_{00} + \gamma_{10}Z_{j0})$ is equal to a single context-specific α_j, this is now the analysis of covariance model (ANCOVA).

p0045 Sometimes it is possible to take some specific parameter in the model and fix it at a known level. Thus, if substantive information at hand indicates that there is no variability to one of the γ terms, it is appropriate to fix it in the model. It is also possible design a combination strategy such as to fix the slope coefficient ($\beta_{j1} = \gamma_{10} + \gamma_{11}Z_{j1}$) and let the intercept coefficient remain a random effect, or to fix the intercept coefficient ($\beta_{j0} = \gamma_{00} + \gamma_{10}Z_{j0}$) and let the slope remain a random effect.

p0050 Another common variation is to assume that $Z_{j0} = 0$ and $Z_{j1} = 0$, but retain the u_{j0} error term:

$$Y_{ij} = \gamma_{00} + \gamma_{01}X_{ij} + u_{j1}X_{ij} + u_{j0} + \varepsilon_{ij} \tag{8}$$

This model asserts that the j categorization is not important for determining the expected effect on Y_{ij}, but that there is an additional source of error from the categories. Hence, specifying the model with only one source of error is to miss a heteroscedastic effect. A specification of this type is typically called a random coefficients regression model.

p0055 Another related variation is to assume that the effect of the within-unit explanatory variable (years of teaching in our example) is uniform across departments. This is equivalent to setting $\gamma_{11}X_{ij}Z_{j1} = 0$, producing:

$$Y_{ij} = \gamma_{00} + \gamma_{01}X_{ij} + \gamma_{10}Z_{j0} + u_{j1}X_{ij} + u_{j0} + \varepsilon_{ij} \tag{9}$$

where sometimes $u_{j1}X_{ij}$ is also set to zero. The common name for this specification is the additive variance components model.

The General Structure of the Hierarchical Linear Model

s0015

p0060 The previously developed model is actually a substantial simplification in that typical models in social science research contain many more explanatory variables at both the within-unit level and the between-unit levels. It is therefore necessary to generalize HLM to incorporate more specification flexibility. First we recast Eq. (4) in matrix terms, such that the dimensional assumptions will be generalized to accommodate more useful specifications. We define a new $\boldsymbol{\beta}_j$ vector according to:

$$\boldsymbol{\beta}_j = \begin{bmatrix} \beta_{j0} \\ \beta_{j1} \end{bmatrix} = \begin{bmatrix} 1 & Z_{j0} & 0 & 0 \\ 0 & 0 & 1 & Z_{j1} \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \\ \gamma_{01} \\ \gamma_{11} \end{bmatrix} + \begin{bmatrix} u_{j0} \\ u_{j1} \end{bmatrix} \tag{10}$$

which is just the vectorized version of Eq. (3). Therefore, it is possible to express Eq. (4) in the very concise form:

$$Y_{ij} = \boldsymbol{\beta}_j' \begin{bmatrix} 1 & X_{ij} \end{bmatrix}' + \varepsilon_{ij} \tag{11}$$

This extra formalism is really not worth the effort for a model of this size; its real utility is demonstrated when there are more explanatory variables at the contextual level. Define k_0 and k_1 to be the number of explanatory variables defined at the second level for β_{j0} and β_{j1}, respectively. Thus far we have had the

restrictions: $k_0 = 2$ and $k_1 = 2$; but we can now generalize this dimension:

$$\boldsymbol{\beta}_j = \begin{bmatrix} \beta_{j0} \\ \beta_{j1} \end{bmatrix} = \begin{bmatrix} 1 & Z_{j01} & Z_{j02} & \cdots & Z_{j0(k_0-1)} & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & Z_{j11} & Z_{j12} & \cdots & Z_{j1(k_1-1)} \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \\ \vdots \\ \gamma_{(k_0-1)0} \\ \gamma_{01} \\ \gamma_{11} \\ \vdots \\ \gamma_{(k_1-1)1} \end{bmatrix} + \begin{bmatrix} u_{j0} \\ u_{j1} \end{bmatrix} \tag{12}$$

The dimension of the matrix of Z variables is now $(2 \times k_0 + k_1)$ and the length of the $\boldsymbol{\gamma}$ vector is $k_0 + k_1$, for any specified values of k_i that are allowed to differ. It is common, and computationally convenient, to assume that the error vector, u_j, is multivariate normally distributed around zero with a given or estimated variance-covariance matrix. Note that the row specifications in the $\mathbf{Z}$ matrix always begin with a 1 for the constant, which specifies a level-one constant in the first row and a level-one restricted explanatory variable in the second row.

p0065 It is important to observe that because the constant in this model is part of the specification, the indices run to $k_0 - 1$ and $k_1 - 1$ to obtain the dimensions k_0 and k_1. Also, when there was only one Z variable specified in the second level of the model, it was sufficient to index simply by the subscripts j and either 0 or 1, as in the first and second equations of (3). However, now that there are an arbitrary number for each second-level equation they must be further indexed by the third value—here, 1 to $k_0 - 1$ or 1 to $k_1 - 1$. Note that each group is no longer required to contain the same mix of second-level explanatory variables. This turns out to be useful in specifying many varying model specifications.

p0070 It is possible that there are also more first-level variables in the model (it is likely, in fact). To accommodate this, we must further generalize the defined matrix structures. Define the $\mathbf{Z}_\ell$ vector as

$$\mathbf{Z}_\ell = \begin{bmatrix} 1 & Z_{j\ell 1} & Z_{j\ell 2} & \cdots & Z_{j\ell(k_\ell - 1)} \end{bmatrix} \tag{13}$$

for $\ell = 1$ to L coefficients in the first-level model, including the constant. Therefore the $\mathbf{Z}$ matrix is now a $(L \times L)$ diagonal matrix according to:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & 0 & 0 & \ldots & 0 \\ 0 & \mathbf{Z}_2 & 0 & \ldots & 0 \\ 0 & 0 & \mathbf{Z}_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \mathbf{Z}_L \end{bmatrix} \tag{14}$$

where each diagonal value is a $\mathbf{Z}_\ell$ vector. This can also be fully written out as an irregular ℓ-diagonalized matrix, but it would be more cumbersome than the given form. Given the new for of the $\mathbf{Z}$ matrix, it is necessary to respecify the g vector as follows:

$$\boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\gamma}_0 \\ \boldsymbol{\gamma}_1 \\ \vdots \\ \boldsymbol{\gamma}_L \end{bmatrix} \tag{15}$$

where each $\boldsymbol{\gamma}_\ell$ is a column vector whose length is determined by the k_ℓ dimension specification. Putting these new structures together gives:

$$\boldsymbol{\beta}_j = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u} \tag{16}$$

and:

$$Y_{ij} = \boldsymbol{\beta}_j' \begin{bmatrix} 1 & X_{ij1} & X_{ij2} & \ldots & X_{ijL} \end{bmatrix}' + \varepsilon_{ij} \tag{17}$$

Thus, the HLM in this form allows any number of first- and second-level explanatory variables, as well as differing combinations across contextual levels. Note also that there is no restriction that number of individual units, n_j, be equal across the contexts (although this can make the estimation process more involved).

The final basic way that the HLM can be made more p0075 general is to add further levels of hierarchy with respect to levels. That is, it is possible to specify a third level in exactly the way that the second level was added by parameterizing the $\boldsymbol{\gamma}$ terms according to:

$$\gamma_{pq} = \delta_{0q} + \delta_{1q} W_{pq} + \upsilon_{pq}$$

where the p subscript indicates a second level of contexts ($p = 1, \ldots, P$), and the q subscript indexes the number of equations ($q = 1, \ldots, Q$) specified at this level (analogous to k at the lower level). In this specification, W_{pq} is a third-level measured explanatory variable and υ_{pq} is the level-associated error term.

Obviously, the specifications can be made very complex at this level.

p0080 The other principle approach to modeling a third level is to specify a Bayesian prior distribution for the $\boldsymbol{\gamma}$ coefficients. These priors are typically assigned normal distributions, but this is not a restriction and many others have been used. As a consequence of linearity, the normal property then ripples down the hierarchy, making estimation relatively easy. This model, then, specifies hierarchies of linear hyperpriors, each of which has its own prior plus an associated matrix of explanatory variables, and only nodes at the highest level of the hierarchy have fixed hyperprior values.

s0020 Estimation of the Hierarchical Linear Model

p0085 No matter how complex the right-hand side of the HLM equation becomes, the left-hand side always consists of Y_{ij}, which is assumed to be normal, with mean equal to the systematic component of the model and variance from the collected error terms. If it were known that the error structure in the model was uncorrelated to explanatory variables, then it would easy to estimate the coefficients with standard maximum likelihood or least squares approaches. Actually we know that, in general, the form of the errors *is* conditional on the levels of the explanatory variables because in Eq. (4) there is the term, $u_{j1}X_{ij}$. In addition, there are increasing numbers of dependencies as the model becomes progressively more complex and realistic.

p0090 If we knew for certain the form of the relationship between the regressors and errors, then it could be expressed through a weighting matrix and general least squares would provide consistent estimates of the coefficients and their standard errors. Unfortunately, this information is rarely available. The classic alternative is to specify a likelihood function and employ a maximum likelihood estimation of the full set of unknown parameters, including variances using Fisher scoring. This is often a cumbersome process, so many software implementations work with the profile likelihood—first estimating the higher order variance terms and only then fixing them in the likelihood function equation for the lower-level parameters. This tends to underestimate the magnitude of the higher-order variance terms because uncertainty is ignored in the first step, leading to overconfident model results. An improved process is to employ restricted maximum likelihood (REML) by integrating out the fixed-effects terms in the calculation of the profile likelihood and, after obtaining the lower-level parameter estimates, recalculating the higher-order variance terms conditional on these. However, the best method is the quasi-Bayesian procedure, empirical Bayes/maximum likelihood (EB/ML). A fundamental principle of Bayesianism is that unknown parameters are treated as random variables possessing their own distributions which can be estimated as a consequence of applying Bayes's law. By analogy, we can consider the unknown HLM estimates as having their own distributions, conditioned on unknown quantities from the higher level of the model. Rather than stipulating explicit priors for the parameters, as a Bayesian would do, it is possible to use a prior suggested by the data, called empirical Bayes.

p0095 The expectation-maximization (EM) algorithm is essential to this estimation process and therefore warrants some description. EM is a flexible and often-used method for incomplete data problems; it is used to fill in missing information, given a specified model. The notion of what is "missing" is general here; it can be unknown parameters, missing data, or both. There are two basic steps. First, we assign temporary data that represent a reasonable guess to the missing data (expectation). Second, we proceed with maximum likelihood estimation of the parameters as if there now existed a complete-data problem (maximization). The algorithm is iterative in the sense that it is now possible to use these parameter estimates to update the assignment of the temporary data values with better guesses, and repeat the process. It can be shown that the EM algorithm gives a series of parameter estimates that are monotonically increasing on the likelihood metric and are guaranteed to converge to a unique maximum point under very general and nonrestrictive regularity conditions. The utility here is that the HLM with linear specifications and normal assumptions is a particularly well-behaved application of EM.

p0100 Detailed summaries of the EB/ML computational procedure for obtaining coefficient estimates and measures of reliability can be found elsewhere. The basic strategy is to obtain estimates of the variance terms using the EM algorithm and the joint likelihood function for the coefficients and the variances, plug these estimates into the top hierarchy of the model, perform maximum likelihood calculations as if these were the correct weightings, and update the estimate of the coefficients by using the mean of the subsequent posterior. This is a very general description of the procedure; there are many nuances that depend on the particular form of the model and configuration of the data.

Critical Advantages of the Hierarchical Linear Model s0025

p0105 HLMs are a compromise between two opposite approaches to clustering. On one side, it is possible to simply pool all the observations and calculate an estimate

of the coefficients of interest as if the between-group effects did not matter. Conversely, it is also possible to aggregate the data by the groups and calculate the coefficient estimates on these aggregations as if they were the primary object of interest. The first approach ignores between-group variation and the second approach ignores within-group variation. It is possible that either of these approaches is entirely appropriate and reliable inferences can be obtained. Of course, if there actually are important differences by groupings, then neither will be correct. HLMs provide a method for producing models that explicitly recognize this distinction by incorporating the nesting of the data into the model specification.

p0110 HLMs also have several specific methodological advantages over standard linear models:

- Hierarchical models are ideal tools for identifying and measuring structural relationships that fall at different levels of the data generating procedure.
- Hierarchical models have virtually no limit to the dimension of their hierarchy.
- Hierarchical models directly express the exchangeability of units.
- Nonhierarchical models applied to multilevel data typically underestimate the variance.
- Hierarchical models facilitate the testing of hypotheses across different levels of analysis.
- Nonhierarchical models can be nested within hierarchical models, allowing a likelihood or Bayes factor test of the validity of the proposed hierarchical structure.

p0115 Although these reasons are compelling, it is only relatively recently that hierarchical models have been actively pursued in the social sciences. This is parallel (and related) to the attachment social scientists have for the linear model in general. What precipitated the change was the dramatic improvement in statistical computing that provided solutions to previously intractable problems. These stochastic simulation tools include the EM algorithm; Markov chain Monte Carlo techniques (MCMC), such as the Metropolis-Hastings algorithm; and the Gibbs sampler, whereby an iterative chain of consecutive computationally generated values is set up carefully enough and run long enough to produce *empirical* estimates of integral quantities of interest from later chain values. Although these approaches are typically associated with Bayesian modeling, such iteration techniques are not *limited* to Bayesian or even hierarchical applications. They do, however, greatly help naturally occurring computational problems in these settings.

See Also the Following Articles

Further Reading

Goldstein, H. (1995). "Multilevel Statistical Models." Edward Arnold, New York.

Heck, R. H., and Thomas, S. L. (2000). "Introduction to Multilevel Modeling Techniques." Lawrence Erlbaum Associates, Mahwah, NJ.

Kreft, I., and de Leeuw, J. (1998). "Introducing Multilevel Modeling." Sage Publications, Newbury Park, CA.

Leyland, A. H., and Goldstein, H. (2001). "Multilevel MQ1 Modelling of Health Statistics." John Wiley & Sons, MQ2 New York.

Lindley, D. V., and Smith, A. F. M. (1972). Bayes estimates for the linear model. *J. Royal Statist. Soc. B* **34,** 1–41.

NeIder, J. A. (1977). A reformulation of linear models (with discussion). *J. Royal Statist. Soc. A* **140,** 48–76.

Raudenbush, S., and Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociol. Educ.* **59,** 1–17.

Raudenbush, S., and Bryk, A. S. (2002). "Hierarchical Linear Models," 2nd Ed. Sage Publications, Newbury Park, CA.

Reise, S. P., and Duan, N. (2001). "Multilevel Models: A Special Issue of Multivariate Behavioral Research." Lawrence Erlbaum Associates, Mahwah, NJ.

Smith, A. F. M. (1973). A general Bayesian linear model. *J. Royal Statist. Soc. B* **35,** 61–75.

Wong, G. Y., and Mason, W. M. (1991). Contextually specific effects and other generalizations of the hierarchical linear model for comparative analysis. *J. Am. Statist. Assoc.* **86,** 487–503.

Manuscript queries and/or remarks - please respond to these points on returning your proofs

(MQ1) OK as edited?
(MQ2) "Modelling" in original? Modeling?