Kelvyn Jones, Myles I. Gould and Robert Watt

Multiple Contexts as Cross-classified Models: The Labor Vote in the British General Election of 1992

Voters make their decisions in social and geographical contexts that can be seen as different levels in an overall data structure. Increasingly these structures are being analyzed by multilevel models, but this approach has so far been limited to structures that are strictly hierarchical. This paper outlines the approach of cross-classified multilevel models in which units at lower levels in the structure can be nested in more than one higher-level unit simultaneously. An appropriate modeling framework is outlined, models are specified, and particular attention is paid to efficient computation. The approach is illustrated through a crossclassified logit analysis of Labor versus Conservative support for a nationally representative sample of voting behavior for the 1992 British General Election. The data is structured so that individual voters at level 1 are nested within constituencies at level 2 which are cross-classified by geographical and functional regionalizations at level 3. A conclusion discusses the general utility of a crossclassified approach to geographically based contextual research, while two technical appendices provide details on model estimation.

INTRODUCTION

Research on voting behavior has long stressed the importance of context (Ennis 1962). Voters may be socialized in their youth in a particular place and subsequently move to another locality where these views are challenged or reinforced. At any one time, a particular voter may receive contextual influences from a number of sources, be it the home, the neighborhood, or the work-place. Different scales of context can readily be recognized with individual voters being influenced by the microcontexts of the home, the mesocontexts of the

The authors thank Harvey Goldstein and Jon Rasbash for allowing them to use the MLn software before its general release. The BES data were obtained through the ESRC Data Archive at the University of Essex. This research is in part supported by ESRC grant H519255005 under the Analysis of Large and Complex Datasets program. Thanks to Sundra Winterbotham for typing the appendices and Bill Johnson for drawing the diagrams.

Kelvyn Jones is professor of geography, Myles Gould is lecturer in health services research, and Robert Watt is a research assistant at the University of Portsmouth.

Geographical Analysis, Vol. 30, No. 1 (January 1998) © 1998 Ohio State University Press Submitted: 10/3/96. Revised version accepted: 6/6/97



FIG. 1. A Strictly Hierarchical Multilevel Structure

neighborhood and the macrocontexts of the constituency, the region, and beyond.

One way of conceiving of these contexts is as a set of "levels" in a data structure. A simple example (Figure 1) is voters at level 1 nested in neighborhoods at level 2 within regions at level 3. Clearly if we ignore any level in an analysis we can say nothing substantive about it, but this is what researchers have previously done. They have *either* worked at the individual scale and used social survey data, or they have analyzed aggregate data at some higher level. The former approach commits the atomistic fallacy (Alker 1969) in omitting context altogether, while the latter approach commits the aggregative fallacy (Robinson 1950) in forgetting that it is individuals who vote, not groups. Moreover, if the levels are of substantive importance, people within each level or context will tend to be autocorrelated. This results in a whole series of technical problems if the hierarchically structured nature of the population with many levels is ignored (Skinner, Holt, and Smith 1989).

The last few years, however, have seen a growing use of a multilevel approach that allows the simultaneous consideration of individual and ecological data, that is, of individuals in particular contexts (Goldstein 1987; Jones 1991). Examples of this approach to voting behavior are given by Jones, Johnston, and Pattie (1992), who examine the voting behavior of individuals nested within constituencies within regions; Jones, Tonkin, and Wrigley (1998), who analyzed the effects of interactions between individual and constituency characteristics; and McMahon and Heath (1992), who use multilevel models to examine the changing class vote in a set of repeated cross-sectional studies of voters in constituencies. Each of these studies, however, has been restricted to context as strict hierarchy. Only models in which each lower-level unit belongs to one (and only one) unit at the next higher level could be estimated. This ignores the reality of multiple overlapping contexts.

The present research is motivated by debate concerning the relative importance of differing overlapping contexts in accounting for voting intentions. This debate revolves around two issues: geographical versus functional contexts, and contextual effects versus compositional differences. The first part of the debate concerns the nature of regional differences and in particular whether the differences are more marked for a "geographical" regionalization than for a "functional" one. According to Johnston, Pattie, and Allsopp (1988), the former refers to "contiguous blocks of territory" which "may be separately identified and treated in certain government policies, and voters there share a culture which leads them to respond in similar ways." The functional classification, in contrast, represents "groups of constituencies with similar characteristics, irrespective of location."

The geographical grouping obviously identifies Wales and Scotland, but

within England we can recognize, for instance, that the Black Country (in the West Midlands) has shown working-class Tory support since Joseph Chamberlain's campaign for tariff reform, while Norfolk was a center for agricultural trade-unionism (Cox 1987). Such long-term geographical distinctiveness may provide an important context for political socialization which may affect current political outcomes. Others have argued for the growing importance of a functional regionalization. Thus Savage (1987, p. 66) argues that

whereas in the past constituencies of a similar type often had different political alignments because of the salience of their local political cultures, this is becoming much less apparent and constituencies of a similar type are behaving in similar ways whatever part of the country they are in.

The second part of the debate concerns context and composition. A key question is the extent to which the observed place differences, based on either geographical or functional regionalizations, are "contextual" effects or merely a result of different types of people living in these places, that is, the result of composition (Hauser 1970; Jones and Duncan 1995). For example, the strong support for Labor in the South Wales valley constituencies may be due to a high percentage of the population being of manual working class who, irrespective of location, generally vote Labor. This argument has been made strongly by Rose and McAllister (1990) who have maintained that the apparent differences between places in their voting behavior are the result of differential social, demographic, and family characteristics. Indeed, Bogdanor (1983, p. 53) has stated that: "an elector would tend to vote the same way as an elector from a similar class in Glasgow regardless of national or locational difference."

Consequently, there is a need for empirical evidence on whether people of similar characteristics vote differently in different places and different types of places. We know that places differ in voting outcomes, but does where a place is positioned in either *locational* (geographical) or *social* (functional) space make a difference?

Cross-classified multilevel models (Goldstein 1994a) have been developed to assess the relative importance of overlapping contexts after allowing for differential composition. Overlapping contexts generate a cross-classified structure in which lower-level units nest within a cross-classification of two or more higherlevel units. In the present study, geographical regions are not nested within functional regions, nor are functional groupings nested within geographical; both groupings are at the same level and are therefore cross-classified. Thus, a possible structure (Figure 2) would be individual voters (at level 1) nested with constituencies (at level 2) nested with geographical regions (at level 3) and functional regions (also at level 3). A model based on such a structure would allow the assessment of contextual effects at higher levels, after the inclusion of individual characteristics of voters.

The paper has both substantive and methodological aims. Methodologically, we consider the specification of a range of cross-classified models and discuss the need for efficient computational strategies. Detailed workings of the underlying iterative generalized least squares (IGLS) algorithm (Goldstein 1986) are considered in two appendices. Substantively, we present a series of results from hierarchical and cross-classified multilevel models. This analysis provides an assessment of the relative importance of the cross-classified contexts of geographical and functional regions while taking account of voter and constituency characteristics. To operationalize the study we use data from the 1992 British General Election Study (BES), and for functional and geographical regionaliza-



FIG. 2. A Multilevel Model with Cross-classification at the Higher Level

tions we use the same definitions as Johnston, Pattie, and Allsopp (1988) which gives us thirty-one of the former and twenty-four of the latter.¹

MODEL SPECIFICATION

Strictly Hierarchical Models

The essence of the multilevel approach is to specify a set of linked models at a number of levels. The approach here is to start with a single-level model for voters and then to develop the specification for the hierarchical case to show some of the complexities of which the technique is capable. Only then is the cross-classified model considered. Finally the discrete nature of the response variable will be taken into account.

In the case of a single-level model for voters, a characteristic equation with a mixture of categorical and continuous variables and systematic and random parameters is

Systematic Random

$$y_i = \beta_0 x_{0i} + \beta_1 x_{1i} + \beta_2 x_{2i} + (\varepsilon_i x_{0i})$$
(1)

where the variables are

- y the response variable, whether voted Labor (1) or Conservative (0);
- x_0 the base category, a set of 1s representing people of average age and from the manual classes;

¹The Johnston, Pattie, and Allsopp (1988) regionalizations were in fact adopted with one exception. The functional regions are exactly the same as Johnston, Pattie, and Allsopp and are a classification derived from forty-one variables covering demographic, socioeconomic, and housing characteristics of the 1981 Census. It groups all 633 constituencies into thirty-one types of constituency, which are listed by Crewe and Fox (1984). Johnston, Pattie, and Allsopp (1988) for their geographical regions used the 22-fold classification developed by *The Economist* to present its election coverage. In effect, these represent a breakdown of Standard Regions into non- and metropolitan areas. The slight modification [following Curtice and Steed (1992)] used here is to ensure a finer geographical division of the southeast outside of London, and this results in a total of twenty-four geographical regions. x_1 1 if nonmanual class, 0 otherwise; x_2 age (centered about mean); and subscript *i* signifies an individual voter.

In this linear probability model, the parameters an be interpreted as follows:

- β_0 the intercept term is the "national" probability of Labor voting for individuals of average age and manual social class;
- β_1 is the differential contrast in the probability of voting Labor for nonmanual class compared to manual social class;
- β_2 is the linear increase in the probability of voting Labor with age;
- (ϵ) signifies the random part which allows for fluctuations around the fixed part.

The systematic part represents the general voting/age and voting/class relations that are fixed and unchanging. The random part allows for fluctuations around the fixed part where the term random simply means "allowed to vary." In this single-level equation, the random part represents all the idiosyncratic aspects of individual voting that have not been included in the systematic part of the model. These residuals are usually summarized in a single variance term σ_{ε}^2 . While individuals are allowed to vary, and the extent of this heterogeneity is summarized by this variance term, place differences are assumed not to exist in this model. Such a single-level model presumes that place does not matter, and the effects for age and class are not allowed to vary from place to place; that is they are "fixed."

We can begin to overcome these problems by allowing voting behavior to vary between constituencies in what is called a random-intercepts multilevel model:

$$y_{ij} = \beta_0 x_{0i} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + (\mu_{0j} x_{0ij} + \varepsilon_{ij} x_{0ij}).$$
⁽²⁾

This has two new features. First, we recognize the hierarchical structure of voters nested within the constituencies, signifying this by the subscripts i and j, respectively. Second, there is an additional term, μ_{0j} , in the random part. This term is associated with constituencies (not voters) and represents a differential effect for the place in which voters are resident. A positive value for μ_{0i} indicates a constituency that generally supports Labor, a negative value a disinclination to do so. There are now two random terms: the constituency-specific random term, μ_{0j} , represents place differences after allowing for age and class composition and individual variation; and the level-1 random term, ε_{ij} , represents individual differences after allowing for age, class, and between-place differences. The "average" probability of voting Labor in each of j places depends on an overall general average (β_0) plus an allowed-to-vary, "random" difference for each place (μ_{0i}) . The probability is no longer fixed but constituency differences are allowed to vary according to a higher-level distribution. Making the usual identical and independently distributed (IID) assumptions (Kennedy 1979) this distribution can be summarized by its overall mean, $\hat{\beta}_0$, and its variance, $\sigma^2_{\mu 0}$. Equation (2) implies that there is a positive "correlation" between any two voters in the same constituency, but a zero "correlation" between any two voters from different constituencies. Thus, the covariance of y_{ij} and y_{i+1j} conditional on the fixed part of the model is the covariance between $(\sigma_{\mu 0}^2 + \varepsilon_{ij})$ and $(\sigma_{\mu 0}^2 + \varepsilon_{i+1j})$. However, since ε_i and ε_{i+1j} are presumed independent, this covariance simply reduces to $\sigma_{\mu 0}^2$. A nonzero value for this term implies (equivalently) between-place variation or within-place similarity, and thereby suggests that geographical context is important in understanding voting outcomes after taking into account the age and class composition of places.

70 / Geographical Analysis

Such a random-intercepts model assumes that places are uniformly high or low in terms of voting Labor. In geographical terms, if this model is correct, place differences in voting can be represented by a single map. Such an assumption may be overly simplistic in suppressing important differences. For example, while manual class status may be associated with Labor voting nationally, there may be constituencies where the converse is true. If this is the case, there would need to be separate maps for each social class, and, if the age effect varies from place to place, for young, middle, and older age groups. Such a complex specification requires a model in which all predictors have associated random terms at level 2:

$$y_{ij} = \beta_0 x_{0ij} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + (\mu_{0j} x_{0ij} + \mu_{1j} x_{1ij} + \mu_{2j} x_{2ij} + \varepsilon_{ij} x_{0ij})$$
(3)

so that there is a general class effect across all constituencies (β_1) , and a differential class effect (μ_{1j}) which is specific to that constituency; there is a general age effect across all places (β_2) , and a differential age effect (μ_{2j}) which is specific to that place. These three constituencies differentials are continuously distributed random variables at level 2. Making IID assumptions, the between-place differences can be summarized in a set of variance terms, so that $\sigma^2_{\mu 0}$ estimates the differences between places for manual-class individuals of average age, $\sigma^2_{\mu 1}$ assesses the extent to which the class differentials vary between places, while $\sigma^2_{\mu 2}$ summarizes the extent to which the Labor/age relationship varies geographically. It would be unnecessarily restrictive to assume that these level-2 random terms are independent of each other, so a covariance term is estimated between each pair of random variables. Thus, if $\sigma_{\mu 0\mu 1}$ is negative, areas of high Labor voting for average-age, manual-class individuals have a relatively low differential rate for nonmanual status individuals of average age. The estimation of these strictly hierarchial models is considered in Appendix A.

Cross-classified Models

The discussion so far has considered only elaborations of the strictly hierarchical model; we now turn to the cross-classified model. For simplicity, consider the "random-intercepts" model in which contextual differences are based only on the base category, but now recognizing the cross-classification of geographical and functional regions:

$$y_{ij(kl)} = \beta_0 x_{0ij(kl)} + \beta_1 x_{1ij(kl)} + \beta_2 x_{2ij(kl)} + (\phi_{0k} x_{0ii(kl)} + \delta_{01} x_{0ii(kl)} + \mu_{0i(kl)} x_{0ii(kl)} + \varepsilon_{ii(kl)} x_{0ii(kl)}).$$
(4)

The response is the probability of voting Labor for individual *i* in constituency *j* in functional region *k* in geographical region *l*, the bracket indicating that subscripts *k* and *l* are at the same level, in this case level 3. The two additional random terms in comparison to equation (2), ϕ_{0k} and δ_{01} , represent the differential effect for Labor voting in functional and geographical regions after taking account of individual class and age, as well as "partitioning out" the variability at the constituency and individual levels. These two additional random terms at level 3 are assumed to be independent of each other, and again making IID assumptions, the distribution of these differentials for the functional and geographical regionalization can be summarized by the appropriate variance terms: $\sigma_{\phi 0}^2$ and $\sigma_{\delta 0}^2$. If either or both these terms are nonzero, this suggests that there are regional contextual differences even after allowing for class and age.

Another way to appreciate what the cross-classified model is trying to achieve is to consider the covariance between voters (the level-1 units) implied by equation (4). The variance of the response for given values of the fixed predictor variables in the cross-classified model of equation (4) is the sum of all the variances at each level:

$$\sigma_{\phi 0}^2 + \sigma_{\delta 0}^2 + \sigma_{\mu 0}^2 + \sigma_{\varepsilon}^2$$

while the covariance between two voters in different constituencies in different regions and different functional groupings is zero by definition. The covariance between voters whose constituency is in the same functional, but different geographical, regionalization is

$$\sigma_{\phi 0}^2 + \sigma_{\mu 0}^2;$$

that between two voters whose constituency is in the same geographical, but different functional, regionalization is

$$\sigma_{\delta 0}^2 + \sigma_{\mu 0}^2,$$

while if their constituency is in the same functional and geographical regionalization, the covariance is

$$\sigma_{\phi 0}^2 + \sigma_{\mu}^2 + \sigma_{\mu 0}^2$$

The relative size of these differing (co)variances in assessing the similarity within, and difference between, the various "groupings" of voters is our fundamental measure of the importance of context. If context is not important there will be no similarity between voters and no difference between places.

An important aspect of multilevel models is that contextuality has been accommodated in the model by expanding the random and not the fixed part (Jones and Bullen 1994). A "fixed part" expansion, in which each context is represented in the model by a set of dummy indicator variables, will typically result in a very large number of separate paramaters to be estimated.² This would result in very inefficient estimation with unstable and hence unreliable parameters. In effect, a separate regression would be fitted to each constituency and to each type of region. In contrast, the multilevel approach conceives the contextual differences not as separate entities but as coming from a distribution and is what Casetti (1986) terms a "stochastic expansion." Consequently, with a randomintercepts model of equation (4), only a single parameter, the variance of the differentials, is being estimated at each level $(\sigma_{\phi 0}^2, \sigma_{\delta 0}^2, \sigma_{\mu}^2)$ and information from throughout the sample is used in this estimation. This specification of contextuality through the random part of the model can result in a marked improvement in the precision of the estimates when, as here, the sample is small in relation to the number of higher-level units. A full discussion of these benefits, is given in Jones and Bullen (1994); a brief technical summary is given in Appendix B.

Categorical Response

Finally in terms of model specification, we have to take into account that the response variable is not a continuous variable but only has, in this case,

² Johnston, Pattie, and Allsopp (1988) fit a model with fifty-three dummy variables for functional and geographical regions in their analysis of the "flow of the vote."

two possibilities: voting Labor or Conservative. The problems and potential solutions are more easily appreciated if we unpack equation (4) into two parts. First, we can distinguish the microlevel model for voters (for ease of presentation the detailed subscripting of the predictor variables will now be dropped):

$$E[y_{ij(kl)}] = \pi_{ij(kl)} = \beta_{0j(kl)} x_0 + \beta_1 x_1 + \beta_2 x_2$$
(5)

where E is the expectation operator, and $\pi_{ij(kl)}$ represents the "true" underlying propensity to vote Labor.

Second, a macrolevel model for the higher levels can be written as

$$\beta_{0j(kl)} = \beta_0 + \phi_{0k} + \delta_{01} + \mu_{0j(kl)} \tag{6}$$

so that the probability of voting Labor (after allowing for class and age) depends on the general "national" rate, and geographical, functional, and constituency differentials. The fact that the response is a binary outcome has no effect on the macroequation and we can continue to summarize the higher-level distributions of the continuous differentials by appropriate variance terms. However, the micromodel requires a binomial random term at level 1, and a nonlinear logit specification (Wrigley 1985).

Beginning with the binomial random distribution, we only observe the $y_{ij(kl)}$, the zero or one outcome, so that in even a perfectly specified micromodel, there will be binomial sampling fluctuations. That is, the $y_{ij(kl)}$ given the $\pi_{ij(kl)}$ will have a distribution with a variance of $\pi_{ij(kl)}(1 - \pi_{ij(kl)})^{0.5}$. This is accommodated in the modeling process by including a "weight" in the level-1 random part, which if there is an exact binomial distribution, will have a variance of one. This is an overly restrictive assumption, however, and the variance can be estimated instead of being constrained to one. This allows for under- and over-dispersion (Collett 1991, chap. 6) which is characterized, respectively, by the level-1 variance associated with the weights being less or greater than unity.

The micromodel of equation (5) represents a linear probability model and as such suffers two drawbacks (Duncan et al. 1994). First, probabilities fitted by the model are unconstrained, and it is possible to get nonsensical estimated values outside the range of 0 and 1. Second, a probability value assumes a linear relationship with the predictors, when it seems more reasonable to anticipate a nonlinear one of the form

$$\pi_{ij(kl)} = \frac{\exp(\beta_{0j(kl)}X_0 + \beta_1X_1 + \beta_2X_2)}{1 + \exp(\beta_{0j(kl)}X_0 + \beta_1X_1 + \beta_2X_2)}.$$
(7)

This nonlinear model can, however, be transformed to a linear one through a logit transformation:

$$L_{ij(kl)} = \left[\log \left(\frac{\pi_{ij(kl)}}{1 - \pi_{ij(kl)}} \right) \right] = \beta_{0j(kl)} x_0 + \beta_{1j} x_1 + \beta_{2j} x_2.$$
(8)

The logit, $L_{ij(kl)}$, represents the logarithm of the odds of voting Labor. As the underlying probability, $\pi_{ij(kl)}$ goes from 0 to 1, the logit goes from minus to plus infinity. This ensures that any estimated probabilities will be bounded between 0 and 1. Highly efficient software, MLn (Rasbash and Woohouse 1995), permits the estimation of a wide range of models and the ready graphic display of results. In practice this multilevel logit model is estimated as a nonlinear model (Goldstein 1991; 1994b).

ESTIMATION, COMPUTATION, AND SOFTWARE

The basic concepts of multilevel models have been known for over twenty years. Indeed the cross-classified model features in the seminal paper of Lindley and Smith (1972) as resulting from "exchangeability" within and between regressions. However, it was not until the mid-1980s that estimation schemes were developed that could deal with realistically sized problems. The problems were twofold: the difficulty of estimating the fixed and random parameters simultaneously, and doing so in a computationally feasible form. Existing programs were extremely limited in the size of the problem they could handle by the need to invert large matrices. Recently, a number of operational methods have been developed, including Goldstein's (1986) IGLS algorithm. This procedure is a highly flexible estimation strategy, and the MLn software associated with it is unique in its ability to fit the complex models discussed here. Our intention here is to provide a succinct overview of the estimation process and to deal with the practicalities of estimating cross-classified models using existing software. Brief technical details of the algorithm are given in Appendix A.

In broad terms, IGLS estimation proceeds as follows (Goldstein 1987, Appendix 2). Initial estimates of the fixed terms are derived by ordinary least squares (OLS) estimation, ignoring the higher-level random terms. The residuals based on this initial fit are then regressed on a set of variables defining the structure of the random part to provide initial estimates of the variance/ covariances. These estimates are then used in a generalized least squares analysis to provide revised estimates of the fixed part, which in turn is used to revise the estimates of the random part, and so on until convergence. Crucially, a difficult estimation problem is decomposed into a sequence of linear regressions that can be solved efficiently and effectively. Goldstein (1986) provides a proof that these estimates are consistent, and that if the terms in the random part follow Gaussian distributions, they are the maximum-likelihood estimates.

The software developed by Goldstein and his collaborators exploits the fact that a strictly hierarchical situation, providing units at a lower level are sorted within the higher level, generates a block-diagonal structure (Figure 3a). Taking the example of voters nested within constituencies nested within functional regions, the covariance structure requires only terms for within constituency and within region variation. The IGLS algorithm and MLn software exploits this block diagonality which results in large sparse, but structured matrices to achieve highly efficient computation (Goldstein and Rasbash 1992). From this perspective, the cross-classified model is problematic in that it destroys this block diagonality (compare Figures 3a and 3b). Consequently, it is seemingly impossible to use existing software to undertake computations leading to model estimation. This is not a trivial matter given the resources that have already been expended in developing the software for the hierarchical case.

Fortunately, recent research (Rasbash and Goldstein 1994) has been able to recast the cross-classified model in a "quasi-hierarchical" form, so that their existing software can be pressed into action by using dummy (indicator) variables, constraints and "pseudo levels." This re-formulation can be seen as involving three steps:

(i) For reasons of computational efficiency, specify the most numerous group of the cross-classification, the one with the larger number of units, at the appropriate level of the cross-classification. In our case functional and geographical



FIG. 3. (a) Three-Level Hierarchical Model Showing the Block-Diagonal Nature of the Covariance Terms; (b) Three-Level Cross-classified Model Showing the Lack of Block Diagonality

regions are cross-classified at level 3, but it is the former that is most numerous with thirty-one units, so it is that classification that is specified at level 3. The data are then sorted as voters within constituencies within functional regions to achieve the block diagonal structure shown in Figure 3a, which portrays the structure given in Figure 1. The fixed and random parts of this hierarchical model can then be readily estimated using the IGLS algorithm and existing software.

(ii) Including geographical regions results in the non-block-diagonal structure of Figure 3b, which portrays the lower part of the structure shown in Figure 2. To overcome this, a pseudo higher level is created with a single group that spans the entire data set. In our case, a group of all the geographical regions is created to form a single unit at level 4. The structure is now again hierarchical.

(iii) The final step is to incorporate all the geographical regions without "disturbing" the block-diagonal structure. This is achieved by creating a set of dummy, indicator (0, 1) variables for each one of the twenty-four regions. These are not entered into the fixed part of the model, but are allowed to be random at the highest level, in this case 4. The variance of these twenty-four random parameters is then estimated from the data but they are all constrained to be equal. This is the desired "random-intercepts" variance term for the between geographical regional variation.

To summarize, a four-level model is specified with functional groupings as the third-level unit and with a single level 4 unit spanning the entire data set. Dummy variables for each geographical region are created with coefficients random at level 4 with variance constrained to be equal during estimation. The fourth level is simply a computational device to allow modeling to proceed with existing hierarchical software.

This is a general procedure that can deal with a number of extensions. For example, if the social class term is allowed to vary over both the functional and geographical regionalizations, this will require, as usual, a variance for the base category, a variance for the class differential and a covariance at level 3 for the functional regions. However, to deal with the geographical regions, there will need to be two sets of twenty-four dummies representing the base and class differentials, and these dummies will have to be allowed to vary at level 4 to produce twenty-four variances for the base category, twenty-four variances for the class differentials and twenty-four coveriances. These seventy-two "parameters" then have to be constrained so as to estimate two variances and a covariance for the geographical regions. Multi-way cross-classifications involving more than two groups are also possible.³ In general, p-way cross-classifications are achieved by including sets of random variables with dummy variables for unit membership at the next p-1 higher levels. The cross-classified models so far specified have been additive, but interaction between the functional and regional groupings is also possible (Goldstein 1995).

This procedure allows existing software to be used but it brings with it two substantial problems: storage requirements and slowness of estimation. The problem here is that both speed and storage are greatly affected by the number of elements in a unit and in this case there is a single unit at level 4 that contains all (2,275) voters. The estimation process can be speeded considerably if instead of a single level 4 unit spanning the complete data, "disjoint sets" are

³For example, there may be two alternative functional regionalizations as well as the single geographical one. As Savage (1987, p. 66) argues, "It would be an extremely interesting project to see if theoretically derived clusters such as those offered by Cook (1983) would provide even better results" [than those used by Johnston, Pattie, and Allsopp (1988)].

found in which there is no overlap (Rasbash and Goldstein 1994). These disjoint sets with no cross-unit membership represent separate combinations of functional and geographical groupings. For example, there may be three underlying sets, North, Mid, and South, each with their distinctive combinations. The crossclassification can then be represented, as before by functional regions at level 3 nested within the three disjoint sets at level 4. Sorting on voters within constituencies, within sets, imposes a block-diagonal structure. The pseudo level 4 can then be specified as a combination of sets (each with less than all the voters) and a smaller number of dummy variables for specific regions. This can lead to major improvements in computational efficiency. This all depends, of course, on finding disjoint, nonoverlapping sets. There may be cases where overlap is present but this is based on a small number of observations. The analyst may wish to omit these cells with small sample sizes, thereby trading storage and computational speed for (hopefully) a slight loss of precision in terms of estimation. In the present study, the samples of entire constituencies would need to be omitted to avoid overlap as no "natural" subgroups were found, and therefore this option was not chosen.

The current version of the MLn software allows the dynamic setting of the number of levels (dependent on RAM) and has a number of commands for estimating cross-classified models. A single command (SETX) creates the dummies variables, the specification of the random part, and sets up the constraints on parameter estimation. There are also commands to search for disjoint sets (XSEARCH and BXSEARCH) and for omitting observations (XOMIT) until a suitable number of disjoint sets are found. Cross-classified models with binomial, multinomial, Poisson, and negative binomial distributions can all now be estimated.⁴

ANALYSIS

The analysis is based on the BES survey which was conducted at the same time as the 1992 General Election (Heath et al. 1993). Data were extracted on voting and voters' characteristics for 2,275 voters in 218 constituencies. These constituencies were then classified into twenty-four geographical and thirty-one functional regions. We begin with what is known as the null variance-components model. Such a model simply decomposes the total variance in the response variable into its various levels without allowing for any predictor variable; that is, compositional differences are not taken into account. Table 1 details the results for two strictly hierarchical models and for the cross-classified model. The only fixed term in the models (β_0), represents the log-odds of voting Labor as opposed to Conservative in the national sample.⁵ When the estimated logit (-0.268) for the hierarchical geographical regional Model A is transformed, the result suggests that the probability of voting Labor is 0.43. While this term varies slightly between the models, none of the estimates are large compared to their standard errors. All the level-1 terms for the "binomial" variance are less than 1, suggesting that there is greater homogeneity within constituencies than is being captured by these variance components models.

⁴Technical details of the binomial logit model are given by Goldstein (1991). In the estimation of logit models, approximations based on the second-order Taylor expansion and Partial Quasi-Likelihood have been used (Goldstein 1994b); "extra-binomial" variation has been allowed for the level-1 random term.

⁵The Scottish constituencies were oversampled but we have not taken account of this in the analysis that follows.

Variance Con	nponents Models			
(a) Model A:	Hierarchical, Geographical Regions			
	Level	Parameter	Estimate	Standard Error
Fixed		β_0	-0.268	0.210
Random	Three Geographical regions	$\sigma_{\delta 0}^2$	0.812	0.304
	Two Constituency	$\sigma_{\mu 0}^2$	1.176	0.177
	One Voter	σ_{ϵ}^{2}	0.920	0.029
(b) Model B:	Hierarchical, Functional Regions	-		
	Level	Parameter	Estimate	Standard Error
Fixed		β_0	-0.212	0.223
Random	Three Functional regions	$\sigma_{\phi 0}^2$	1.216	0.385
	Two Constituency	$\sigma_{\mu0}^2$	0.830	0.143
	One Voter	σ_{ϵ}^2	0.9281	0.029
(c) Model C:	Cross-classified, Geographical, and Fu	inctional Regions		
	Level	Parameter	Estimate	Standard Error
Fixed		β_0	-0.161	0.161
Random	Three Geographical regions	$\sigma^2_{\delta 0}$	0.199	0.093
	Three Functional regions	$\sigma_{\phi 0}^2$	0.538	0.183
	Two Constituency	$\sigma_{\mu 0}^2$	0.382	0.077
	One Voter	σ_{ϵ}^{2}	0.753	0.023

TABLE 1

Examining the higher-level random terms of Model A suggests that there are noticeable differences between (or equivalently similarities within) both constituencies and geographical regions, although it is at the former level that the larger random term is found. The estimated differential for each geographical region is given in Table 2. The estimated percentage voting Labor varies from a high of 74 percent in the Industrial Northeast to a low of just 13 percent in the South Coast region. There are large apparent differences between regions, with the estimates confirming a general a north-south division with the exceptions of the Rural North favoring the Conservatives and Inner London favoring Labor. A map of the regions is given in Johnston, Pattie, and Allsopp (1988).

The random terms for the hierarchical model of the functional regions (Model B) show that there are substantial differences between functional regions. Indeed, the estimate of the variance at the functional level in Model B is greater than that for constituencies. The scale of the differences between different functional regions can be appreciated from Table 3. The extremes of support for Labor now extend from 12 percent in Prosperous areas with little industry to 80 percent in Areas with the poorest domestic conditions. The general pattern is a direct association between Labor voting and deprivation but there are exceptions in Poor inner-city areas (relatively anti-Labor) and Modestly affluent urban Scotland (relatively pro-Labor). There is a marked contrast between industrial and metropolitan areas on one hand, and rural and suburban regions on the other.

Model C (Table 1c) gives the estimates for the cross-classified model; it can be seen that the greatest variation between units occurs for the functional regions. This suggests that voters who live in different regions and constituencies but live in places with the same social characteristics will tend to vote the same way. The geographical regional variance is much reduced and is now only barely more

TABLE 2

Differentials for	or Geograp	hical Regions
-------------------	------------	---------------

-		Mode	A I	Mode	l C	Mode	l D	Mode	I E
No.	Region	Logit	Rank	Logit	Rank	Logit	Rank	Logit	Rank
1	Strathclyde	1.21	22	0.27	19	0.26	16	0.30	18
2	East-central Scotland	0.69	19	0.21	17	0.31	19	0.35	20
3	Rural Scotland	-0.43	9	-0.21	7	-0.20	8	-0.22	8
4	Rural North	-0.74	6	-0.13	8	-0.10	11	-0.11	11
5	Industrial Northeast	1.30	24	0.72	24	0.61	24	0.61	24
6	Merseyside	1.24	23	0.58	23	0.29	17	0.29	17
7	Greater Manchester	0.40	16	0.00	14	-0.20	7	-0.25	7
8	Rest of Northwest	0.38	15	0.31	20	0.32	20	0.27	16
9	West Yorkshire	0.24	14	-0.08	10	-0.01	14	-0.02	13
10	South Yorkshire	0.98	21	0.26	18	0.30	18	0.32	19
11	Rural Wales	0.50	18	0.43	22	0.32	21	0.40	22
12	Industrial South Wales	0.48	17	0.13	16	0.42	23	0.46	23
13	West Midlands Conurbation	-0.60	7	-0.47	3	-0.41	3	-0.39	3
14	Rest of West Midlands	-0.83	4	-0.44	4	-0.33	5	-0.37	5
15	East Midlands	-0.10	13	-0.02	13	-0.01	13	-0.01	14
16	East Anglia	-0.28	11	0.05	15	0.09	15	0.10	15
17	Devon and Cornwall	-0.43	10	-0.08	11	-0.15	10	-0.15	10
18	Wessex	-0.90	3	-0.48	2	-0.46	2	-0.48	2
19	Inner London	0.85	20	0.36	21	0.36	22	0.39	21
20	Outer London	-0.23	12	-0.07	12	-0.09	12	-0.09	12
21	South Metropolitan	-0.96	2	-0.42	5	-0.37	4	-0.38	4
22	South Coast	-1.59	1	0.59	1	-0.50	1	-0.50	1
23	North Metropolitan	-0.45	8	-0.24	6	-0.29	6	-0.31	6
24	West Metropolitan	-0.75	5	-0.11	9	-0.19	9	-0.22	9

than twice its standard error.⁶ Comparing the differentials in Models A and B for the geographical regions (Table 2) shows that although the rankings remain broadly the same, the absolute size of the differentials is much reduced. What previously appeared to be differences based on geographical location are seen to be based to a considerable extent on geographical regions containing different types of places. In contrast, the differentials for functional regions based on Models B and C (Table 3) show complex changes. While some effects are reduced, others show little change, while others show an increase. Marked differential support for Labor is now seen for Clydeside and Scottish industrial areas while Conservative support is strong in Agricultural areas and Areas of modest influence with some industry.

In summary, the analysis so far suggests that while there are some differences between geographical regions, there are sizeable differences between functional groupings.⁷ The question is now whether these apparent contextual differences are really a result of differential sociodemogrpahic composition. Table 4 shows the results of fitting Model D in which a wide range of predictor variables for individual voters are included in the fixed part. The choice of level-1 predictor variables is informed by reviews of previous British general elections such as

⁶ The use of ratios of estimated coefficients to their standard errors should not be compared with a Z-ratio when the number of higher level units is small. The procedure is used here as an approximate screening device and not as an exact test. A likelihood deviance test was used to assess the relative improvements in model fit.

 $^{^{7}}$ Comparisons of the relative size of variances have to be made carefully as they will reflect sample size. In this case there is a rough equality between the number of functional and geographical groupings.

TABLE	E 3			;							
Differe	strials for Functional Regions										
		Mode	B	Model	U	Mode	D	Model	Э	Model	<u></u> ш
	Functional classification	¢ _{0k} Logit	Rank	∳₀k Logit	Rank	∳ _{0k} Logit	Rank	¢ _{0k} Logit	Rank	∳ok Logit	Rank
1	Metropolitan inner-city areas with immigrants	1.1	27	0.65	22	0.46	25	0.41	23	0.1	22
61	Industrial areas with immigrants	0.89	25	0.9	26	0.66	28	0.64	29	0.02	18
3	Poorest immigrant areas	0.71	23	0.31	19	0.18	19	0.18	18	-0.04	15
4	Intermediate industrial areas	0.53	20	0.83	25	0.25	20	0.27	20	-0.25	ę
ñ	Old industrial and mining towns	0.83	24	1.47	29	0.37	22	0.38	22	-0.23	S
9	Textile areas	0.7	22	0.72	23	0.46	24	0.53	28	-0.33	61
7	Areas with poorest domestic conditions	1.57	31	1.06	27	1.16	31	1.12	31	-0.25	4
œ	Conurbation local authority housing	1.06	26	1.35	28	0.32	21	0.31	21	-0.08	14
6	Black country	-0.55	6	-0.54	13	-0.18	12	-0.21	12	0.14	24
10	Maritime industrial areas	1.5	29	0.76	24	0.56	27	0.48	25	-0.11	11
11	Poor inner-city areas	-0.49	12	-0.45	14	-0.43	œ	-0.43	æ	0.07	20
12	Clydeside	1.52	30	2.13	31	0.92	30	0.74	30	0.17	25
13	Scottish industrial constituencies	1.23	28	2.07	30	0.71	29	0.52	27	0.43	30
14	Scottish rural areas	-0.97	7	-1.74	9	-1.02	Ļ	-0.91	-	0.0	17
15	High-status inner-metropolitan areas	0.49	19	0.32	20	0.54	26	0.49	26	0.0	16
16	Inner-metropolitan areas	0.66	21	0.43	21	0.38	23	0.43	24	-0.14	6
17	Outer London suburbia	-0.96	80	-1.41	œ	-0.44	6	-0.36	10	-0.09	13
18	Very high-status areas	-1.54	2	-1.85	വ	-0.7	ъ	-0.48	1	-0.12	10
19	Conurbation white-collar areas	-0.54	10	-0.63	12	-0.39	10	-0.25	11	-0.15	œ
20	City constituencies with service employment	-0.53	11	-0.83	10	-0.36	11	-0.36	6	0.04	19
21	Resort and retirement areas	-1.43	с С	-2.14	e	-0.76	4	-0.81	61	0.39	29
22	Areas of recent growth and modern housing	0.2	18	0.27	18	-0.05	13	0.02	15	-0.22	9
23	Stable industrial towns	-0.03	15	-0.4	15	0.05	15	-0.09	13	0.37	28
24	Small towns and rural hinterlands	0.15	17	0.17	17	0.06	16	0.21	19	-0.43	I
25	Southern urban constituencies	-0.33	14	-0.68	11	0.09	17	0.14	17	-0.2	2
26	Areas of modest affluence with some industry	-1.25	ъ	-2.2	61	-0.78	ო	-0.71	ы	0.13	23
27	Metropolitan industrial areas	-0.41	13	-0.95	6	0.09	18	-0.01	14	0.22	27
28	Modestly affluent urban Scotland	0.11	16	-0.36	16	0.04	14	0.07	16	-0.1	12
29	Areas of rapid growth	-1.38	4	-1.89	4	-0.7	9	-0.71	4	0.18	26
30	Prosperous towns with little industry	-1.76	I	-1.58	7	-0.84	61	-0.71	9	0.1	21
31	Agricultural areas	-1.07	9	-2.41	1	-0.68	2	-0.79	ę	0.46	31

TABLE 4

Cross-classified Model with Predictors

	Mode	l D	Mod	Model E	
Terms	Estimate	Standard Error	Estimate	Standard From	
Intercent	-0.058	0.051	0.049	0.040	
Age-ser interactions	-0.038	0.031	-0.048	0.243	
Female are	-0.025	0.094	0.094	0.00	
Female age quadratic	0.020	0.024	-0.024	0.02	
Malo	-0.00003	0.0002	-0.0004	0.0002	
Malo ago	0.001	0.101	0.379	0.0136	
Male age quadratio	0.030	0.004	0.090	0.000	
Education	-0.0001	0.0003	-0.001	0.0003	
Higher education	-0.408	0.176	0.411	0.172	
A and 0 level	-0.403	0.178	-0.411	0.175	
Tanura	-0.234	0.150	-0.207	0.130	
Local authority rent	1 921	0.150	1.910	0.156	
Private rent	0.488	0.109	1.210	0.150	
Income	0.400	0.191	0.479	0.100	
f6_19k	0.484	0.163	0.402	0.160	
More than fook	0.404	0.103	0.495	0.160	
Less than fok	-0.400	0.101	-0.420	0.130	
Unknown	0.400	0.190	0.495	0.194	
Class	-0.107	0.134	-0.160	0.191	
Public-sector salariat	-0 375	0 903	0.426	0.020	
Private-sector salariat	-1.934	0.203	-0.430	0.230	
Routine nonmanual	-0.675	0.210	-1.507	0.241	
Petty hourgeoisie	-0.075	0.155	-0.000	0.131	
Foreman	-1.001	0.200	-1.700	0.277	
Skilled manual	-0.244	0.240	-0.243	0.240	
Unknown	-0.122	0.100	-0.124	0.100	
Employment	-0.300	0.329	-0.904	0.323	
Unemployed	0 707	0.961	0 792	0.950	
	0.707	0.201	0.725	0.239	
I and accorrentical regions					
Intercent σ^2	0.180	0.001	0.102	0.100	
Covariance grass	0.100	0.091	0.195	0.100	
Upper class σ^2			0		
I evel functional regions		—	0		
Intercept σ^2	0.450	0.167	0 500	0 100	
Coverience au	0.439	0.107	0.300	0.190	
Upper alors -2	—		-0.130	0.142	
Upper class $\sigma_{\vec{r}}$	—	—	0.206	0.183	
Level 2 constituency	0.000	0.007	0.000	0.117	
$\sum_{\mu = 0}^{n} \sum_{\mu = 0}^{n} $	0.230	0.087	0.323	0.117	
$Covariance \sigma_{\mu 0\mu 1}$	_		-0.273	0.164	
Upper class σ_{μ_1}	—	—	0.790	0.349	
Level 1 Voters $\sigma_{\mu 0}^2$					
Intercept σ_{e0}^{z}	0.945	0.029			
Lower class $\sigma_{\epsilon l}^2$	—	—	0.982	0.033	
Upper class σ_{e2}^2		—	0.982	0.031	

Denver (1994). With the exception of age, which is a continuous variable centered around the mean, the remaining predictors are all categories. These are represented in the model by a set of indicator variables that are contrasted with the base category of a 46-year-old woman without educational qualifications who lives in an owner-occupied household whose head is employed, receives a "middle" annual income (£12–20,000), and who is classified as unskilled working class. That is the set of individual characteristics that occur most frequently: the stereotypical individual. The intercept is the estimated log-odds of voting Labor as opposed to Conservative for the stereotypical individual across Britain and gives a value, when transformed, of 49 percent favoring Labor. The estimate for the age-sex interactions must be interpreted as a "set" of results. Young women are generally pro-Labor but this declines with age in a roughly linear fashion (the quadratic age term for females is not large). Young men in contrast are generally more pro-Conservative than young women, but the relationship with age is nonlinear. Middleaged men are pro-Labor, older men are less so, and the decline in support is more pronounced than that for women.

Turning now to the rest of the estimates of Table 4, it can be seen that many of them exceed their standard error by more than a factor of two. Both those with Higher education and O and A level (post-16-) qualifications are significantly anti-Labor in comparison with the base category. Both tenure contrasts are also significant with a marked difference between Local authority tenants and Owner Occupiers, the base category. With the exception of the Unknown category, all the income contrasts are also significant and follow the anticipated pattern of low-income support for Labor. While the differences between Publicsector salariat, Foreman, and Skilled manual and the base category (Unskilled manual) are not significant, all other contrasts are. They are all anti-Labor, with the most marked effects being found for the Private-sector salariat, and the Petty bourgeoisie. Finally, unemployed individuals are significantly pro-Labor in contrast to those in employment.⁸

The consequence of including these demographic and social variables for voters in the cross-classified model is to reduce all the higher-level variances, and to make the level-1 variance closer to a binomial distribution. The smallest higherlevel variance is for the geographical regions. Transforming the logits of Table 2 reveals that the probability of voting Labor for the stereotypical voter varies from 0.36 in the South Coast to 0.63 in the Industrial Northeast. The largest higher-level variance remains the between-functional regional variance which is still more than twice its standard error. This suggests that context in the form of the type of constituency is an important element in understanding voting after account is taken of differential composition. Examining the estimates for each functional region (Table 3) shows that the differentials have generally decreased as the variables are included (the exception is High-status inner-metropolitan). However, they remain substantial so that for the stereotypical voter the probability of voting Labor varies at the extreme from 0.75 (Areas with poorest domestic conditions) to 0.25 (Scottish rural areas).⁹ The differences between functional regions have been reduced but remain substantial despite taking account of a wide range of predictor variables.

The estimates of the final Model E to be fitted to the data are given in Table 4.¹⁰ The fixed part remains defined as in Model D, but the random part is allowed to be considerably more complex. A new dummy variable (Upper class) is created with a 1 to signify the combined class categories of Public- and Private-sector salariat and Petty bourgeoisie. This new variable is not included in the fixed part (to do so would cause exact multicollinearity) but it is allowed to vary over all three higher-level contexts by the inclusion of variance and covariance

⁸ The estimates for class, tenure, and unemployment agree closely with those found for the 1987 election by Jones, Johnson, and Pattie (1992).

⁹For comparison the greatest fixed effects for individual class are the difference between unskilled manual (49 percent) and petty bourgeoisie (15 percent) when all the other predictor variables are held at their stereotypical value.

 $^{^{10}}$ Each iteration of Model \dot{E} took approximately five minutes on a 66 MhZ PC; seven iterations were required for convergence.

terms. At level 1 two additional dummy variables are required to allow for differential heterogeneity between upper- and lower-class voters in comparison with the binomial distribution.¹¹

Examining the fixed part estimates of Table 4 reveals little substantial difference from the simpler model. In the random part only the base category variance for the geographical regions is estimated as different from zero, thereby implying that this type of contextuality is not differentiated in terms of class. This is perhaps not surprising given that the model is trying to estimate three parameters on the basis of just twenty-four regions. Comparing the place-specific differentials with the previous model (Table 2) shows little change, with the Industrial Northeast being the most pro-Labor, while the South Coast is the most anti. For the functional regions a nonzero variance and covariance are estimated for the "Upper" contrast, although only the intercept variance is more than twice its standard error; three parameters are being estimated on just thirty-one regions. The variance for the lower-class category is given directly by the variance associated with the intercept, and this is 0.50. The variance for the upper-class category is given by 0.5 + 2(-0.15) + 0.206 [that is, the sum of the variance associated with the intercept, two times the covariance, and the variance for the upper-class differential (Goldstein 1995)]. The result of this calculation (0.4) shows that the betweenplace variation for the two functions of class are approximately the same, but the negative covariance suggests that where the differential working-class vote is positive, there is a tendency for the upper-class differential to be negative (the correlation between the two sets of place specific differentials is -0.33).¹² This is shown in Table 3 with, for example, the logit for the Small towns showing a positive value for the base category (higher than usual support for Labor from the lower classes), but a negative logit for the Upper-class contrast (higher than usual anti-Labor voting by the Upper class). Transforming these values into probabilities, the range of Labor support among the working class goes from 0.26 in Scottish rural areas to 0.77 in Areas with poorest domestic conditions. For the Private-sector salariat, the range is from 0.08 in Scottish rural areas to 0.41 in the Clydeside constituencies. The biggest difference between the classes is found in Textile areas where the probabilities for the working class and Private-sector salariat are 0.62 and 0.25, respectively. The smallest difference is in Agricultural areas where the probabilities are 0.29 and 0.16. These results at the level of the functional regions must, of course, be treated with caution due to the nonsignificance of the upper-class variance and covariance.

At the constituency level, the variance for the upper-class variable is more than twice its standard error. Moreover, the negative correlation (-0.57) between the upper and lower differentials is stronger for constituencies than for functional regions. This suggests a complex geography of constituency preference even after allowing for demographic and social characteristics of individuals, with a tendency for places that are pro-Labor for the working class to be relatively even more anti-Labor for the upper class. At level 1, both classes are found to have the same estimate which is close to the binomial assumption of a unit variance.

CONCLUSIONS

Two sets of conclusions are appropriate: substantive and methodological. In substantive terms, there appears good evidence that voting cannot simply be

¹¹ At level 1, for a number of technical reasons (Bullen, Jones, and Duncan 1997) the random part is specified as variances for two separate dummies; a covariance is not needed as it is impossible for a voter to belong simultaneously to the upper and lower classes.

¹² Derived as the ratio of the covariance to the square root of the product of the variances.

reduced to individual characteristics. These results show that any analysis of British voting that does not take place into account is inadequate. Voting depends not only on who you are (class and age), what you have (tenure and employment), but also on where you live in the contexts of the history, traditions, and economic experience of different types of places. A strong contextuality of voting remains after a wide range of demorgraphic and sociostructural variables are included in the models. There is evidence that this contextuality is complex and differentiated. While differences between geographical regions are not great, the remaining differences between functional regions and constituencies are substantial and, to some extent, the size of effect is differentiated by class. Further research needs to focus on accounting for these differentials by the inclusion of ecological variables and their interactions with individual characteristics in the manner of Jones, Tonkin, and Wrigley (1998).

Methodolgocially, it has been shown that the multilevel model provides a coherent framework for studying individuals and their overlapping contexts. This represents a major breakthrough as geography is often concerned with multiple contexts that do not form neat hierarchies. An effective methodology that tackles a wide range of complex designs is beginning to be developed. Thus, cross-classified structures will also occur in panel or longitudinal designs of individuals who migrate from one locality to another so that subjects are partially crossed with localities rather than strictly nested within them (Raudenbush 1993). The models as currently developed include continuous and multiple category response as well as the binary categorical response considered here. Indeed, it is even possible to deal with multivariate cross-classified models where there are sets of responses that may be continuous, categorical, or even a mixture (Duncan, Jones, and Moon 1996). However, for the full potential of these models to be achieved, there is a need for larger and larger data sets with particular emphasis being placed on sampling a large number of higher-level units (Jones 1994). The BES, for example, has seen a substantial decline in the number of constituencies sampled between 1987 and 1992. Larger data sets, although essential, bring their own problems in terms of efficient computation. While this in part can be resolved by faster hardware with more memory, there remains a need to develop even more efficient computational algorithms. This is particularly the case when there are cross-classifications involving a large number of units, such as panel study which takes into account changing household structure. In summary, multilevel modeling allows us to explore some of the complexity that we know exists in reality (Rose 1974) and, in so doing, provides a technical framework for substantive research questions involving highly differentiated geographies.

APPENDIX A. IGLS ESTIMATION

We can begin with the usual single-level model [equation (2)] recast in matrix terms:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{A1}$$

where **Y** is a vector of *n* observations of the response; **X** is the matrix of the predictor variables, the first column of which, x_0 , is a column of 1s; ε is a vector of random terms; and β is a vector of unknown parameters whose first element, β_0 , is the intercept. The conditional expectation of the variance-covariance matrix, **V**, is given by

$$E(\mathbf{V}) = (\varepsilon \varepsilon^T) \tag{A2}$$

where T represents the transpose. Making the usual assumptions for the random term of a mean of zero $(E(\varepsilon) = 0)$ and a constant variance (σ_{ε}^2) , the (n-by-n) variance-covariance matrix has the form:

$$E(\mathbf{V}) = \begin{pmatrix} \sigma_{\varepsilon}^{2} & 0 & 0 & \cdots & 0\\ 0 & \sigma_{\varepsilon}^{2} & 0 & \cdots & 0\\ 0 & 0 & \sigma_{\varepsilon}^{2} & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & \sigma_{\varepsilon}^{2} \end{pmatrix}$$
(A3)

or equivalently,

$$E(\mathbf{V}) = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & \vdots \\ 0 & 0 & 1 & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \vdots & \vdots & \vdots & \vdots & \vdots \\ \end{pmatrix} \sigma_{\varepsilon}^{2}.$$
(A4)

The main diagonal shows that each observation is presumed to have the same variance (that is, homogeneity and no heteroskedasticity) while the off-diagonal elements, the covariances, are presumed to be zero (that is independent and not autocorrelated).

An exact solution to (A1) is found by minimizing the sum of squared residuals (in matrix terms $\hat{\epsilon}^T \hat{\epsilon}$) to give the least squares estimate:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$
(A5)

where ^ and -1 represent an estimate and a matrix inverse respectively.

The multilevel model can be cast in matrix terms as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\varepsilon} \tag{A6}$$

with Z, the only additional term in comparison to (A1), being a design matrix that structures the random part. Consequently ε now represents not a single random term but potentially a set of terms at each level of the model. The expectation of the variance-covariance matrix is now given by

$$E(\mathbf{V}) = ([\mathbf{Z}\varepsilon][\mathbf{Z}\varepsilon]^T)$$
(A7)

and this does not reduce to (A3) and (A4) for we can anticipate both autocorrelation and heterogeneity.

If the elements of V are known, however, a well-established result (for example, Johnston 1972, chap. 7) is that estimates of the fixed part of the general equation (A6) can be obtained by generalized least squares:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}.$$
(A8)

The key element of Goldstein (1986) is to assume that β is known and to use this

together with the known structure of ${\bf V}$ to derive estimates of the variance and covariances of the random terms.

To make the discussion concrete, let us use the simplest possible two-level hierarchical model, the random intercepts multilevel model with a single level-1 random term:

$$y_{ij} = \beta_0 x_0 + \beta_1 x_1 + (\mu_j + \varepsilon_{ij}) \tag{A9}$$

and deal with a data set where there are just two people in two places. Sorting on place gives

$$y_{11} = \beta_0 x_0 + \beta_1 x_1 + (\mu_1 + \varepsilon_{11})$$

$$y_{21} = \beta_0 x_0 + \beta_1 x_1 + (\mu_1 + \varepsilon_{21})$$

$$y_{12} = \beta_0 x_0 + \beta_1 x_1 + (\mu_2 + \varepsilon_{12})$$

$$y_{22} = \beta_0 x_0 + \beta_1 x_1 + (\mu_2 + \varepsilon_{22})$$
(A10)

If β is known, then the vector of "raw" residuals, **W**, the difference between the actual and fitted values, will also be known:

$$\mathbf{W} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \tag{A11}$$

or for our minimal data set,

$$w_{11} = (y_{11} - \beta_0 x_0 - \beta_1 x_1) = (\mu_1 + \varepsilon_{11})$$

$$w_{21} = (y_{21} - \beta_0 x_0 - \beta_1 x_1) = (\mu_1 + \varepsilon_{21})$$

$$w_{12} = (y_{12} - \beta_0 x_0 - \beta_1 x_1) = (\mu_2 + \varepsilon_{12})$$

$$w_{22} = (y_{22} - \beta_0 x_0 - \beta_1 x_1) = (\mu_2 + \varepsilon_{22})$$
(A12)

The expectation of the variance-covariance matrix (A7) can now be rewritten as W equates to $\mathbf{Z} \epsilon$:

$$E(\mathbf{V}) = \mathbf{W}\mathbf{W}^{T} = \begin{pmatrix} w_{11}w_{11} & w_{11}w_{21} & w_{11}w_{12} & w_{11}w_{22} \\ w_{21}w_{11} & w_{21}w_{21} & w_{21}w_{12} & w_{21}w_{22} \\ w_{12}w_{11} & w_{12}w_{21} & w_{12}w_{12} & w_{12}w_{22} \\ w_{22}w_{11} & w_{22}w_{21} & w_{22}w_{12} & w_{22}w_{22} \end{pmatrix}$$
(A13)

Vectorizing the lower half of this symmetric matrix gives (the operation "vech"):

$$E(\mathbf{V}) = \begin{pmatrix} w_{11}^2 \\ w_{21}w_{11} \\ w_{22}^2 \\ w_{12}w_{11} \\ w_{12}w_{21} \\ w_{12}^2 \\ w_{22}w_{11} \\ w_{22}w_{21} \\ w_{22}w_{12} \\ w_{22}^2 \end{pmatrix}$$
(A14)

Another approach to this variance-covariance matrix is to decompose it into its constituent elements and associated parameters given the basic assumptions of multilevel models. Returning to equation (A13) and noting how the elements of w_{ij} are comprised of the level-1 and level-2 random terms (A12), the expected variance-covariance is given by a symmetric matrix:

$$E(\mathbf{V})$$

$$= \begin{pmatrix} (\mu_{1} + \varepsilon_{11})(\mu_{1} + \varepsilon_{11}) \\ (\mu_{1} + \varepsilon_{21})(\mu_{1} + \varepsilon_{11})(\mu_{1} + \varepsilon_{21})(\mu_{1} + \varepsilon_{21}) \\ (\mu_{2} + \varepsilon_{12})(\mu_{1} + \varepsilon_{11})(\mu_{2} + \varepsilon_{12})(\mu_{1} + \varepsilon_{21})(\mu_{2} + \varepsilon_{12})(\mu_{2} + \varepsilon_{12}) \\ (\mu_{2} + \varepsilon_{22})(\mu_{1} + \varepsilon_{11})(\mu_{2} + \varepsilon_{22})(\mu_{1} + \varepsilon_{21})(\mu_{2} + \varepsilon_{22})(\mu_{2} + \varepsilon_{22}$$

For the random-intercepts hierarchical model of (A9) it is assumed that

- (i) μ_i and ε_{ij} are uncorrelated; that is, they have zero covariances;
- (ii) the variance of y_{ij} conditional on the values of the predictors and fixed coefficients is the variance of $(\mu_i + \varepsilon_{ij})$; that is, $\sigma_{\mu}^2 + \sigma_{\varepsilon}^2$;
- (iii) observations within a place are correlated, so that the conditional covariance between two measurements in the same place is given by

$$\operatorname{cov}(\mu_j + \varepsilon_{ij}, \mu_j + \varepsilon_{kj}) = \operatorname{cov}(\mu_j, \mu_j) = \operatorname{variance}(\mu_j) = \sigma_{\mu}^2$$

when $j = j$ and $i \neq k$;

(iv) observations in different places are uncorrelated, so that conditional covariance between two measurements in different places is given by

$$\operatorname{cov}(\mu_i + \varepsilon_{ij}, \mu_k + \varepsilon_{ik}) = 0, \text{where } j \neq k.$$

Consequently, the expected variance-covariance in (A15) can be rewritten as

$$E(\mathbf{V}) = \begin{pmatrix} \sigma_{\mu}^{2} + \sigma_{\varepsilon}^{2} & & \\ \sigma_{\mu}^{2} & \sigma_{\mu}^{2} + \sigma_{\varepsilon}^{2} & & \\ 0 & 0 & \sigma_{\mu}^{2} + \sigma_{\varepsilon}^{2} & \\ 0 & 0 & \sigma_{\mu}^{2} & \sigma_{\mu}^{2} + \sigma_{\varepsilon}^{2} \end{pmatrix}$$
(A16)

with a block diagonal matrix for each place. This can be separated into two components representing a covariance matrix at level-2 and level-1:

$$E(\mathbf{V}) = \begin{pmatrix} \sigma_{\mu}^{2} & & \\ \sigma_{\mu}^{2} & \sigma_{\mu}^{2} & \\ 0 & 0 & \sigma_{\mu}^{2} & \\ 0 & 0 & \sigma_{\mu}^{2} & \sigma_{\mu}^{2} \end{pmatrix} + \begin{pmatrix} \sigma_{\varepsilon}^{2} & & \\ 0 & \sigma_{\varepsilon}^{2} & \\ 0 & 0 & \sigma_{\varepsilon}^{2} & \\ 0 & 0 & 0 & \sigma_{\varepsilon}^{2} \end{pmatrix}$$
(A17)

or equivalently,

$$E(\mathbf{V}) = \begin{pmatrix} 1 & & \\ 1 & 1 & \\ 0 & 0 & 1 & \\ 0 & 0 & 1 & 1 \end{pmatrix} \sigma_{\mu}^{2} + \begin{pmatrix} 1 & & \\ 0 & 1 & \\ 0 & 0 & 1 & \\ 0 & 0 & 0 & 1 \end{pmatrix} \sigma_{\varepsilon}^{2}.$$
 (A18)

Vectorizing (vech) the lower half of these symmetric matrices gives

$$\begin{pmatrix} \sigma_{\mu}^{2} + \sigma_{\epsilon}^{2} \\ \sigma_{\mu}^{2} + \sigma_{\epsilon}^{2} \\ 0 \\ 0 \\ \sigma_{\mu}^{2} + \sigma_{\epsilon}^{2} \\ 0 \\ \sigma_{\mu}^{2} + \sigma_{\epsilon}^{2} \\ \sigma_{\mu}^{2} + \sigma_{\epsilon}^{2} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 \\ 1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \sigma_{\mu}^{2} \\ \sigma_{\epsilon}^{2} \end{pmatrix}$$
(A19)

Comparing the versions of V given in equations (A14) and (A19), it can be seen that the squares are the estimates of $\sigma_{\mu}^2 + \sigma_{\epsilon}^2$ and the cross-product terms are estimates of σ_{μ}^2 . If β is known and V needs to be estimated, the w_{ij} s are known but σ_{μ}^2 and σ_{ϵ}^2 are not. This suggests that we formulate a model structure in which the response consists of the w_{ij} s, the parameters to be estimated are the level-2 and level-1 variances, and the predictors are the (1,0) elements of equation (A19). That is,

Response
 Predictors Parameters

$$\begin{pmatrix} w_{11}^2 \\ w_{21}w_{11} \\ w_{22}^2 \\ w_{12}w_{11} \\ w_{12}w_{21} \\ w_{12}^2 \\ w_{12}^2 \\ w_{22}w_{11} \\ w_{22}w_{21} \\ w_{22}w_{12} \\ w_{22}^2 \\ w_{22}^2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}$$
 (A20)

Letting $\hat{\beta}^*$ be the parameter vector (the variances), \mathbf{Y}^* be the response vector (that is, squares and products of the raw residuals), \mathbf{X}^* the design (1,0) matrix linking \mathbf{Y}^* to \mathbf{V} , and \mathbf{V}^* is the conditional covariance matrix of \mathbf{Y}^* , then we can obtain estimates by using generalized least squares:

$$\hat{\beta}^* = (\mathbf{X}^{*T} \mathbf{V}^{*-1} \mathbf{X}^*)^{-1} \mathbf{X}^{*T} (\mathbf{V}^*)^{-1} \mathbf{Y}^*.$$
(A21)

As the estimates of σ_{μ}^2 are based only on cross-products within places, there needs to be a reasonable number of places for effective estimation.

When neither β nor V is known, the iterative generalized least squares estimates are those that simultaneously satisfy both (A8) and (A21). The overall algorithm therefore works by deriving an initial estimate for V in equation (A8) by ignoring higher-level structure and assuming the constant variance of (A3), OLS estimates can then be derived by equation (A5) which in turn allows the estimation of square and products of the raw residuals as in (A14). These are then fed into the generalized least squares equation of (A21) to derive initial estimates of the level-2 and level-1 variances, which in turn allows the generalized least squares estimation of the fixed parameters through equation (A8). The process then iterates between (A8) and (A21) until convergence.

When there are more than two random variables at level 1 and/or level 2, the block diagonals have a more complicated structure but the same general IGLS procedure can still be used (Goldstein 1987, Appendix 3.1). The V matrix can also be replaced by any known or estimated covariance matrix so it is possible to model further dependency (temporal or spatial) amongst the level-1 responses (Goldstein 1995, chap. 7). The iterative nature of the algorithm has been exploited in the *MLn* software to allow manipulation and calculation of variables between cycles. This is partially useful in the calculation of generalized multilevel models (such as logit link and binomial level-1 random term for categorized outcomes) that require that the variances are a function of the predicted response (Goldstein 1991) and when there is a need to improve estimation by using predictive quasi-likelihood (Breslow and Clayton 1993).

The generalized least squares equations (A8 and A21) require the inversion of the V and V^{*}. This potentially results in very large storage overheads and an n^2 time dependency. Considerable research effort has gone into developing procedures that render these equations in a computationally more tractable form. In particular, procedures have been developed to exploit the known structure of V, so that this matrix can be evaluated block by block. Consequently, computations can be arranged so that maximum storage required for the two-level model is equal to the number of coefficients that are random multiplied by the number of level-1 units in the largest level-2 unit (Goldstein and Rasbash 1992).

All the discussion has so far focussed on purely hierarchically structured data. The advance of Rasbash and Goldstein (1994) is to recast the cross-classified model within this framework. The key to the hierarchical model is to exploit the known structure of V in the estimation of the variance-covariances of the random parameters. Returning to equation (A17) and the two-level hierarchical model, it is possible to see that V is composed of two elements:

$$E(\mathbf{V}) = \mathbf{V}_{2(2)} + \mathbf{V}_{1(2)} \tag{A22}$$

where $V_{1(2)}$ refers to the level-1 contribution in a two-level model. In this particular case with a single random term at level-1 and level-2, the two matrices are defined as

Level 1 :
$$\mathbf{V}_{1(2)} = \sigma_{\varepsilon}^{2} \mathbf{I}_{n}$$

Level 2 : $\mathbf{V}_{2(2)j} = \sigma_{\mu}^{2} \bigoplus_{j} \mathbf{J}_{(n_{j})}$

where \mathbf{I}_n is the identity matrix for all *n* observations, \oplus is the direct sum operator, and $\mathbf{J}_{(n_j)}$ is an $n_j \times n_j$ matrix of ones for all observations within place *j*. This patterning can clearly be seen in equation (A18).

More generally, the variance-covariance matrix of the multilevel model of equation (A6) can be written as

$$E(\mathbf{V}) = \mathbf{V}_{2(2)} + \mathbf{V}_{1(2)} \tag{A23}$$

$$E(\mathbf{V}) = \left(\bigoplus_{J} (\mathbf{Z}_{j}^{2} \Omega_{2} \mathbf{Z}_{j}^{(2)T})\right) + \left(\bigoplus_{i} (\mathbf{Z}_{ij}^{(1)} \Omega_{1} \mathbf{Z}_{ij}^{(1)T})\right)$$
(A24)

where Ω_1 and Ω_2 are the random parameter matrices at level 1 and level 2, while $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$ are the corresponding "design" matrices for the random part. Thus, for the variance-components random-intercepts model [equation (A9)] the "omega" matrices are scalars, consisting of just σ_{ϵ}^2 and σ_{μ}^2 , but for a "random slopes" model:

$$y_{ij} = \beta_0 + \beta_1 + (\mu_j x_0 + \mu_{ij} x_0 + \mu_{ij} x_1 + \varepsilon_{ij}), \tag{A25}$$

the variance-covariance matrix will be given by

$$E(\mathbf{V}) = \bigoplus_{j} \left(x_{0}^{(j)} x_{1}^{(j)} \begin{pmatrix} \sigma_{\mu_{0}}^{2} & \sigma_{\mu_{0}\mu_{1}} \\ \sigma_{\mu_{0}\mu_{1}} & \sigma_{\mu_{1}}^{2} \end{pmatrix} x_{0}^{(j)} x_{1}^{(j)} \right)^{T} + \mathbf{V}_{1}$$
(A26)

where V_1 remains defined as $\sigma_{\varepsilon}^2 I_n$. The V matrices for higher levels are similarly constructed in a recursive manner, so that for a three-level model,

$$E(\mathbf{V}) = (\bigoplus_k \mathbf{Z}_k^{(3)} \mathbf{\Omega}_3 \mathbf{Z}^{(3)T}) + \mathbf{V}_2 + \mathbf{V}_1.$$
(A27)

Turning now to a two-level cross-classified model,

$$y_{i(jk)} = \beta_0 + \beta_1 x_1 + (\delta_k + \mu_j + \varepsilon_{i(jk)}),$$
(A28)

the variance-covariance of the observations is written as

$$E(\mathbf{V}) = \mathbf{Z}_{1}^{(2)} \boldsymbol{\Omega}_{(1)2} \mathbf{Z}_{1}^{(2)T} + \mathbf{Z}_{2}^{(2)} \boldsymbol{\Omega}_{(2)2} \mathbf{Z}_{2}^{(2)T} + \mathbf{V}_{1(2)}$$
(A29)

where the Ω s are random parameter matrices, and the Zs are design matrices. Each of the design matrices are of the size *n* by the number of relevant higherlevel units and consists of dummy variables that are set to 1 if the *n*th observation is nested within the higher-level unit, 0 otherwise. Thus, the two-level cross-classified model of equation (A29) can be seen as a special case of the three-level model (A27) with a single level-2 unit nested within a single level-3 unit. Alternatively, (A29) can be rewritten as

$$E(\mathbf{V}) = \mathbf{Z}^{(2)} \Omega_2 \mathbf{Z}^{(2)T} + \mathbf{V}_{1(2)}$$
(A30)
where $\mathbf{Z} = (\mathbf{Z}_1^{(2)} \mathbf{Z}_2^{(2)})$
and $\Omega_2 = \begin{pmatrix} \Omega_{(1)2} & 0 \\ 0 & \Omega_{(2)2} \end{pmatrix}$

which is a special case of equation (A24), the two-level model. In both cases, it is possible to use the composite residuals (in the form of WW^T) and the known structure of the design and random parameter matrices to obtain random-part estimates, through generalized least squares.

APPENDIX B. SHRINKAGE

The effects of shrinkage are seen most straightforwardly in a null, randomintercepts, two-level hierarchical model:

$$y_{ij} = \beta_0 + (\mu_i + \varepsilon_{ij}) \tag{B1}$$

where

 σ_{μ}^2 is the variance of the level-2 random term;

 σ_{ϵ}^2 is the variance of the level-1 random term.

As discussed in Appendix A, both these variances are derived from the IGLS algorithm, through using the composite residuals:

$$\hat{w}_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \hat{\beta}_0. \tag{B2}$$

Treating these parameters and variables as known, we have

$$cov(w_{ij}, \mu_j) = variance(\mu_j) = \sigma_{\mu}^2;$$

$$cov(w_{ij}, \varepsilon_{ij}) = \sigma_{\varepsilon}^2;$$

$$variance(w_{ij}) = \sigma_{\mu}^2 + \sigma_{\varepsilon}^2.$$
(B3)

It is therefore possible to consider w_{ij} , μ_j , and ε_{ij} as random variables, each having a mean of zero, and known variances and covariances. Consequently, it is possible to regress any residual on the observed w_{ij} and derive the expected value of the residual on the basis of the observed value and the random parameters, regression being just another name for conditional expectation. Since the residuals are independent between levels, they may be estimated separately at each level. Goldstein (1987, p. 21) gives the expected values for level 1 as

$$\varepsilon_{ij} = \frac{\sigma_{\varepsilon}^2}{\sigma_{\varepsilon}^2 + \sigma_{\mu}^2} \, w_{ij} \tag{B4}$$

so that the composite residual is deflated by the ratio of the level-1 random variation to the total random variation. At level 2, the expected values are given by

$$\mu_j = \frac{n_j \sigma_\mu^2}{n_j \sigma_\mu^2 + \sigma_\varepsilon^2} \,\overline{w}_j \tag{B5}$$

where n_j is the number of people in place j, and \overline{w}_j is the observed place mean residual defined as $(\Sigma_i w_{ij}/n_j)$.

Examining (B5), the effects of shrinkage can be more readily appreciated by dividing both parts of the ratio by n_j and replacing the observed place mean residual by the difference between the OLS overall mean for all places (β_0^*) and the OLS mean for place $j(\beta_{0j}^*)$:

$$\mu_j = \frac{\sigma_{\mu}^2}{\sigma_{\mu}^2 + \sigma_{\epsilon}^2/n_j} (\beta_{0j}^* - \beta_0^*).$$
(B6)

This allows the multilevel, place-specific intercept to be expressed as a function of the OLS grand intercept and the "shrunken" OLS place-specific intercept:

$$\beta_{0j} = \beta_0^* + \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\epsilon^2/n_j} (\beta_{0j}^* - \beta_0^*).$$
(B7)

The shrinkage ratio is then seen as reflecting the reliability of the OLS mean. As the number of people in a place decreases so there is less information about that place and greater shrinkage. The OLS intercept takes no account of how reliable the data are on each place, the same formula is used regardless of whether there are two or two thousand people in place. In contrast, the multilevel intercept will be shrunk toward the overall intercept for all places when it is based on little information, but it will retain its value when it is based on reliable information. Multilevel estimates consequently represent a form of precision-weighted estimation, where the degree of shrinkage is dependent on the information in the data for each place (n_i) and the degree of similarity between places (σ_u^2) .

Goldstein (1987, Appendix 3.2) gives the matrix formulation for (B4) and (B5) when there is more than one random term at any level; he also provides equations for deriving the standard errors of the residuals for both comparative and diagnostic purposes. The effects of shrinkage when there are random intercepts and slopes at level 2 are discussed by Paterson (1990), and Jones and Bullen (1994). They show that the degree of shrinkage is determined by the combination of three factors:

- The number of units with a place: when this is low, there will be greater shrinkage of both the differential intercept and slope to their overall fixed part equivalents;
- (ii) The variation of predictor variable: when this is limited within a place, there
 will be greater shrinkage of the place-specific slope to the overall fixed
 slope;
- (iii) The sign and size of the covariation between the random terms; for example, when this is large and positive, high differential intercepts will be supported by high differential slopes and the effects of shrinkage on both terms will thereby be reduced; however, a place-specific, unreliable large positive differential slope will be shrunk toward the overall slope if the differential intercept is negative, and the covariance between slopes and intercepts remains positive.

Multilevel estimates therefore pool information across places and "borrow strength," so the place-specific relations that are poorly estimated on their own benefit from the information for other places (Jones and Bullen 1994).

LITERATURE CITED

Alker, H. S. (1969). "A Typology of Ecological Fallacies." In Quantitative Ecological Analysis, edited by M. Dogan and S. Rokan. Cambridge, Mass.: MIT Press.

Bogdanor, V. (1983). Multiparty Politics and the Constitution. Cambridge: Cambridge University Press.

- Breslow, N. E., and D.G. Clayton (1993). "Approximate Inference in Generalized Linear Models." Journal of American Statistical Association 88, 9–25.
- Bullen, N., K. Jones, and C. Duncan (1997). "Modelling Complexity: Analysing between Individual and between Place Variation." Environmental and Planning: A 29, 585-609.
- Casetti, E. (1986). "The Dual Expansion Method: An Application for Evaluating the Effects of Population Growth on Development." IEEE Transactions on Systems, Man, and Cybernetics SMC-16(1), 29-39.

Collett, D. (1991). Modelling Binary Data. London: Chapman and Hall.

- Cox, K. R. (1987). "The Individual, the Social, and Reconceptualizing Contextual Effects." Political Geography Quarterly 6, 41-43.
- Crewe, I., and A. Fox (1984). British Parliamentary Constituencies: A Statistical Compendium. London: Faber and Faber.
- Curtice, J., and M. Steed (1992). "Appendix 2: The Results Analysed." In *The British Election of* 1992, by D. Butler and D. Kavanagh. Basingstoke: St. Martin's Press.
- Denver, D. (1994). Elections and Voting Behaviour in Britain, 2d ed. New York: Phillip Allen.
- Duncan, C., M. Gould, K. Jones, and G. Moon (1994). Using Multilevel Models. Mimeo, Department of Geography, University of Portsmouth.
- Duncan, C., K. Jones, and G. Moon (1996). "Health-related Behaviour in Context: A Multilevel Modelling Approach." Social Science and Medicine 42, 817–30.
- Ennis, P. H. (1962). "The Contextual Dimension in Voting." In Public Opinion and Cogressional Elections, edited by W. N. McPhee and W. A. Glaser. Glencoe: Free Press.
- Goldstein, H. (1986). "Multilevel Mixed Linear Model Analysis Using Iterative Generalized Least Squares." Biometrika 73, 43-56.
- (1987). Multilevel Models in Educational and Social Research. London: Charles Griffin.
- (1991). "Nonlinear Multilevel Models, with an Application to Discrete Response Data." *Biometrika* 78, 45-51.

- Goldstein, H., and J. Rasbash (1992). "Efficient Computational Procedures for the Estimation of Iterative Generalized Least Squares." Computational Statistics and Data Analysis 13, 63-71.
- Hauser, R. M. (1970). "Context and Consex: A Cautionary Tale." American Journal of Sociology 75, 645-64.
- Heath, A., et al. (1993). British Election Study 1992: User Guide. ESRC Data Archive, University of Essex, Colchester.
- Johnson, J. (1972). Econometric Methods. Tokyo: McGraw-Hill.
- Johnston, R. J., C. J. Pattie, and J. G. Allsopp (1988). A Nation Dividing? London: Longman.
- Jones, K. (1991). "Multilevel Models for Geographical Research." Concepts and Techniques in Modern Geography. Norwich: Environmental Publications.
- ------- (1994). "Analysing and Designing Extensive Studies in which Place Matters: A Comment on Neighbourhood Incivilities and the Study of Crime in Place." Area 26, 79-81.
- Jones, K., and N. J. Bullen (1994). "Contextual Models of Urban House Prices: A Comparison of Fixedand Random-Coefficient Models Developed by Expansion." *Economic Geography* 70, 252-72.
- Jones, K., and C. Duncan (1995). "Individuals and Their Ecologies: Analysing the Geography of Chronic Illness within a Multilevel Modelling Framework." *Health and Place* 1, 27-40.
- Jones, K., R. J. Johnston, and C. J. Pattie (1992). "People, Places, and Regions: Exploring the Use of Multilevel Modelling in the Analysis of Electoral Data." British Journal of Political Science 22, 343-80.
- Jones, K., P. Tonkin, and N. Wrigley (1998). "Individual and Constituency Characteristics: Analysing the Geography of Voting Behaviour within a Multilevel Modelling Framework." *Transactions of the Institute of British Geographers*, forthcoming.
- Kennedy, P. (1979). A Guide to Econometrics. Oxford: Martin Robertson.
- Lindley, D. V., and A. F. M. Smith (1972). "Bayes Estimates for the Linear Model." Journal of the Royal Statistical Society (Series B) 34, 1-41.
- McMahon, D., and A. Heath (1992). "Class and Party in Britain: Preliminary Results with a Multilevel Logit Model." Multilevel Modelling Newletter 4, 5–8.
- Paterson, L. (1990). "The Use of Multilevel Models to Estimate School Effects." Mimeo, Centre for Educational Sociology, Edinburgh.
- Rasbash, J., and H. Goldstein (1994). "Efficient Analysis of Mixed Hierarchical and Cross-classified Random Structures." *Journal of Educational and Behavioural Statistics* 19, 337-50.
- Rasbash, J., and G. Woodhouse (1995). MLn Command Reference: Version 1.0. University of London: Multilevel Models Project, Institute of Education.
- Raudenbush, S. W. (1993). "A Crossed Random Effects Model for Unbalanced Data with Application in Cross-sectional and Longitudinal Research." Journal of Educational Statistics 18, 321-49.

- Robinson, W. S. (1950). "Ecological Correlations and the Behavior of Individuals." American Sociological Review 15, 351-57.
- Rose, R. (1974). "Britain: Simple Abstractions and Complex Realities." In *Electoral Behaviour: A Compa*rative Handbook, edited by R. Rose. New York: Free Press.
- Rose, R., and I. McAllister (1990). The Loyalties of Voters. London: Sage Publications.
- Savage, M. (1987). "Understanding Political Alignments in Contemporary Britain: Do Localities Matter?" Political Geography Quarterly 6, 53-76.
- Skinner, C. J., D. Holt, and T. M. F. Smith, eds. (1989). Analysis of Data from Complex Surveys. Chichester: Wiley.
- Wrigley, N. (1985). Categorical Data Analysis for Geographers and Environmental Science. Harlow: Longman.