Contextual Models of Urban House Prices: A Comparison of Fixed- and Random-Coefficient Models Developed by Expansion

Article *in* Economic Geography · September 1994

DOI: 10.2307/143993

Kelvyn Jones
University of Bristol

Contextual Models of Urban House Prices: A Comparison of Fixed- and Random-Coefficient Models Developed by Expansion*

Kelvyn Jones and Nina Bullen
Department of Geography, University of Portsmouth, Portsmouth, Hampshire PO1 3HE, U.K.

Abstract: Contextual quantitative models are receiving considerable attention in the geographic literature in the form of models developed by the "expansion method." The majority of these models have involved the expansion of the fixed part of the model and are in many cases equivalent to standard ANOVA and ANCOVA procedures. Models from a "multilevel perspective," in which the expansion takes place in the random part of the model, have also been developed. Although these two methods converge conceptually, major differences in the form of the models can be specified and estimated. This paper develops contextual models of house prices using both approaches. We argue that varying-parameter multilevel models, which recognize that houses are nested within districts, are preferable to the usual single-level, fixed-coefficient models. We illustrate both approaches with data on house prices in London.

Key words: expansion method, multilevel model, contextual models, house prices.

The recent increase in research on contextual quantitative models is moving away from traditional approaches which "deny geography" (Foster 1991, 140). The key feature of such models is parameter variability. Instead of a response variable being related to a predictor by a universal constant, the relationship is allowed to vary according to context. In geography, much of the impetus for such developments has come from the work of Casetti (1972, 1991, 1992), using what he calls the "expansion method." The core of this method is a three-step procedure: (1) an initial model is specified in terms of parameter relationships between a response and a set of predictors; (2) the model is expanded by further specifying that at least one of these initial parameters varies according to another "contextual" model; and (3) the initial equation and expansion equation are combined to form a "terminal" model, which is then estimated.

A particular feature of the original work, still found in the majority of literature today (Casetti and Jones 1992), is that the modeling of context is achieved by expanding what is called the "fixed" part of the model. At the same time as the original Casetti (1972) paper was published, another approach, in which the expansion occurs in the "random" part of the model, was beginning to be developed (Lindley and Smith 1972). This approach—now variously called "variance components analysis" (Longford 1986), "multilevel modeling" (Goldstein 1987), and "hierarchical linear modeling" (Raudenbush and Bryk 1986; Bryk and Raudenbush 1992)—remained for many years an attractive concept, but not a practical procedure owing to the difficulty of developing a valid estimation strategy. While models based on the expansion of

* We would like to thank Harvey Goldstein for his comments on an earlier draft; the usual disclaimers apply. Kelvyn Jones would like to acknowledge a Nuffield Social Science Research Fellowship 1991–92. Material from Crown Copyright Records, made available through the Department of the Environment and the ESRC Data Archive, University of Essex, has been used by the permission of the controller of Her Majesty's Stationery Office.

the fixed part could readily be calibrated using standard software and Ordinary Least Squares (OLS), it was not until the mid-1980s that multilevel estimation procedures were developed (Goldstein 1986; Longford 1986; Mason, Wong, and Entwistle 1984). Indeed, implementation in general purpose statistical software has not yet occurred, although specialist software is now available for multilevel modeling (Bryk et al. 1986; Longford 1986; Prosser, Rasbash, and Goldstein 1991). Not surprisingly, multilevel applications are only now beginning to appear in the geographic literature (Bondi and Bradford 1990; Davies, Martin, and Penn 1988; Duncan, Jones, and Moon 1993; Jones 1991a, 1991b, 1992a, 1992b; Jones and Bullen 1993; Jones, Moon, and Clegg 1991; Jones, Johnston, and Pattie 1992; Jones and Moon 1990, 1991a, 1991b).

Our aim is to compare the two approaches of expanding the fixed and the random part, grounding the explanation in the specification of models of urban house prices, a subject previously approached by Can (1990) from the traditional expansion paradigm and by Jones (1991a) and Jones and Bullen (1992) from the multilevel perspective. We develop a series of fixed coefficient models of house prices and a comparable series of multilevel models. After comparing the two in an empirical illustration of house prices in London, we conclude that the multilevel, random-coefficient formulation of the expansion model has a number of advantages.

Contextual Models of Urban House Prices: Fixed-Part Expansions

The simplest bivariate model can be specified as:

$$Y_i = \alpha X_{0i} + \beta X_{1i} + \epsilon_i \qquad (1)$$

where

Y is the observed price;

i is the subscript denoting each property;

X_1 is a predictor variable: a single housing attribute, such as the size of the property measured by lot size or number of rooms and expressed as a difference from the mean;

β is the associated regression slope term, which represents the price-size relation;

ϵ is the random term; making the usual assumptions of a mean of zero, constant variability and no autocorrelation, this random variable may be summarized by a single variance term, σ^2_ϵ;

α is the regression intercept term associated with the constant term, X_0, which consists entirely of ones. Since the housing attribute variable is expressed as deviations from the mean, the intercept represents the price for the average property in the sample—that is, the price when X_1 is zero.

This population model is thus comprised of two parts: fixed and random. The fixed part, representing the general underlying systematic relationship between price and size, consists of two fixed or unchanging parameters: the intercept and slope. The random part consists of a distribution that allows for the price of individual houses to vary around the intercept after taking account of size. The model can therefore be more fully written as

$$Y_i = \underbrace{\alpha X_{0i} + \beta X_{1i}}_{\text{Fixed}} + \underbrace{\epsilon X_{0i}}_{\text{Random}}. \qquad (2)$$

Alternative Expansions

The expansion can be made in relation to any or all of the fixed regression terms, but to begin with it is the intercept term that will be expanded here. The expansion can take a number of forms. One expansion is according to a trend-surface model based on a polynomial expansion of locational coordinates. According to such a model, the price of the "average of X_1" property is not constant across the city but varies smoothly and continuously from place to place. While such a formulation may in many cases be a suitable one (for example, Miles, Stow, and Jones 1992), it can be anticipated that the spatial variations of house prices are discrete and complex, particularly in the inner city, where gentrified districts are juxtaposed with the poorest areas (Goodman 1981). Such a concept can be accommodated by the expansion being allowed to vary according to discrete space, in which the intercept is related to a set of indicator or dummy variables with a 1 indicating district membership, 0 otherwise. For example, if there are 33 districts, a set of 32 dummy variables ($D_1, \ldots, D_{32}$) will be needed at the expansion stage:

$$\alpha = \alpha_0 X_{0i} + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \ldots + \alpha_{32} D_{32i} \tag{3}$$

with each of the 32 terms ($\alpha_1 \ldots \alpha_{32}$) representing the district differentials from the base district price of α_0. Another alternative is where the expansion is made in terms of a contextual variable that varies over space:

$$\alpha = \alpha_0 X_{0i} + \alpha_1 Q_i \tag{4}$$

where Q measures district quality. Can (1990) operationalizes this concept by using the first principal component based on nine socioeconomic variables, equating the district with a census block. In this expansion, the price of the typical property varies according to the characteristics of the district in which it is located.

Alternative Terminal Models

A terminal model is associated with each type of expansion. For the discrete-space expansion of equation (3), the terminal model is

$$Y_i = \alpha_0 X_{0i} + \beta X_{1i} + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \ldots + \alpha_{32} D_{32i} + \epsilon X_{0i} \tag{5}$$

while for the district-quality contextualized model of equation (4), the terminal model is

$$Y_i = \alpha_0 X_{0i} + \beta X_{1i} + \alpha_1 Q_i + \epsilon X_{0i}. \tag{6}$$

So far, all the expansions have been in terms of the intercept, and the terminal model is consequently a straightforward combination of both the initial and expansion models. The slope term can also be expanded. For example, for the discrete-space expansion it may be postulated that the size relation is a function of the district:

$$\beta = \beta_0 X_{0i} + \beta_1 D_{1i} + \beta_2 D_{2i} + \ldots + \beta_{32} D_{32i}. \tag{7}$$

When this is combined with the initial model, the following terminal model results:

$$Y_i = \alpha X_{0i} + (\beta_0 X_{0i} + \beta_1 D_{1i} + \beta_2 D_{2i} + \ldots + \beta_{32} D_{32i}) X_{1i} + \epsilon X_{0i} \tag{8}$$

or equivalently (note the interactions):

$$Y_i = \alpha X_{0i} + \beta_0 X_{1i} + \beta_1 D_1 X_{1i} + \beta_2 D_2 X_{1i} + \ldots + \beta_{32} D_{32} X_{1i} + \epsilon X_{0i}. \tag{9}$$

Moreover, the initial model (equation 1) can be combined with the discrete-space intercept expansion (equation 3) and the slope expansion (equation 7). The resultant terminal hyperparameter model,

$$Y_i = \alpha_0 X_{0i} + \beta_0 X_{1i} + \alpha_1 D_{1i} + \beta_1 D_1 X_{1i} + \alpha_2 D_{2i} + \beta_2 D_2 X_{1i} + \ldots + \alpha_{32} D_{32i} + \beta_{32} D_{32} X_{1i} + \epsilon X_{0i} \tag{10}$$

suggests a complex geography of house prices based on localized interactions between district and housing attributes.

A distinctive feature of this approach is that all of the expansion has occurred in the fixed part of the model, for the expansion models (equations 3, 4, and 7) contain no random terms, whereas the terminal models (equations 5, 6, 8, 9, and 10) have only the single random term associated with the intercept derived from the initial model. Consequently, all these terminal models with a single random term can be estimated with standard procedures such as ordinary least squares. Indeed, the terminal model of equation (5) is nothing other than ANOVA and the terminal model of equation (10) is the ANCOVA model (Silk 1977). Examining the latter model more carefully shows that this expansion is equivalent to fitting a separate regression model between price and size for each district. This may appear a comforting and desirable result, but it means that an overall model is not really being fitted to all the data simultaneously. In terms of the estimated coefficients, the results will be the same regardless of whether an overall model is fitted or 33 separate ones; in this respect, nothing is being gained by stringing them together in one model. The only difference concerns the variance of the random term, for fitting an ANCOVA model with a single variance term assumes that this variance is unchanging throughout the city. If this is indeed the case, the ANCOVA approach will be more efficient than a set of separate regressions; this may affect the estimated standard errors of the coefficients. We will return to this discussion after considering an alternative form of expansion based on developing the random part of the model.

Contextual Models of Urban House Prices: Random-Part Expansions

From the multilevel perspective, such fixed-part expansions are problematic. This is most easily appreciated by looking at the district-quality terminal model of equation (6). District quality has been made an attribute of each housing unit, with no distinction between houses and the districts in which they are located. Districts and houses are treated as equivalent observations, although houses are likely to be more numerous than districts, and houses within a district are likely to be more similar than houses in a different district. When there is only one observation per district, the within-place variation is totally confounded with the between-place variation and no separate estimate of these distinct components is possible. This is not a problem that can be solved by merely including relevant subscripts because there are two distinct levels of analysis, houses (level 1) and districts (level 2), while the expansion methods, as usually applied, have presumed a single level. These problems can be overcome by modifying the three-step expansion method so that the model at the expansion stage is specified at the appropriate, higher level.

Random-Intercepts Multilevel Model

Returning to the initial model and algebraically detailing the house and district levels:

$$Y_{ij} = \beta_0 X_{0ij} + \beta_1 X_{1ij} + \epsilon_{ij} X_{0ij} \qquad (11)$$

it is now clear in this micromodel that the price of house i in district j is specified as a function of citywide price of the typical property (β_0), the cost of an extra unit of size (β_1), and the price associated with the idiosyncratic elements of the individual house (ϵ). To achieve the equivalent multilevel model to that of the discrete-space fixed expansion of equation (3), the intercept term first has to be indexed:

$$Y_{ij} = \beta_{0j} X_{0ij} + \beta_1 X_{1ij} + \epsilon_{ij} X_{0ij} \qquad (12)$$

so that it can be allowed to vary in a higher, level-2, between-district macromodel:

$$\beta_{0j} = \beta_0 + \mu_{0j}. \qquad (13)$$

β_{0j}, the price of the typical house in district j, is seen as a function of the citywide price, β_0, plus a differential for each district, μ_{0j}. The micromodel is the within-place equation, while the macromodel is the between-place equation, in which one of the parameters of the within-place model is the response. The expansion can now be seen to have occurred at the proper level, in a statistical model that has a response, an intercept, and an appropriate random term. The micro- and macromodels combine to form

$$Y_{ij} = \beta_0 X_{0ij} + \beta_1 X_{1ij} + (\mu_{0j} X_{0ij} + \epsilon_{ij} X_{0ij}) \tag{14}$$

where brackets are used to indicate the random terms. The two sets of random terms, $\mu_0 + \epsilon$ (Goldstein 1987), are assumed to be independent of each other. Under the usual assumptions, their distributions can be summarized by the variance terms, σ^2_ϵ at level 1, $\sigma^2_{\mu 0}$ at level 2. Not surprisingly, this model, in which the intercepts are allowed to vary according to a distribution, is known as the random-intercepts model.

Fully Random Multilevel Models

Figure 1a shows the random-intercepts model; for clarity, only seven districts are shown. Some districts attract a premium (a positive value for μ_0), whereas others are relatively cheap (a negative value for μ_0). These parallel lines imply that the price-size relationship is the same in every district, so that some places are uniformly more expensive than others. This is a restrictive specification; Figures 1b–1d, in which the steepness of the lines is allowed to vary from place to place, present possible alternatives. In Figure 1b the pattern is such that district makes very little difference for small-sized properties, but the price differential for large properties is substantial. In contrast, Figure 1c shows relatively large place-specific differentials for small properties but not for large ones. The final graph, Figure 1d, with its criss-crossing, represents a complex interaction between size and place: some districts are relatively cheap for small houses and others are relatively cheap for large ones.

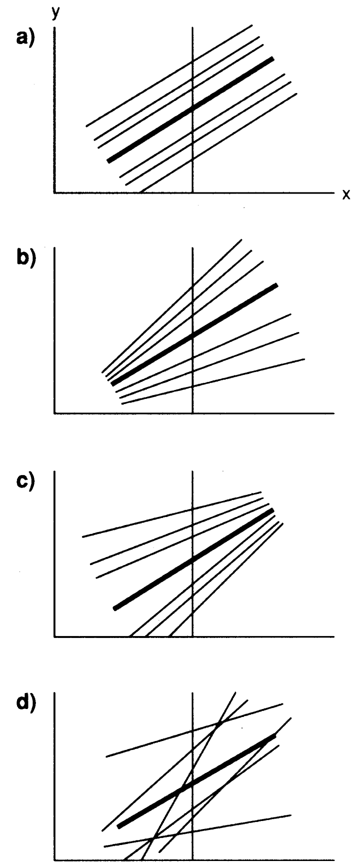

Figure 1. Varying relationships between house price and size.

The complexities of Figures 1b to 1d are achieved in a fully random model in which the slopes as well as the intercepts are allowed to vary. This is achieved by specifying an additional macromodel:

$$\beta_{1j} = \beta_1 + \mu_{1j} \tag{15}$$

which conceives the district slope terms as an average citywide slope plus a differential. The combination of the initial model and the two macromodels produces

$$Y_{ij} = \beta_0 X_{0ij} + \beta_1 X_{1ij} + (\mu_{0j} X_{0ij} + \mu_{1j} X_{1ij} + \epsilon_{ij} X_{0ij}). \tag{16}$$

The μ_{1j} terms are another set of level-2 random terms, and, making the usual assumptions, they can be summarized in a single variance term, $\sigma^2_{\mu 1}$. This model now has six parameters requiring estimation: the two fixed terms representing the overall, citywide intercept (β_0) and slope (β_1); and four random terms, σ^2_ϵ at level 1, $\sigma^2_{\mu 0}$, $\sigma^2_{\mu 1}$, and $\sigma_{\mu 0 \mu 1}$ at level 2. The covariance term, $\sigma_{\mu 0 \mu 1}$, allows the random intercepts and slopes to covary according to a higher-level, joint distribution.

Higher-Level Distributions

The key feature of the multilevel models is that they specify the differential intercepts and slopes not as fixed, separate, and independent, as in the usual expansion model, but as coming from a distribution at a higher level. These distributions concern districts (not houses) and result from treating places as a sample drawn from a population. Figures 2 and 3 reflect aspects of these higher-level distributions that underlie the different graphs of Figure 1. Figure 2 shows a "dotplot" for the distributions of the slopes and intercepts, whereas Figure 3 shows a "scatterplot" of the joint distribution. Not surprisingly, the random-intercepts model of Figure 1a is distinguished from Figures 1b to 1d by the lack of a distribution for the slopes, while the latter three plots are distinguished by the nature of the joint distribution—that is, their covariation. In Figure 1b the cost of an increase in size is greatest in places where the average house is expensive. That is, a steep slope is associated with a high differential intercept, or, equivalently, there is positive covariation between the intercepts and the slopes. In contrast, districts in Figure 1c, where the average house is very expensive, do not command a great deal more for an increase in size. A high differential intercept is generally associated with a shallow slope, or, equivalently, negative covariation. Finally, the complex criss-crossing of Figure 1d is a result of the lack of any pattern or covariation between the intercepts and slopes. The high cost associated with a particular district tells us nothing about

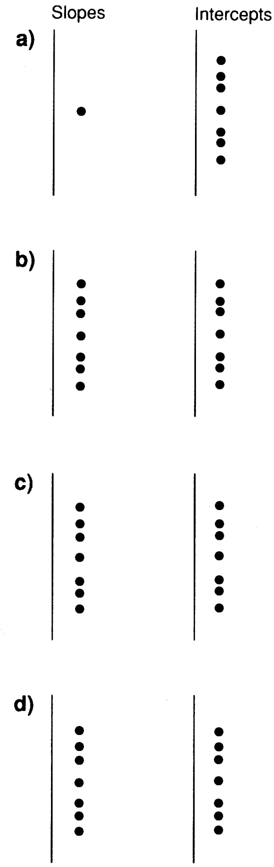

Figure 2. Dotplots of the higher-level distributions underlying Figure 1.

the marginal cost associated with an increase in size in that district.

Table 1 summarizes the higher-level distributions underlying Figure 1 in terms of means, variances, and covariances. Clearly, the means of the distribution are simply the usual intercept and slope representing the overall citywide relationship. The variance/covariances of the higher-level random terms therefore succinctly summarize parameter variability. Indeed, if all the variance terms of the higher-level distributions are effectively zero, there is no contextuality and no need for macromodels; urban house prices are adequately described in terms of the initial model, based solely on house attributes.

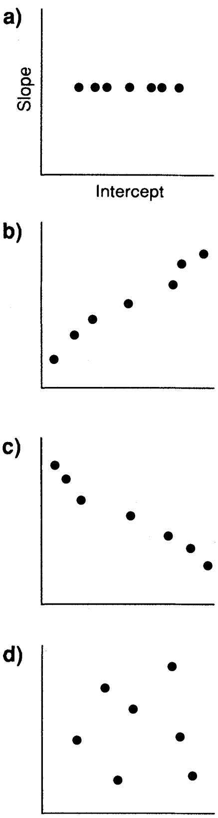

Figure 3. Scatterplots of the higher-level distributions underlying Figure 1.

Estimation

In this form, the fully random model is equivalent to the unbalanced (a different number of houses in each district) random-effects ANCOVA. Until the mid-1980s the lack of general estimation procedures and suitable software severely limited the application of such inherently multilevel models. 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, three alternative operational methods have been developed that give equivalent results: the EM algorithm (Mason, Wong, and Entwistle 1984), a Fisher-scoring algorithm (Longford 1987), and an Iterative Generalized Least Squares (IGLS) procedure (Goldstein 1986). Goldstein's procedure is a highly flexible estimation strategy, and the software associated with it, ML3 (Prosser, Rasbash, and Goldstein 1991), is unique in its ability to fit complex models of the form, to be discussed later (Jones 1992a, 1992b).

In broad terms, IGLS estimation proceeds as follows. Initial estimates of the fixed terms are derived by OLS, ignoring the higher-level random terms. The squared 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.

Extending the Model

All the multilevel elaborations have so far been based on a bivariate two-level model, but this can be extended in a number of ways. For example, additional variables can be included on individual house characteristics and included in a revised micromodel. These may be continuous variables or discrete, and the effects of both these types of variables can be allowed to vary at a higher level by specifying additional macromodels (Jones 1991a; Jones and Bullen 1992) so that, for example, the differential cost of a detached property may vary from place to place. The two-level model can also

Table 1

Figure 1 Represented as Parameters for Two Higher-Level Distributions

	Intercepts		Slopes		Intercept/Slope
Plot	Mean β_0	Variance $\sigma^2_{\mu 0}$	Mean β_1	Variance $\sigma^2_{\mu 1}$	Covariance $\sigma_{\mu 0 \mu 1}$
a	+	+	+	0	0
b	+	+	+	+	+
c	+	+	+	+	-
d	+	+	+	+	0

Note: + is positive difference from zero
- is negative difference from zero.

be extended to higher levels so that houses (level 1) are nested within districts (level 2), which are nested within cities (level 3). Time as well as space can also be included as a level. For example, in a true longitudinal study (Wrigley 1986), time at level 1 can represent the measurement occasion nested within the individual house at level 2. The ML3 (Prosser, Rasbash, and Goldstein 1991) software, as the name implies, can handle complex models with up to three levels, while VARCL (Longford 1987) can handle up to nine levels in a random-intercepts model.

Another extension is to include variables that are measured only at the higher level. For example, if district average prices are thought to reflect district quality, the random-intercept macromodel of equation (13) can be respecified to include an extra term:

$$\beta_{0j} = \beta_0 + \gamma_1 Q_j + \mu_{0j} \tag{17}$$

with the resultant combined model

$$Y_{ij} = \beta_0 X_{0ij} + \beta_1 X_{1ij} + \gamma_1 Q_j + (\mu_{0j} X_{0ij} + \epsilon_{ij} X_{0ij}). \tag{18}$$

In a similar fashion, the district slope term can also be related to district quality:

$$\beta_{1j} = \beta_1 + \gamma_2 Q_j + \mu_{1j} \tag{19}$$

with the combined model requiring an interaction term

$$Y_{ij} = \beta_0 X_{0ij} + \beta_1 X_{1ij} + \gamma_1 Q_j + \gamma_2 Q_j X_{1ij} + (\mu_{0j} X_{0ij} + \mu_{1j} X_{1ij} + \epsilon_{ij} X_{0ij}). \tag{20}$$

The Models Compared

We have described the two approaches largely in their own terms. One may think that the key difference between the two approaches (treating district differences as random terms instead of fixed) brings nothing but more compact symbolization. It would appear that apparently identical models can be specified by either methodology. Thus, the random-intercepts model of Figure 1a would appear to be specified by either the fixed-expansion model of equation (5) or by the random-part expansion of equation (14). Similarly, a model in which there are district-specific intercepts and slopes appears to be specified by either equation (10) or by (16). In this section we cover a number of important differences under four topics: the importance of precision-weighted estimates, heterogeneity, the correct estimation of district-level variables, and autocorrelation. In order to compare the two different forms of model specification directly, we shall calibrate illustrative models on a 5 percent sample of Building Society mortgages collected for the Department of the Environment in the United Kingdom. These data represent the total London sample for the first quarter of 1990, January to March inclusive. The response variable is price in thousands of pounds, the single predictor is size of the property, measured by the number of rooms. The data structure has two levels, with properties nested within 33 districts.[1]

Precision-Weighted Estimation

The key conceptual difference between the two types of specification is that while the fixed-part coefficients are estimated separately, the random-part differentials are conceived as coming from a distribution. This conceptualization results in three practical benefits:

(1) pooling information between districts, with all the information in the data being used in the combined estimation of the fixed and random part; in particular, the overall slope and intercept terms are based on the information for all districts;
(2) borrowing strength, whereby district-specific relations which are poorly estimated on their own benefit from the information for other places;
(3) precision-weighted estimation, whereby unreliable district-specific fixed estimates are differentially down-weighted or shrunk toward the overall citywide estimate; a reliably estimated within-place relation will be largely immune to this shrinkage.

To illustrate the empirical usefulness of these ideas, we first fit a null model without any predictors. The appropriate random-intercepts multilevel model with no predictors is

$$Y_{ij} = \beta_0 X_{0ij} + (\mu_{0j} X_{0ij} + \epsilon_{ij} X_{0ij}). \quad (21)$$

The seemingly equivalent terminal model for a fixed-part expansion is

$$Y_i = \sum_{j=1}^{33} \beta^*_{0j} D_{ij} + \epsilon_i X_{0i} \quad (22)$$

where * is used to distinguish these coefficients from their random-part equivalents β_{0j} (that is, $\beta_0 + \mu_{0j}$).[2] The random-part expansion was estimated by IGLS and the software package ML3; the ANOVA model was estimated by OLS, also using ML3. The summary estimates for the multilevel model are given in Table 2. They suggest that there is significant between-district variance around the overall London citywide average of £84,270 (the level-2 variance of 284.1 is more than twice its estimated standard error). Partitioning the total variance (284.1 + 3112) into each level indicates that some 8 percent of the variation is between districts, while 92 percent of the variation is between houses. Districts are somewhat different in their average price, but not surprisingly there is great internal variation.

Table 3 provides estimates of these district average prices obtained by both procedures, together with the ranked absolute differences between the two types of estimates and the number of houses in each district (n_j); the district weights (w_j) will be explained later. Obviously, the two procedures do not give equivalent results. The most substantial difference is for Sutton; the fixed estimate at £207,000 is some £90,000 greater than the IGLS random-part estimate. A detailed examination shows that there is a general shrinkage of the OLS estimates to derive the IGLS equivalents,

[1] Further details of the sample and more extensive modeling will be found in Jones and Bullen (1992). We acknowledge that districts are unlikely to be coterminous with housing submarkets; we are currently pursuing research with a finer spatial disaggregation. This is only likely to emphasize the differences between the two approaches owing to the need to estimate a very large number of separate estimates if the fixed-part expansion is used.

[2] In this form of the fixed model, the coefficients associated with each indicator variable represent the district intercept and not the district differential, as previously; this form of the model is chosen to facilitate comparison between the two approaches.

Table 2

Multilevel Estimates for the Null Model

Term	Estimate (£000s)	Standard Error
Fixed part		
β_0: overall mean intercept	84.27	
Random part		
Level 2		
$\sigma^2_{\mu 0}$: intercept variance	284.1	117.9
Level 1		
σ^2_{ϵ}: intercept variance	3112	

whereby extreme large and small OLS estimates are shrunk toward the IGLS estimates that have a smaller absolute magnitude. In a number of districts the two estimates are identical.

To account for these differences, we turn to the work of Paterson (1990). He shows that the district-specific random intercept is a weighted combination of the fixed district intercept and the overall multilevel intercept

$$\beta_{0j} = w_j\beta^*_{0j} + (1 - w_j)\beta_0 \qquad (23)$$

while the overall multilevel intercept is a weighted average of all the fixed intercepts:

$$\beta_0 = (\Sigma w_j\beta^*_{0j}) / \Sigma w_j. \qquad (24)$$

The weights are ratios of the true between-district parameter variance to the total variance, which additionally includes sampling variance resulting from observing a sample within each district. Consequently, the weights represent the reliability or precision of the fixed terms:

$$w_j = \sigma^2_{\mu 0} / (v^2_j + \sigma^2_{\mu 0}) \qquad (25)$$

where the random sampling variance of the fixed parameter is

$$v^2_j = \sigma^2_\epsilon / n_j \qquad (26)$$

and n_j is the number of sample observations in the district. When there are genuine differences between the districts and a large district sample size, the sampling variance will be small in comparison to total variability, the associated weight will be close to 1, the fixed term can be reliably estimated, and the random term will be close to the fixed term. As the sampling variance of the fixed parameter increases, however, the weight will be less than one and the multilevel estimate will increasingly be influenced by the overall intercept based on pooling all the information across all districts. Moreover, as is made clear by equation (24), the overall multilevel intercept is a weighted average in which the fixed terms that can be most reliably estimated have the greatest influence. In summary, for the null model, the random-part estimates and the fixed-part estimates will be the same either when there is a large sample or when the fixed term is close to the overall city-wide mean. Put simply, the fixed-part estimate does not differentiate whether two or two thousand houses are involved; the random-part equivalent does.

Returning to Table 3, we can now account for the shrinkage. The pronounced change for Sutton is a result of a very small sample size (4) and the large difference between the fixed-term estimate (£207,000) and the overall London mean (£84,000). In contrast, Redbridge, with only 11 houses in the sample, does not experience any change, as the fixed estimate is already close to the London average. Interesting pairs of contrasts are: Newham and Waltham Forest, which have the same sample size and hence weights, but the former experiences greater shrinkage, as its fixed estimate is further away from the London average; Kingston and Wandsworth, which have the same fixed-part estimate (at £91,000), but the

Table 3

OLS and IGLS Estimates for the Null Model

District	n_j	w_j	Intercepts Fixed	Intercepts Random	Difference	Rank
City	3	0.21	79	83	−4	17.5
Barking	6	0.35	39	68	−29	2
Barnet	31	0.73	93	91	2	26
Bexley	6	0.35	84	84	0	32.5
Brent	32	0.74	73	76	−3	22
Bromley	15	0.57	97	92	5	13.5
Camden	17	0.60	76	79	−3	22
Croydon	18	0.62	72	76	−4	17.5
Ealing	23	0.67	78	80	−2	26
Enfield	34	0.75	80	81	−1	30
Greenwich	26	0.70	58	66	−8	10.5
Hackney	9	0.45	54	70	−16	6
Hammersmith	16	0.59	103	95	8	10.5
Haringey	12	0.52	76	80	−4	17.5
Harrow	22	0.66	81	82	−1	30
Havering	4	0.26	59	77	−18	5
Hillingdon	27	0.71	100	96	4	17.5
Hounslow	26	0.70	99	94	5	13.5
Islington	22	0.66	90	88	2	26
Kensington	12	0.52	124	105	19	3.5
Kingston	8	0.42	91	87	4	17.5
Lambeth	18	0.62	81	82	−1	30
Lewisham	38	0.77	66	70	−4	17.5
Merton	10	0.47	89	87	2	26
Newham	28	0.71	63	69	−6	12
Redbridge	11	0.50	85	85	0	32.5
Richmond	26	0.70	130	117	13	7
Southwark	9	0.45	66	76	−10	8.5
Sutton	4	0.26	207	117	90	1
Tower Hamlets	13	0.54	42	61	−19	3.5
Waltham Forest	28	0.71	75	78	−3	22
Wandsworth	39	0.78	91	89	2	26
Westminster	16	0.59	109	99	10	8.5

former experiences the greater shrinkage as it is based on a smaller sample.[3]

The second illustration concerns a model with a single predictor, X_1, number of rooms deviated around its mean, which is specified as a fully random multilevel model:

$$Y_{ij} = \beta_0 X_{0ij} + \beta_1 X_{1ij} + (\mu_{1j} X_{1ij} + \mu_{0j} X_{0ij} + \epsilon_{ij} X_{0ij}) \quad (27)$$

and as a fixed-part expansion ANCOVA model:

[3] Illustrative calculation for district 1, the City of London: an estimate of the sampling variance is derived using equation (26) and the multilevel estimates of the level-1 variance:

$$\upsilon_1^2 = \sigma_\epsilon^2 / n_1 = 3112 / 3 = 1037.$$

The weights are estimated using equation (25) and the multilevel estimates of the level-2 between-district variance:

$$\begin{aligned} w_1 &= \sigma_{\mu 0}^2 / (\upsilon_1^2 + \sigma_{\mu 0}^2) \\ &= 284.1 / (284.1 + 1037) \\ &= 0.21. \end{aligned}$$

Finally, the district-specific shrunken estimate is calculated using the precision-weighting formula of equation (23):

$$\begin{aligned} \beta_1 &= w_1 \beta_1^* + (1 - w_1)\beta_0 \\ &= 0.21{*}\ 79 + (1 - 0.21){*}84 \\ &= 83. \end{aligned}$$

$$Y_i = \sum_{j=1}^{33} \beta_{0j}^* D_{ij} + \sum_{j=1}^{33} \beta_{1j}^* D_{ij} X_{1ij} + \epsilon_i X_{0i}. \quad (28)$$

Both models were again estimated using ML3, and the summary results from the multilevel model are shown in Table 4. The London-wide average price for an average-sized property is £80,490, while across London an extra room costs over £14,000. All the level-2 variance/covariance terms are more than twice their estimated standard errors, which implies significant differences between districts in the prices for an average-sized property and in the marginal cost of an extra room. The significant and positive covariance term indicates that where the district average price is high, the marginal cost is also high. Graphically the results conform to Figure 1b, where spatial differences are most marked for large properties. The level-1 between-house variation is reduced in comparison to the null model, as the number of rooms is included in the model.

Results from both specifications are given in Table 5. The most marked differences are again for Sutton, with both the fixed-part slope and intercept being shrunk away from highly unlikely and unreasonable values; the fixed-part estimates suggest that in this district one would pay minus £203,600 for an average-sized house and £150,000 for each extra room! There are marked changes for other districts, too, with unreasonable fixed negative slopes (paying less for a larger property) being shrunk in three cases (Barking, Havering, and Tower Hamlets) to positive multilevel estimates, while two other negative fixed slopes are shrunk considerably but remain negative (Southwark, Westminster). These changes are demonstrated most clearly in Figure 4. The plots show the covariation of the intercepts and slopes (as in Figure 3), with the size and direction of the shrinkage being shown by a line joining the fixed-part estimates, shown by a solid circle to their random-part equivalents, shown by an open square. As a result of the very large changes experienced by Sutton, the second plot excludes this outlier, to demonstrate that many other district estimates undergo shrinkage.

The shrinkage in this fully random model is controlled, as before, by both the district sample size and the distance of the fixed-part estimate from the overall multilevel average. Additionally, the shrinkage is determined by the amount of information contained in the predictor values for each district, and borrowing strength now occurs between intercepts and slopes. Equivalent equations to those used earlier can be produced to explain this multidimensional shrinkage, but they must now be cast in a matrix form (Paterson 1990).

Table 4

Multilevel Estimates for the Fully Random Model

Term	Estimate	Standard Error
Fixed part		
β_0: overall mean intercept	80.49	
$\beta 1$: overall mean slope for number of rooms	14.22	2.9
Random part		
Level 2		
$\sigma^2_{\mu 0}$: intercept variance	195.3	81.9
$\sigma_{\mu 0 \mu 1}$: intercept/slope covariance	176.6	62.6
$\sigma^2_{\mu 1}$: slope variance	209.6	68.1
Level 1		
σ^2_{ϵ}: intercept variance	2127	

Table 5

OLS and IGLS Estimates for Model with One Predictor

District	Intercepts Fixed	Intercepts Random	Intercepts Difference	Intercepts Rank	Slopes Fixed	Slopes Random	Slopes Difference	Slopes Rank
City	81.9	80.7	1.20	31	14.30	14.40	−0.10	33
Barking	38.7	65.2	−26.00	4	−8.40	0.24	−8.70	5
Barnet	84.7	82.1	2.65	24	13.60	14.30	−0.70	26
Bexley	79.1	80.5	−1.30	30	15.60	14.40	1.26	21
Brent	73.4	74.6	−1.10	32	7.69	8.15	−0.40	31
Bromley	79.1	87.8	−8.60	10	29.50	25.20	4.23	8
Camden	79.5	76.4	3.12	23	5.07	7.27	−2.10	16
Croydon	72.7	74.8	−2.10	26	7.77	8.50	−0.70	28
Ealing	76.9	74.7	2.12	27	3.66	6.83	−3.10	11
Enfield	80.1	83.8	−3.60	22	23.90	21.40	2.48	14
Greenwich	58.0	65.2	−7.10	12	1.63	0.96	0.67	30
Hackney	58.0	73.5	−15.00	6	6.88	9.19	−2.30	15
Hammersmith	113.0	100.0	12.40	8	37.20	33.50	3.70	10
Haringey	83.5	79.3	4.18	20	10.20	11.30	−1.10	23
Harrow	81.4	82.5	−1.10	33	18.20	17.30	0.88	24
Havering	63.7	69.8	−6.10	15	−21.00	2.73	−24.00	2
Hillingdon	86.8	94.7	−7.90	11	41.60	35.20	6.45	6
Hounslow	92.9	94.4	−1.40	29	34.30	31.30	2.94	12
Islington	89.5	83.2	6.27	14	12.70	13.90	−1.10	22
Kensington	138.0	107.0	30.50	3	40.10	37.50	2.56	13
Kingston	90.0	83.3	6.70	13	14.20	16.20	−2.00	17
Lambeth	81.2	78.9	2.29	25	9.91	11.20	−1.30	19
Lewisham	65.5	70.8	−5.30	19	8.59	7.90	0.69	29
Merton	84.4	82.8	1.60	28	16.80	16.50	0.32	32
Newham	64.7	70.8	−6.00	16	8.00	7.26	0.73	27
Redbridge	70.1	79.0	−8.80	9	18.10	14.00	4.07	9
Richmond	110.0	104.0	5.41	18	39.70	38.90	0.84	25
Southwark	29.4	63.1	−33.00	2	−27.00	−5.30	−22.00	3
Sutton	−203.6	99.6	−302.00	1	150.00	35.91	114.09	1
Tower Hamlets	39.8	64.7	−24.00	5	−2.60	2.49	−5.10	7
Waltham Forest	72.5	76.2	−3.60	21	12.90	11.60	1.27	20
Wandsworth	92.7	86.8	5.98	17	13.90	15.40	−1.50	18
Westminster	75.6	62.6	13.00	7	−27.00	−17.00	−9.80	4

The vector of the multilevel intercept and slope terms for each district is given by

$$\beta_j = \begin{bmatrix} \beta_{0j} \\ \beta_{1j} \end{bmatrix} \tag{29}$$

and its fixed-part counterpart by β^*_j, while the overall multilevel intercept and slope are given by

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \tag{30}$$

A 2-by-2 matrix $\mathbf{V}_j$ gives the random-sample variances and covariances for the vector of the fixed-part terms for each district:

$$\mathbf{V}_j = \sigma_\epsilon^2 \begin{bmatrix} n_j & \Sigma x_{1ij} \\ \Sigma x_{1ij} & \Sigma x_{1ij}^2 \end{bmatrix}^{-1} \tag{31}$$

and $\mathbf{T}$ is the variance/covariance matrix for the district-level random part of the multilevel model:

$$T = \begin{bmatrix} \sigma_{\mu 0}^2 & \sigma_{\mu 0 \mu 1} \\ \sigma_{\mu 0 \mu 1} & \mu_{\mu 1}^2 \end{bmatrix}. \tag{32}$$

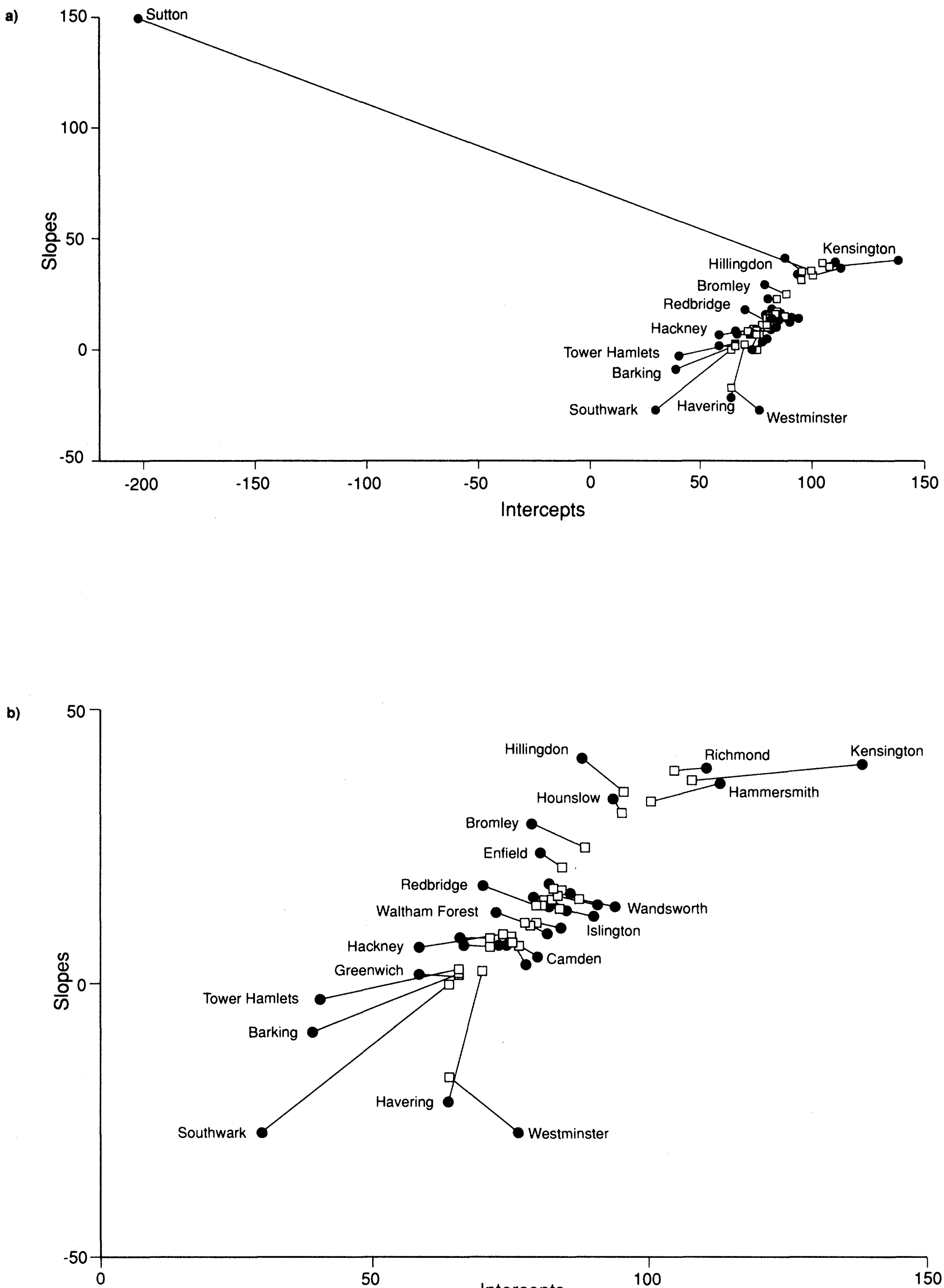

Figure 4. Slopes and intercepts for the fixed and random specifications.

The district-specific weights are now a 2-by-2 matrix:

$$\mathbf{W}_j = \mathbf{T}(\mathbf{V}_j + \mathbf{T})^{-1} \tag{33}$$

and the random-part district term can again be seen as a weighted combination of the fixed district term and the overall citywide multilevel term:

$$\beta_j = \mathbf{W}_j\beta_j^* + (\mathbf{I} - \mathbf{W}_j)\beta \tag{34}$$

where $\mathbf{I}$ is the identity matrix.

Taking just the multilevel intercept, it is possible to "unpack" the elements of this equation to give

$$\beta_{0j} = w_{00j}\beta_{0j}^* + w_{01j}\beta_{1j}^* + (1 - w_{00j})\beta_0 + w_{01j}\beta_1 \tag{35}$$

or, equivalently, by regrouping terms:

$$\beta_{0j} = [w_{00j}\beta_{0j}^* + (1 - w_{00j})\beta_0] + w_{01j}(\beta_{1j}^* - \beta_1). \tag{36}$$

Comparing this shrinkage with its unidimensional equivalent of equation (23) reveals that an extra term (outside the square brackets) is involved. This is the contribution to the shrinkage of the intercept from the difference between the fixed slope and the overall multilevel slope. It can be shown (Paterson 1990) that the weight w_{01j} depends both on the covariance between the multilevel intercepts and slopes (through the off-diagonal elements of $\mathbf{T}$) and on the mean size of the predictor variable within a district in comparison to the overall mean (through the matrix $\mathbf{V}$). As there is a strong positive covariance between the slope and intercept terms in the estimated multilevel model (Table 4), the weight in this case will be relatively large and negative for districts with a low mean house size, and large and positive for districts with a high mean number of rooms.

Examining the case of Sutton in more detail, we now can understand the causes of such extreme shrinkage. The sample data for this district consists of just four houses; three houses are the same size, with 7 rooms, and have a mean price of £170,000; the remaining house has just one more room but has been sold at £320,000. Consequently, the separate estimation of the fixed-part slope term gives a very large slope term for the cost of the extra room (£150,000). The negative value for the intercept, the cost of a house with 4.5 rooms (the London average), is a result of the extrapolation of the very steep slope outside the range of the district data. The small sample size and the lack of variation in house size within the district result in large sampling variances (equation 31), so that fixed terms and slope may be expected to be unreliable. Consequently, the weights, w_{00}, for this district will approach zero, and the random-part estimate will be highly influenced by the overall multilevel terms that are far from the fixed-part estimate. Moreover, there will also be a substantial contribution from the difference in the slope term. The average size of the rooms in this district is higher than the citywide average, so that a relatively large positive weight (w_{01}) is associated with the difference between the fixed district slope and the overall multilevel slope. This contribution will also help pull the fixed intercept away from its extreme negative value.

A contrasting case is offered by the Richmond district. Examining the results for the null model in Table 3, it may be seen that the multilevel intercept is based on a fair amount of shrinkage (the seventh largest absolute difference), despite this district having a relatively large sample size of 26. The relatively large difference between the fixed term and the overall multilevel average is the cause of this shrinkage. A district average price of £130,000 appears somewhat large when the overall city average is only £84,270. In the fully random model, however, the shrinkage experienced by the intercept is relatively less marked, being ranked only eighteenth. The estimate for a high district intercept is being supported by a high district slope (and, of course, vice versa). The shrinkage does not move the intercept so markedly toward the center of the distribution of intercepts because there is evidence from the slope that the district really is atypical.

Shrinkage estimates, derived from specifying a higher-level distribution for the random terms, allows the data to determine an appropriate compromise between specific estimates for different places and the overall fixed estimate that pools information across places over the entire sample. The estimation procedure is both specific and general. Moreover, although it is not possible to estimate the separate relation at all if there are fewer houses than the number of parameters in the district-level model, the multilevel specification does not waste this information but incorporates it into the estimation of the overall estimate. While single-level estimation is troubled when there are many areas and therefore many separate estimates, the multilevel approach thrives on having many districts on which to base an estimate of the variance of their associated differentials (provided, of course, that there is also sufficient information within places).[4] A number of studies reviewed by Morris (1983) have revealed the empirical usefulness and quality of the shrinkage estimates. Efron and Morris (1977) provide an excellent tutorial on shrinkage estimates, and Jones and Moon (1991b) review the use of such procedures in disease mapping.

[4] To obtain reliable estimates of both the within- and between-place variation, a compromise is needed between the number of lower- and higher-level units. This is most efficiently achieved in a multistage design (Goldstein 1991b; Jones 1992a; Skinner, Holt, and Smith 1989). To obtain reliable estimates of place differences, we need many places. Having many houses provides information on the price-size relation within a place, but many places are needed to assess the differences between places. As a rough guide, it has been suggested that a minimum of 25 higher-level units are needed (Paterson and Goldstein 1991).

Heterogeneity

An endemic problem of geographic analysis is spatial heterogeneity. Such heterogeneity may be of substantive interest, while the failure to accommodate it in the model leads to imprecise estimation of the terms that are included. Anselin and Griffith (1988) argue that such heterogeneity has two distinct aspects: structural instability from spatially varying parameters and heteroscedasticity from misspecification, resulting in a nonconstant variance for the random term.

The first source of the problem is, of course, what contextual models are designed to overcome. For the fully random model of equation (27), the total variance between the districts at level 2 is the sum of two random variables:

$$\begin{aligned}\mathrm{Var}(\mu_{0j}X_{0ij} &+ \mu_{1j}X_{1ij}) \\ &= \sigma^2_{\mu0}X_{0ij} \\ &+ 2\sigma_{\mu0\mu1}X_{1ij} \\ &+ \sigma^2_{\mu1}X_{1ij}^{\ 2} .\end{aligned} \tag{37}$$

Clearly, the total variance is not constant but a quadratic function of the predictor variable. Figure 5 shows this graphically by plotting the total variance at level 2 against the number of rooms, using the estimates of Table 4. We now can see that the greatest variation between districts is for large and small properties; the small-

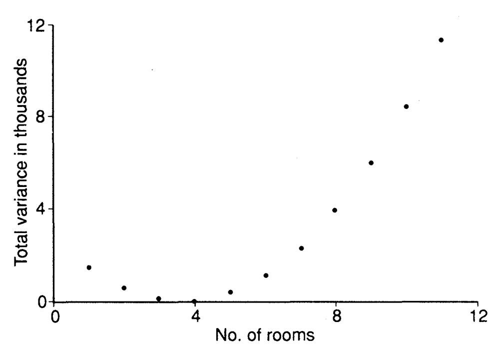

Figure 5. Variance heterogeneity at level 2.

est spatial variation in prices is for properties of average size.

From this perspective, multilevel modeling can be seen as a technique that models the structure of the variation that is not accounted for by the fixed part of the model but does not assume a constant variance that can be summarized in a single parameter. Moreover, there is no reason to anticipate that a single parameter fully captures the random variation in the micromodel. The relationship between price and size within a district, for example, might be inherently more variable for large properties. Such a proposition can be accommodated by expanding the random part at level 1, to give a multilevel model with an additional random term for room size:

$$Y_{ij} = \beta_0 X_{0ij} + \beta_1 X_{1ij} + (\mu_{1j} X_{1ij} + \mu_{0j} X_{0ij} + \epsilon_{0ij} X_{0ij} + \epsilon_{1ij} X_{1ij}). \tag{38}$$

Instead of using a transformation, which is a common approach to heteroscedasticity (Cox and Jones 1981; Forrest 1991), the nonhomogeneity is being directly modeled. Table 6 gives the IGLS estimates for this model obtained by using ML3. Both additional level-1 terms, the variance and the covariance, are positive and considerably more than twice their standard error, thereby suggesting that individual prices are indeed more variable for large properties. Most importantly, the inclusion of these terms has affected the estimates for the other terms. In particular, in comparison to Table 4, the level-2 slope variance has been substantially reduced. What appeared to be differences between districts in the price-size relation is in part a misspecification of the micromodel.

Recently, Goldstein, Prosser, and Rasbash (1990) have developed a very general multilevel formulation in which the random term at level 1 is a function of some or all of the fixed coefficients. They achieve this by creating a pseudo variable whose values depend on the current values of the fixed coefficients. This created variable is then allowed to vary in the random part. Such a complex specification is made possible by the flexibility of their ML3 software, which permits modifications to be made to variables between each iteration of the estimation of the fixed and random parts. In terms of specific models, they consider how to estimate a "constant coefficient of variation" model, in which the response variance is proportional to the square of the expected value of the response. Such a specification presumes that the heterogeneity is a function of all the predictors; but they are also able to realize more general formulations with homoscedasticity based on some of the variables, as well

Table 6

Multilevel Estimates for the Model with Complex Level-1 Variation

Term	Estimate	Standard Error
Fixed part		
β_0: overall mean intercept	78.77	
β_1: overall mean slope for number of rooms	11.44	2.0
Random part		
Level 2		
$\sigma^2_{\mu 0}$: intercept variance	242.3	88.8
$\sigma_{\mu 0 \mu 1}$: intercept/slope covariance	99.3	45.7
$\sigma^2_{\mu 1}$: slope variance	61.0	31.7
Level 1		
$\sigma^2_{\epsilon 0}$: variance	1262	
$\sigma_{\epsilon 0 \epsilon 1}$: covariance	470.6	60.5
$\sigma^2_{\epsilon 1}$: variance	413.1	63.6

as models that involve higher-level variables. Thus, the variance of the level-1 random term could be specified as an increasing or decreasing function of district quality. This ability to model heterogeneity is of particular importance when the response is a proportion or is discrete (Goldstein 1991a, 1991b; Jones, Johnston, and Pattie 1992).

Higher-Level Variables

Higher-level variables such as district quality are treated differently in the fixed- and random-part expansions. In the fixed-part expansion (equation 4), there is no random term, but in the macro-equations of equations (17) and (19) there is. The latter multilevel specification correctly recognizes that the variability of the district-level differences (the response in the macromodel) is composed of two parts (Bryk and Raudenbush 1992). As discussed in relation to equation (25), the true parameter variance reflects the real differences between places, and the sample variance is a consequence of working with a sample of houses within each district. The former potentially can be accounted for by including contextual higher-level variables in the macromodel, while the latter is unexplainable in terms of any predictor variables. Multilevel models distinguish these two different levels of variation through the random terms, but the explanatory power of any predictors defined at the higher level is likely to be incorrectly estimated if a single-level model is used (Aitken and Longford 1986, 15). Bryk and Raudenbush (1992) provide several empirical illustrations of this effect.

Autocorrelation

Autocorrelation or dependence is a key feature of any geographic analysis. If there is no autocorrelation in the response, it is unlikely that the variable would command geographic attention. Moreover, with hierarchically structured data sets such as houses nested in districts, autocorrelation can be anticipated as the norm, for geographic data points within the same cluster are likely to be similar. In the survey-analysis literature, this effect is known as intraclass or intercluster correlation (Skinner, Holt, and Smith 1989). Standard statistical techniques, however, assume that data are independent. Nonfulfilment of this assumption results in misestimated precision, resulting in incorrect standard errors, confidence limits, and statistical tests. If positive autocorrelation is not taken into account, there is an increased risk of finding differences and relationships where none exists and of building overcomplicated models (Cliff and Ord 1981; Skinner, Holt, and Smith 1989).

The basic random-intercepts model of equation (14) implies a positive correlation between the house prices for any two properties in the same district, but a zero correlation between the prices of any two properties from different districts. Thus, the covariance of Y_{ij} and Y_{i+1j} is the covariance between $(\mu_{0j} + \epsilon_{ij})$ and $(\mu_{0j} + \epsilon_{i+1j})$, and since ϵ_{ij} and ϵ_{i+1j} are assumed to be independent, this covariance simply reduces to $\sigma^2_{\mu 0}$. The variance of Y_{ij} or Y_{i+1j}, conditional on the fixed part of the model, is $(\sigma^2_\epsilon + \sigma^2_{\mu 0})$, so that the autocorrelation is given by

$$p = \sigma^2_{\mu 0} / (\sigma^2_\epsilon + \sigma^2_{\mu 0}). \qquad (39)$$

This coefficient measures cluster homogeneity—that is, the degree of similarity of property prices within districts after taking account of predictor variables. It can vary between 1, when all the variation is between districts ($\sigma^2_\epsilon = 0$), and 0, when all the variation is between houses ($\sigma^2_{\mu 0} = 0$). If this ratio is zero, there is no autocorrelation and only a single-level model is needed. But if the ratio is not zero, the multilevel approach exploits this dependence to derive improved estimates, while the standard errors of the estimates are "adjusted" to take account of the autocorrelation (Goldstein 1987). As we have already seen for the null random-intercepts model, the ratio is 0.08 and the estimate for $\sigma^2_{\mu 0}$ is more than twice its standard error, suggesting that a multilevel model is needed for these data. While OLS estimation is severely troubled by dependence, the multilevel approach explicitly takes the autocorrelation into account during the modeling process.

The multilevel models so far considered are based on two key assumptions of strict hierarchy and independence among the level-1 units. Goldstein (1987, 80–83) has developed estimation schemes for more complex structuring, where each house at level 1 is nested in two higher-level units at level 2; such a structure is known as a cross-classification. Moreover, he has shown (1987, 83–84) that his estimation procedure can also be used when there is dependence among the level-1 units. Specifically, he develops a model for autoregressive time-series analysis, but in principle the method could be extended to spatial series, provided that it is possible to specify the covariance matrix between the level-1 units (Goldstein, pers. comm.).

Conclusions

Multilevel procedures have a number of features that make them attractive: the assumption of inherent autocorrelation; standard errors and confidence intervals not based on misestimated precision; the ability to specify models with complex heteroscedasticity; precision-weighted estimation based on pooling data and borrowing strength; and the proper assessment of the relative importance of higher-level variables. Processes operating at several levels over discrete space present a strong argument for using multilevel modeling. If the hoped for "paradigmatic intersections" (J. P. Jones III 1992) with research on localities, structuration, realism, and the new regional geography is to be achieved, the multilevel variant of the expansion method deserves considerable attention by geographers. By working simultaneously at the individual and contextual levels, these analytic models begin to reflect the social organization of life. By providing estimates of both the average effect of a variable over a number of settings and the extent to which that effect varies over settings, these models provide a means of thick quantitative description. The complexity of the real world of people and places, both with a history, is not ignored in the pursuit of a universal equation, specified at a single level.

References

Aitken, M., and Longford, N. 1986. Statistical modelling in school effectiveness studies (with discussion). *Journal of the Royal Statistical Society, Series A* 149:1–43.

Anselin, L., and Griffith, D. A. 1988. Do spatial effects really matter in regression analysis? *Papers of the Regional Science Association* 65:11–34.

Bondi, L., and Bradford, M. 1990. Applications of multilevel modelling to geography. *Area* 22:256–63.

Bryk, A. S., and Raudenbush, S. W. 1992. *Hierarchical linear models.* London: Sage.

Bryk, A. S.; Raudenbush, S. W.; Seltzer, M.; and Congdon, R. T. 1986. *An introduction to HLM: Computer program and user's guide.* Chicago: Department of Education, University of Chicago.

Can, A. 1990. The measurement of neighborhood dynamics in urban house prices. *Economic Geography* 66:254–72.

Casetti, E. 1972. Generating models by the expansion method: Applications to geographical research. *Geographical Analysis* 4:81–91.

———. 1991. The investigation of parametric drift by expanded regression: Generalities, and a family-planning example. *Environment and Planning A* 23:1046–61.

———. 1992. The dual expansion method: An application for evaluating the effects of population growth on development. In *Applications of the expansion method,* ed. E. Casetti and J. P. Jones III, 10–41. London: Routledge.

Casetti, E., and Jones, J. P. III, eds. 1992. *Applications of the expansion method.* London: Routledge.

Cliff, A. D., and Ord, J. K. 1981. *Spatial processes, models and applications.* London: Pion.

Cox, N. J., and Jones, K. 1981. Exploratory data analysis. In *Quantitative geography*, ed. N. Wrigley and R. J. Bennett, 135–43. London: Routledge.

Davies, R. B.; Martin, A. M.; and Penn, R. 1988. Linear modelling with clustered observations: An illustrative example of earnings in the engineering industry. *Environment and Planning A* 20:1069–84.

Duncan, C.; Jones, K.; and Moon, G. 1993. Do places matter? A multilevel analysis of regional variations in health behaviour in Britain. *Social Science and Medicine* 37:725–33.

Efron, B., and Morris, C. 1977. Stein's paradox in statistics. *Scientific American*, May, 119–27.

Forrest, D. 1991. An analysis of house-price differentials between English regions. *Regional Studies* 25:231–38.

Foster, S. A. 1991. The expansion method: Implications for geographic research. *Professional Geographer* 43:131–42.

Goldstein, H. 1986. Multilevel mixed linear model analysis using iterative generalised least squares. *Biometrika* 73:43–55.

———. 1987. *Multilevel models in educational and social research*. London: Charles Griffin.

———. 1991a. Nonlinear multilevel models. *Biometrika* 78:45–52.

———. 1991b. Multilevel modelling of survey data. *The Statistician* 40:235–44.

Goldstein, H.; Prosser, R.; and Rasbash, J. 1990. Multilevel models in which the random variation is a function of the fixed coefficients. Institute of Education, University of London. Mimeo.

Goodman, A. C. 1981. Housing submarkets within urban areas: Definitions and evidence. *Journal of Regional Science* 21:175–85.

Jones, J. P. III. 1992. Paradigmatic dimensions of the expansion method. In *Applications of the expansion method,* ed. E. Casetti and J. P. Jones III, 42–62. London: Routledge.

Jones, K. 1991a. Specifying and estimating multilevel models for geographical research. *Transactions of the Institute of British Geographers* 16:148–59; Amendment to "Specifying and estimating multilevel models for geographical research." *Transactions of the Institute of British Geographers* 17:246.

———. 1991b. *Multilevel models for geographical research. Concepts and techniques in modern geography.* Norwich, U.K.: Environmental Publications.

———. 1992a. Using multilevel models for survey analysis. In *Survey and statistical computing*, ed. A. Westlake, R. Banks, C. Payne, and T. Orchard, 231–42. Amsterdam: Elsevier.

———. 1992b. Review of multilevel software: ML3. *Computers in Higher Education Economics Review* 17:36–41.

Jones, K., and Bullen, N. 1993. A multilevel analysis of the variations in domestic property prices: Southern England 1980–87. *Urban Studies* 30:1409–26.

Jones, K.; Johnston, R. J.; and Pattie, C. J. 1992. People, places, regions: Exploring the use of multilevel modelling in the analysis of electoral data. *British Journal of Political Science* 22:343–80.

Jones, K., and Moon, G. 1990. A multilevel model approach to immunisation uptake. *Area* 22:264–71.

———. 1991a. Multilevel assessment of immunisation uptake as a performance measure in general practice. *British Medical Journal* 303:28–31.

———. 1991b. Medical geography: Multilevel perspectives. *Progress in Human Geography* 15(4):437–43.

Jones, K.; Moon, G.; and Clegg, A. 1991. Ecological and individual effects in childhood immunisation uptake: A multilevel approach. *Social Science and Medicine* 33:501–8.

Lindley, D. V., and Smith, A. F. M. 1972. Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B* 34:1–41.

Longford, N. T. 1986. VARCL: Interactive software for variance components analysis. *Professional Statistician* 5:28–32.

———. 1987. A fast-scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika* 74:817–27.

Mason, W. M.; Wong, G. Y.; and Entwistle, B. 1984. Contextual analysis through the multilevel linear model. In *Sociological methodology,* ed. S. Leinhardt, 72–103. San Francisco: Jossey-Bass.

Miles, M.; Stow, D. A.; and Jones, J. P. 1992. Incorporating the expansion method into remote sensing-based water quality analyses. In *Applications of the expansion method,* ed. E. Casetti and J. P. Jones III, 279–96. London: Routledge.

Morris, C. 1983. Parametric empirical Bayes.

Journal of the American Statistical Association 78:47–65.

Paterson, L. 1990. The use of multilevel models to estimate school effects. Centre for Educational Sociology, Edinburgh. Mimeo.

Paterson, L., and Goldstein, H. 1991. New statistical methods for analysing social structures: An introduction to multilevel models. *British Educational Research Journal* 17: 387–93.

Prosser, R.; Rasbash, J.; and Goldstein, H. 1991. *ML3: Software for three-level analysis*. London: Institute of Education, University of London.

Raudenbush, S. W., and Bryk, A. S. 1986. A hierarchical model for studying school effects. *Sociology of Education* 59:1–17.

Silk, J. 1977. *Analysis of covariance and comparison of regression lines. Concepts and techniques in modern geography* 20. Norwich: GeoBooks.

Skinner, C. J.; Holt, D.; and Smith, T. F., eds. 1989. *Analysis of data from complex surveys*. Chichester: John Wiley.

Wrigley, N. 1986. The era of longitudinal data analysis. *Progress in Human Geography* 10:84–102.