Dimension Reduction in Hierarchical Linear Models[1]

Yoshio Takane[1] and Michael A. Hunter[2]

[1]McGill University
1205 Dr.Penfield Avenue, Montreal, Quebec, Canada H3A 1B1
[2]University of Victoria
P. O. Box 3050, Victoria, British Columbia, Canada V8W 3P5

Summary: In many disciplines of social sciences, data are often hierarchically structured. Academic performance may be measured of students who are nested in classes which are in turn nested within schools. Multi-level analysis based on the hierarchical linear model (HLM) has been effectively used to capture the hierarchical nature of such data. Most of the existing studies that employ HLM, however, use only a few predictor variables at all levels, because interpretation of parameters in HLM will become increasingly more difficult as the number of parameters increases. To alleviate the difficulty, we propose a method of reducing the dimensionality of the parameter space in HLM in a manner similar to reduced-rank regression models. We describe the two-level HLM, present a parameter estimation procedure and suggest where the rank-reduction may be applied. An example is given to illustrate the proposed method.

1. Introduction

In many fields of social sciences, data collected often have hierarchical structures. For example, academic performance may be measured of students who are nested in classes which, in turn, are nested within schools. Measurements may be repeatedly taken of an attribute from individuals grouped by the region of their domicile, and so on. Multi-level analysis based on hierarchical linear models (HLM) has been effectively used to capture the hierarchical nature of such data (Bock, 1989; Bryk & Raudenbush, 1992; Goldstein, 1987; Hox, 1995).

Most of the existing studies that employ HLM, however, use only a few (typically, one or two) predictor variables at all levels. This is primarily because interpretation of parameters in HLM becomes increasingly more difficult with the increasing number of parameters in the model. Numerical difficulties often encountered in fitting HLM with a moderate number of predictor variables may be another contributing factor to this practice. To alleviate the difficulty, we propose a method of reducing the dimensionality of the parameter space in HLM. This is done in a manner similar to reduced-rank regression models (Anderson, 1951) or equivalently redundancy analysis

[1]In Nishisato, S. et al. (Eds.), *Measurement and Multivariate Analysis* (pp. 145-154). Tokyo: Springer Verlag

(Van den Wollenberg, 1977). The dimension reduction improves the interpretability of model parameters, particularly when there are a large number of parameters in the model.

2. The method

2.1 The model

We consider a two-level hierarchical linear model (HLM). Extensions to higher-level HLM are straightforward. Throughout this paper, we assume a fixed-effect model. While this is a bit unconventional, it may be justified on the ground that the dimension reduction is most pertinent in the exploratory mode of data analysis.

Suppose there are N_j $(j = 1, \dots, J)$ subjects (the first-level units) nested within J groups (the second-level units). Let $\mathbf{y}_j$ denote an N_j-component vector of observations on the dependent variable for the j-th group, and let $\mathbf{X}_j$ denote an N_j by P matrix of the P first-level predictor variables for the j-th group. Although this is not an absolute requirement, we assume, for simplicity, that the number of the first-level predictor variables, P, is the same across all J groups. Let $\mathbf{w}_j$ denote an M-component vector of the second-level predictor variables for the j-th group. Then, the first-level model of the two-level HLM can be written, for a specific group j, as

$$\mathbf{y}_j = \mathbf{1}_{N_j}\mathrm{b}_{0j} + \mathbf{X}_j\mathbf{b}_{1j} + \mathbf{e}_j, \tag{1}$$

where $\mathbf{1}_{N_j}$ is an N_j-component vector of ones, b_{0j} is the intercept parameter, $\mathbf{b}_{1j}$ is the vector of slope parameters, and $\mathbf{e}_j$ is the vector of disturbance terms. Subscripts 0 and 1 are used to distinguish between the intercept and the slope parameters for which separate second-level models are postulated:

$$\mathrm{b}_{0j} = \mathrm{c}_{00} + \mathbf{c}'_{01}\mathbf{w}_j + u_{0j} \quad (\text{or } \ \mathrm{b}_{0j} = \mathrm{c}_{00} + \mathbf{w}'_j\mathbf{c}_{01} + u_{0j}), \tag{2}$$

and

$$\mathbf{b}_{1j} = \mathbf{c}_{10} + \mathbf{C}'_{11}\mathbf{w}_j + \mathbf{u}_{1j} \quad (\text{or } \ \mathbf{b}_{1j} = \mathbf{c}_{10} + (\mathbf{I}_P \otimes \mathbf{w}'_j)\mathbf{c}_{11} + \mathbf{u}_{1j}), \tag{3}$$

where c's are the second-level regression parameters, u's are the second-level disturbance terms, $\mathbf{I}_P$ is the identity matrix of order P, $\mathbf{c}_{11} = \mathrm{vec}(\mathbf{C}'_{11})$, and $\otimes$ indicates a Kronecker product. Substituting (2) and (3) for b_{0j} and $\mathbf{b}_{1j}$ in (1) leads to

$$\mathbf{y}_j = \mathbf{1}_{N_j}(\mathrm{c}_{00} + \mathbf{w}'_j\mathbf{c}_{01} + u_{0j}) + \mathbf{X}_j(\mathbf{c}_{10} + (\mathbf{I}_p \otimes \mathbf{w}'_j)\mathbf{c}_{11} + \mathbf{u}_{1j}) + \mathbf{e}_j. \tag{4}$$

Let $\mathbf{y}$, $\mathbf{1}_N$, $\mathbf{u}_0$, $\mathbf{u}_1$, and $\mathbf{e}$ be super-vectors of $\mathbf{y}_j$, $\mathbf{1}_{N_j}$, u_{0j}, $\mathbf{u}_{1j}$, and $\mathbf{e}_j$ $(j = 1, \dots, J)$, respectively. Let $\mathbf{G}$ and $\mathbf{D}_X$ denote block diagonal matrices with $\mathbf{1}_{N_j}$ and $\mathbf{X}_j$ as the j-th diagonal blocks, and let

$$\mathbf{W}'_0 = (\mathbf{w}_1, \dots, \mathbf{w}_J), \tag{5}$$

and

$$\mathbf{W}_1' = (\mathbf{I}_P \otimes \mathbf{w}_1, \ldots, \mathbf{I}_P \otimes \mathbf{w}_J). \qquad (6)$$

Then, the model for all observations for all J groups can be written as

$$\mathbf{y} = \mathbf{1}_N \mathrm{c}_{00} + \mathbf{G}\mathbf{W}_0\mathbf{c}_{01} + \mathbf{G}\mathbf{u}_0 + \mathbf{X}\mathbf{c}_{10} + \mathbf{D}_X\mathbf{W}_1\mathbf{c}_{11} + \mathbf{D}_X\mathbf{u}_1 + \mathbf{e}, \qquad (7)$$

where $\mathbf{1}_N = \mathbf{G}\mathbf{1}_J$, and $\mathbf{X} = \mathbf{D}_X(\mathbf{1}_J \otimes \mathbf{I}_P)$.

The above model is not identified. To remove redundancies in the model we successively make the seven terms in the model mutually orthogonal. The model then decomposes the data vector $\mathbf{y}$ into seven mutually orthogonal components with each component having specific interpretation. The first term in (7) is the intercept term. The next two terms represent between-groups effects, of which the second term pertains to the portions of the between-groups effects that can be accounted for by the second-level predictor variables ($\mathbf{w}_j$'s), and the third term to the portions left unaccounted for by the second-level predictor variables. The remaining four terms represent within-groups effects, the first one of which (the fourth term in (7)) pertains to the main effects of the first-level predictor variables ($\mathbf{X}$), the next (the fifth term) to the within-groups interactions between the first- and the second-level predictor variables, and the sixth term represents the portions of the interaction effects between groups and the second-level predictor variables left unaccounted for by the fourth and the fifth terms. The last term represents the within-groups effects that cannot be explained by any systematic effects in the model.

2.2 Estimation

Since the model (7) is linear in parameters assumed to be of the fixed-effects type, least squares (LS) estimates of parameters (c's and u's) can be obtained in a straightforward manner. In what follows, SS_i denotes the sum of squares (SS) accounted for by the i^{th} term in (7). We put a hat on a symbol to indicate a least squares estimate. Below, where the regular inverse cannot be taken, it may be replaced by the Moore-Penrose inverse.

We fit one term at a time sequentially. The remaining terms are orthogonalized to all previously fitted terms. Let $\bar{y} = \mathbf{1}_N'\mathbf{y}/N$, the grand mean. Then,

$$\hat{\mathrm{c}}_{00} = \bar{y}. \qquad (8)$$

$\mathrm{SS}_1 = N\bar{y}^2$. We define $\mathrm{SS}_t = \mathbf{y}'\mathbf{y} - \mathrm{SS}_1$ (the total SS). To make the second term orthogonal to the first, we redefine the second term as $\mathbf{G}\tilde{\mathbf{W}}_0\mathbf{c}_{01}$, where $\tilde{\mathbf{W}}_0$ denote the columnwise centered $\mathbf{W}_0$. For later use, we also define the vector of deviation scores from the grand mean, $\mathbf{y}^* = \mathbf{y} - \mathbf{1}_N\bar{y}$. This vector represents the portions of $\mathbf{y}$ left unaccounted for by the grand mean. Then,

$$\hat{\mathbf{c}}_{01} = (\tilde{\mathbf{W}}_0'\mathbf{G}'\mathbf{G}\tilde{\mathbf{W}}_0)^{-1}\tilde{\mathbf{W}}_0'\mathbf{G}'\mathbf{y}^*. \qquad (9)$$

SS_2 is obtained by $\mathrm{SS}_2 = \mathbf{y}^{*\prime}\mathbf{G}\tilde{\mathbf{W}}_0\hat{\mathrm{c}}_{01}$. Define $\bar{\mathbf{y}}^* = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{y}^*$. This is the vector of group means in the form of deviations from the grand mean. Then,

$$\hat{\mathbf{u}}_0 = \bar{\mathbf{y}}^* - \tilde{\mathbf{W}}_0\hat{\mathbf{c}}_{01}. \qquad (10)$$

SS_3 is obtained by $\mathrm{SS}_3 = \mathbf{y}^{*\prime}\mathbf{G}(\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'\mathbf{y}^* - \mathrm{SS}_2$. We define $\mathrm{SS}_b = \mathrm{SS}_2 + \mathrm{SS}_3$ (the between-groups SS). Let $\mathbf{y}^{(w)} = \mathbf{y}^* - \mathbf{G}\bar{\mathbf{y}}^*$. which is the vector of deviation scores from the group means and represents the portions of $\mathbf{y}^*$ (or equivalently, the portions of $\mathbf{y}$) left unaccounted for by $\mathbf{G}$. We columnwise center $\mathbf{X}_j$ within each group and denote it by $\mathbf{X}_j^*$. We then form a supermatrix of $\mathbf{X}_j^*$'s by putting them columnwise and denote it by $\mathbf{X}^*$. Then,

$$\hat{\mathbf{c}}_{10} = (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}\mathbf{y}^{(w)}. \qquad (11)$$

SS_4 is given by $\mathrm{SS}_4 = \mathbf{y}^{(w)\prime}\mathbf{X}^*\hat{\mathbf{c}}_{10}$. Let $\mathbf{D}_{X^*}$ denote a block diagonal matrix with $\mathbf{X}_j^*$'s as diagonal blocks. To make the fifth term orthogonal to all the previous terms, we redefine it as $\mathbf{D}_{\mathbf{X}^*}\tilde{\mathbf{W}}_1\mathbf{c}_{11}$, where $\tilde{\mathbf{W}}_1 = \mathbf{W}_1 - (\mathbf{1}_J \otimes \mathbf{I}_P)(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\mathbf{D}_{X^*}\mathbf{W}_1$. Then,

$$\hat{\mathbf{c}}_{11} = (\tilde{\mathbf{W}}_1'\mathbf{D}_{X^*}'\mathbf{D}_{X^*}\tilde{\mathbf{W}}_1)^{-1}\tilde{\mathbf{W}}_1'\mathbf{D}_{X^*}'\mathbf{y}^{(w)}. \qquad (12)$$

SS_5 is then given by $\mathrm{SS}_5 = \mathbf{y}^{(w)\prime}\mathbf{D}_{X^*}\tilde{\mathbf{W}}_1\hat{\mathbf{c}}_{11}$. Let $\mathbf{y}^{(w)*} = (\mathbf{D}_{X^*}'\mathbf{D}_{X^*})^{-1} \times \mathbf{D}_{X^*}'\mathbf{y}^{(w)}$. Then,

$$\hat{\mathbf{u}}_1 = \mathbf{y}^{(w)*} - (\mathbf{1}_J \otimes \mathbf{I}_P)\hat{\mathbf{c}}_{10} - \tilde{\mathbf{W}}_1\hat{\mathbf{c}}_{11}. \qquad (13)$$

SS_6 is obtained by $\mathbf{y}^{(w)\prime}\mathbf{D}_{X^*}\hat{\mathbf{u}}_1$. Finally,

$$\hat{\mathbf{e}} = \mathbf{y}^{(w)} - \mathbf{D}_{X^*}\hat{\mathbf{u}}_1. \qquad (14)$$

This vector represents what's left unaccounted for by all the systematic effects in the model. Let $\mathrm{SS}_w = \mathbf{y}^{(w)\prime}\mathbf{y}^{(w)}$ (the within-groups SS). Then, $\mathrm{SS}_7 = \mathrm{SS}_w - (\mathrm{SS}_4 + \mathrm{SS}_5 + \mathrm{SS}_6)$, and $\mathrm{SS}_t = \mathrm{SS}_b + \mathrm{SS}_w$.

2.3 Dimension reduction

There are several kinds of c parameters. They are c_{00} (1×1), $\mathbf{c}_{01}$ $(M \times 1)$, $\mathbf{c}_{10}$ $(P \times 1)$, and $\mathbf{c}_{11}$ $(PM \times 1)$. None of them depend on j. The $\mathbf{c}_{11}$ represents regression coefficients for the interactions between the first- and the second-level predictor variables. Its estimate may be arranged in the form of $\mathbf{C}_{11}'$ $(P \times M)$ and may be subjected to rank reduction, which is done by generalized singular value decomposition (GSVD) of $\hat{\mathbf{C}}_{11}'$ (see, for example, Takane & Shibayama, 1991). That is,

$$\hat{\mathbf{C}}_{11}' = \mathbf{TDV}', \qquad (15)$$

where $\mathbf{T}$ and $\mathbf{V}$ satisfy $\mathbf{T}'\mathbf{MT} = \mathbf{I}$ and $\mathbf{V}'\mathbf{NV} = \mathbf{I}$ and $\mathbf{D}$ is diagonal and positive-definite (*pd*). Matrices $\mathbf{M}$ and $\mathbf{N}$ are called metric matrices, assumed *pd.* Appropriate metric matrices for GSVD of $\hat{\mathbf{C}}_{11}'$ are $\mathbf{M} = \mathbf{X}^{*\prime}\mathbf{X}^{*}$ on the row side and $\mathbf{N} = \tilde{\mathbf{W}}_0'\tilde{\mathbf{W}}_0$ on the column side. The GSVD of $\hat{\mathbf{C}}_{11}'$ with metrics $\mathbf{M}$ and $\mathbf{N}$ can be obtained by ordinary SVD of $\mathbf{R}_M'\hat{\mathbf{C}}_{11}'\mathbf{R}_N$, where $\mathbf{R}_M$ and $\mathbf{R}_N$ are arbitrary square root factors of $\mathbf{M}$ and $\mathbf{N}$, respectively. Let $\mathbf{R}_M'\hat{\mathbf{C}}_{11}'\mathbf{R}_N = \mathbf{T}^*\mathbf{D}^*\mathbf{V}^{*\prime}$ be the ordinary SVD of $\mathbf{R}_M'\hat{\mathbf{C}}_{11}'\mathbf{R}_N$. Then, the desired GSVD is obtained by $\mathbf{T} = (\mathbf{R}_M')^{-1}\mathbf{T}^*$, $\mathbf{V} = (\mathbf{R}_N')^{-1}\mathbf{V}^*$, and $\mathbf{D} = \mathbf{D}^*$. Matrix $\hat{\mathbf{C}}_{11}'$ may be appended by $\hat{c}_{00}$, $\hat{\mathbf{c}}_{01}$, and $\hat{\mathbf{c}}_{10}$ to form a super-matrix

$$\hat{\mathbf{C}} = \begin{bmatrix} \hat{c}_{00} & \hat{\mathbf{c}}_{01}' \\ \hat{\mathbf{c}}_{10} & \hat{\mathbf{C}}_{11}' \end{bmatrix}. \tag{16}$$

This augmented matrix may be subjected to GSVD with metrics $\mathbf{M} = [\mathbf{1}_N, \mathbf{X}^*]'[\mathbf{1}_N, \mathbf{X}^*]$ and $\mathbf{N} = [\mathbf{1}_J, \tilde{\mathbf{W}}_0]'[\mathbf{1}_J, \tilde{\mathbf{W}}_0]$. According to our experience, however, this procedure tends to facilitate the mean tendency to dominate the solution.

The u parameters are group specific. Still, one my define a P by J matrix (assuming that P remains the same across the J groups),

$$\hat{\mathbf{U}}_1 = [(\mathbf{X}_1^{*\prime}\mathbf{X}_1^*/N_1)^{1/2}\mathbf{u}_1, \ldots, (\mathbf{X}_J^{*\prime}\mathbf{X}_J^*/N_J)^{1/2}\mathbf{u}_J],$$

which may be subjected to a rank reduction by ordinary SVD. Again, $\hat{\mathbf{u}}_0'$ may be appended to $\hat{\mathbf{U}}_1$ to form a super-matrix, which may be subject to a joint rank reduction.

When N_j is constant across all J groups (say, $N_j = n$ for all j), $\hat{\mathbf{e}}$ may also be rearranged into a J by n matrix $\hat{\mathbf{E}}$, which may be subjected to a rank reduction by SVD. This may be considered a kind of error analysis, which often helps detect which crucial factors are missing in the fitted model.

3. An illustrative example

For illustration we use part of the data from the British Social Attitudes Panel Survey, 1983-1986 (Wiggins, Ashworth, O'Muircheartaigh, & Galbraith, 1990) on attitudes toward abortion. In a panel survey, subjects are asked to respond to the same set of questions on several occasions. This allows the stability of responses over time to be assessed, while allowing any changes in the responses to be linked to the individual characteristics of subjects.

Two hundred sixty four subjects were each interviewed four times approximately one year apart on their attitudes toward abortion. The exact format of the questions was as follows:

Here are a number of circumstances in which a woman might consider an abortion. Please say whether or not you think the law should allow an abortion in each case. Should abortion be allowed by law?

(1) The woman decides on her own she does not wish to have the child.

(2) The couple agree they do not wish to have the child.
(3) The woman is not married and does not wish to marry the man.
(4) The couple cannot afford any more children.
(5) There is a strong chance of a defect in the baby.
(6) The woman's health is seriously endangered by the pregnancy.
(7) The woman became pregnant as a result of rape.

The number of items endorsed by each subject was counted and used as a measure of his/her favorableness toward abortion. The repeated measurements within subjects ($N_j = 4$) were taken as the first-level units, and the subjects ($J = 264$) were taken as the second-level units.

The second-level predictor variables used were:

1. Gender: 1) male, or 2) female.
2. Age classified into five groups: 1) up to 29 years of age, 2) up to 39, 3) up to 49, 4) up to 59, or 5) 60 and over.
3. Religion: 1) catholic, 2) protestant, 3) other, or 4) no religion.

These variables describe subject characteristics which were assumed stable throughout the four-year period of study. For our analysis, they were coded into 11 dummy variables.

The first-level predictor variables used were:

1. Year of measurements: 1) 1983, 2) 1984, 3) 1985, or 4) 1986.
2. Political party: 1) conservative, 2) labour, 3) liberal, 4) other, or 5) none.
3. Self-assessed social class: 1) middle class, 2) upper working class, or 3) lower working class.

These variables pertain to the repeated measurements within subjects and were considered time-variant. They were dummy-coded into 12 variables as in the case of the second-level predictor variables. (Both sets of predictor variables are linearly dependent, and consequently, the Moore-Penrose inverses were used to obtain LS estimates, where necessary.)

Table 1 gives the breakdown of the total SS (SS_t) explained by different terms in model (7).

Table 1. The breakdown of the total SS.

Source	% SS	SS	df	MS	F
SS_2	8.5%	304	8	38.0	4.5[1]
SS_3	60.0%	2134	255	8.4	
SS_4	2.8%	98	9	10.9	8.4[2]
SS_5	2.9%	103	72	1.4	1.1[2]
SS_6	25.8%	920	711	1.3	
SS_7	0.0%	0	0		

[1] Against SS_3.
[2] Against SS_6.

SS_2 represents the portions of SS_t accounted for by the main effects of the second-level predictor variables. SS_4 represents the portions accounted for by the main effect of the first-level predictor variables, and SS_5 represents the portions that can be accounted for by the interaction between the first- and the second-level predictor variables. The proportions of SS_t that can be accounted for by these effects only add up to 14.2%. However, tested against SS_3, SS_2 is statistically significant ($p < .01$), and tested against SS_6, SS_4 is significant ($p < .01$), although SS_5 is not significant ($p > .05$). SS_3 represents the between-subjects SS left unaccounted for by the second-level predictor variables, while SS_6 represents the SS due to the subjects by first-level predictor variables interaction effects left unaccounted for by SS_4 and SS_5. These two SS's account for 85.8% of SS_t. SS_7 is equal to zero in the present case, because the number of the first-level predictor variables is larger than the number of repeated measurements per subject, and the interaction between the subjects and the first-level predictor variables captures all the within-subjects effects.

The following tables give estimates of regression coefficients for the main effect of the second-level (Table 2) and that of the first-level (Table 3) predictor variables:

Table 2. Estimate of $\mathbf{c}_{01}$.

Plotting Labels	Variable Categories	Estimates $\hat{\mathbf{c}}_{01}$
1. Gender		
w_{11}	male	.066
w_{12}	female	-.066
2. Age		
w_{21}	˜29 yrs.	-.106
w_{22}	˜39 yrs.	.294
w_{23}	˜49 yrs.	-.206
w_{24}	˜59 yrs.	.109
w_{25}	60˜ yrs.	-.091
3. Religion		
w_{31}	catholic	-.778
w_{32}	protestant	.323
w_{33}	other	-.333
w_{34}	no religion	.788

Table 3. Estimate of $\mathbf{c}_{10}$.

Plotting Labels	Variable Categories	Estimates $\hat{\mathbf{c}}_{10}$
1. Year		
x_{11}	1983	,158
x_{12}	1984	-.488
x_{13}	1985	.017
x_{14}	1986	.313
2. Political Party		
x_{21}	conserv.	.006
x_{22}	labour	.167
x_{23}	liberal	-.101
x_{24}	other	.064
x_{25}	none	-.135
3. Social Class		
x_{31}	middle	-.069
x_{32}	upper w.	-.099
x_{33}	lower w.	.168

By eliminating one predictor variable at a time from the model and refitting the reduced model, we can assess the unique contribution of that variable. We found that the religion was the only variable which was statistically significant among the second-level predictor variables ($SS_{Religion} = 227.4$ with 3 df; $F(3, 255) = 9.0$). Estimated coefficients in Table 2 indicate that catholics are least favorable, and those without any religious affiliation tend to be most favorable to abortion. Somewhat unexpectedly, there was little gender difference ($SS_{Gender} = 4.3$ with 1 df; $F(4, 255) = 1.1$), and there was little variation over age ($SS_{Age} = 36.1$ with 4 df; $F(4, 255) = .5$). Among the first-level predictor variables, we found the year of measurements was the only variable significantly affecting the attitude toward abortion ($SS_{Year} = 93.2$ with 3 df; $F(3, 711) = 23.9$). Table 3 indicates that people were on average less favorable toward abortion in 1984, and more favorable in 1986, although the reason for this is not readily apparent. However, there was little systematic trend over time. Neither political alignment ($SS_{Party} = 2.6$ with 4 df; $F(4, 711) = .5$) nor self-assessed social class ($SS_{Class} = 4.5$ with 2 df; $F(2, 711) = 1.7$) had substantial effects on attitude toward abortion.

As has been noted earlier, SS_5 representing the overall size of the interaction effects between the first- and the second-level predictor variables was not statistically significant. Still, some portions of the interaction effects may be significant. A strategy used above for testing the significance of the main effects is a bit cumbersome because there are so many of them (132), which are also partially redundant. This is where dimensionality reduction may help. We can simplify the pattern of the interaction effects by applying SVD to $\hat{\mathbf{C}}_{11}$. Figure 1 displays the two-dimensional configuration resulting from a reduced-rank approximation to $\hat{\mathbf{C}}_{11}$. The two dimensions account for approximately 75% of the sum of squares pertaining to the interaction effects. Plotting symbols are given in the first columns of Tables 2 and 3. (The first-level predictor variables are indicated by "x", and the second-level variables by "w", followed by the variable number and the category number.) Since what is analyzed here is the weights applied to the interaction effects between the first- and the second-level predictor variables, we look for 1) which variables come in similar directions relative to the origin (indicated by "+"), and 2) which variables come in the opposite directions. For example, w23 (protestant) and x14 (year 1986) come close to each other, indicating that protestants in the year 1986 favored abortion more than that can be expected from separately being protestants (who tend to be more favorable to abortion than average) and that the year the data were taken was 1986 (the year that people tend to be more favorable to abortion than the average year). On the other hand, w31 (catholic) and x25 (no party support) come on the opposite side. This means that catholics with no party support were even less favorable to abortion than that was expected from separately being catholics (who tend to be less favorable to abortion than average) and that the subjects had no political parties to support (who were also less favorable to abortion than average). We can make many other similar observations.

We also analyzed $\hat{\mathbf{U}}_1$ as suggested above. However, this relates to the subject differences left unaccounted for by those already taken into account in the model, and interpretation was extremely difficult without additional information about the subjects.

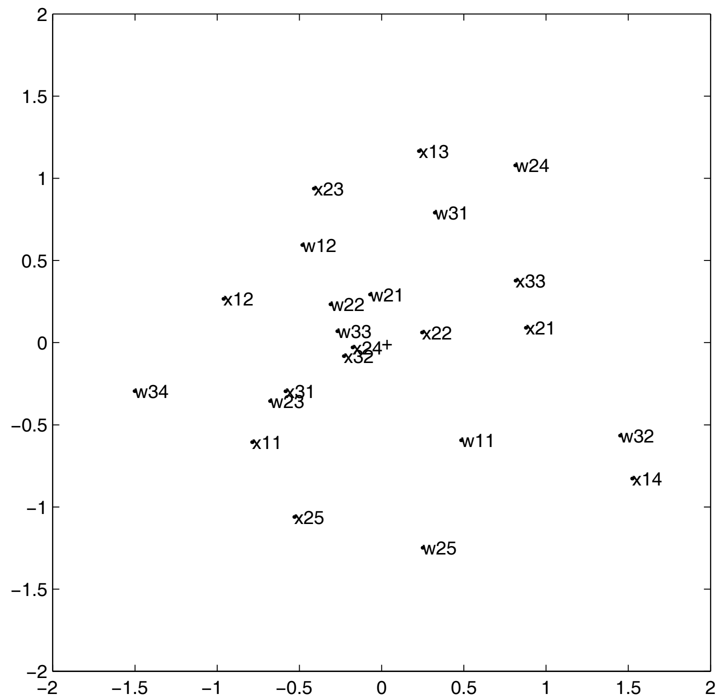

Figure 1: The two-dimensional configuration of the interaction effects.

4. Discussion

Reduced-rank approximations of some of the regression parameters in HLM seem useful, particularly when the dimensionality of the parameter space is very high. There are a number of things that may be done to further facilitate the use of HLM:

1. Although only the two-level HLM has been discussed in this paper, similar things can be done for higher-level HLM, although the list of terms in the model gets longer very quickly. For example, there are 15 terms in the three-level HLM. We have already derived LS estimates of parameters associated with these terms, although, due to the space limitation, they cannot be presented here.

2. Bootstrap (e.g., Efron, 1982) or other resampling techniques could be used to assess the stability of individual parameters, which may in turn be used to test their significance. Since the normality assumption is almost always in suspect in survey data, the resampling methods may also be useful to benchmark the distribution of the conventional statistics used in HLM. The number of components to be retained in the reduced-rank approximation may be determined by permutation tests in a manner similar to Takane

& Hwang (in press) who developed a permutation procedure for testing the number of significant canonical correlations.

3. Additional (linear) constraints can be readily incorporated in the LS estimation procedure. This allows the statistical tests of the hypotheses represented by the constraints.

4. We exclusively discussed the univariate HLM in this paper. However, the proposed method can easily be extended to the multivariate cases. We may incorporate structures representing the hypothesized relationships among the multiple dependent variables or use the dimension reduction technique to structure the multiple dependent variables.

5. When the u parameters are assumed to be random rather than fixed, errors are no longer statistically independent. The dependence structure among the errors may be estimated from the initial estimates of parameters (obtained under the independence assumption), which may then be used to re-estimate the parameters, and so on. This leads to an iterative estimation procedure similar to that used in Hwang & Takane (2001) in the context of structural equation modeling. The derived estimates are more efficient than those obtained under the independence assumption.

Acknowledgments

The work reported in this paper has been supported in part by NSERC individual operating grants (A6394) to the first author.

References:

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, **22**, 327-351.

Bock, R. D. (1989). *Multilevel analysis of educational data*. San Diego, CA: Academic Press.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage Publications.

Efron, B. (1982). *The Jackknife, the bootstrap and other resampling plans*. Philadelphia: SIAM.

Hwang, H., & Takane, Y.(2001). Structural equation modeling by extended redundancy analysis. This volume.

Goldstein, H. I. (1987). *Multilevel models in educational and social research*. London: Oxford University Press.

Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam: TT-Publikaties.

Takane, Y., & Hwang, H. (in press). Generalized constrained canonical correlation analysis. *Multivariate Behavioral Research*.

Takane, Y., & Shibayama, T. (1991). Principal component analysis with external information on both subjects and variables. *Psychometrika*, **56**, 97-120.