
The Effect of Different Forms of Centering in Hierarchical Linear Models

Ita G. G. Kreft
California State University, Los Angeles

Jan de Leeuw
University of California, Los Angeles

Leona S. Aiken
Arizona State University

Author note: Correspondence should be sent to Ita G.G. Kreft, Division of Educational Foundations and Interdivisional Studies, School of Education, California State University, 5151 State University Drive, Los Angeles, CA 90032. We thank Rod McDonald and John Tukey for very helpful discussions on the topics treated in this paper. A slightly different version of this paper will appear in the first 1995 issue of Multivariate Behavioral Research. Research of the second author supported through the National Sciences Foundation Grant No. DMS-9208758.

ABSTRACT

Multilevel models are becoming increasingly used in applied educational social and econometric research for the analysis of hierarchically nested data. In these random coefficient regression models the parameters are allowed to differ over the groups in which the observations are nested. For computational ease in deriving parameter estimates, predictors are often centered around the mean. In nested or grouped data the option of centering around the grand mean is extended with an option to center within groups or contexts. Both are statistically sound ways to improve parameter estimation. In this paper we study the effects of these two different ways of centering, in comparison to the use of raw scores, on the parameter estimates in random coefficient models. The conclusion is that centering around the group mean amounts to fitting a different model from that obtained by centering around the grand mean or by using raw scores. The choice between the two options for centering can only be made on a theoretical basis. Based on this study we conclude that centering rules valid for simple models, such as the fixed coefficients regression model, are no longer applicable to more complicated models, such as the random coefficient model. We think researchers should be made aware of the consequences of the choice of particular centering options.

INTRODUCTION

Multilevel models are becoming increasingly used in applied educational and econometric research (see Bock, 1989, Bryk and Raudenbush, 1992, and Goldstein, 1987) to analyze hierarchically nested data. Micro-level units, such as workers or students, are nested within macro-level units, such as industries or schools. In multilevel models separate predictors characterize the micro-level units, the individuals, and the macro-level units, the groups or contexts. The assumptions regarding the coefficients of the model depend upon the level of the predictors. The coefficients of all but the highest level predictors may be treated as random, hence the name random coefficient models, while those of the highest level are always treated as fixed. Treating a coefficient as random means that the coefficient is permitted to vary across the units at the next higher level. It also means, however, that the different values of the coefficient are interpreted as different realizations of the same random variable. This treatment of coefficients as random, instead of in the traditional way as fixed, is a result of the fact that we interpret our groups as a sample of possible groups, and we want to make inferences to the total population of such groups.

The available software packages for the analysis of hierarchically nested data (GENMOD, HLM, ML3 and VARCL) differ in the way they process the raw data. The most popular option in HLM (Bryk et al., 1988) is to center on the context mean. ML3 (Prosser et al., 1992) offers many choices, among them to center on the grand mean or the group mean. The manuals for ML3 and HLM specify as reasons for centering that it may facilitate interpretation, and is also useful to improve numerical performance of the estimation algorithm. One of the improvements mentioned in the literature (e.g. Belsley, 1991, Chapter 6) is that it removes non-essential ill-conditioning caused by the choice of origin for the regressors. In VARCL (Longford, 1990) variables are automatically replaced by their centered version in deviations from the grand mean, not the group mean. The result are reparameterized in terms of the original data, which makes the VARCL output the same as that of GENMOD (Mason et al., 1991), a program that uses raw data (see Kreft et al., 1990).

With formulas and examples we illustrate the different effects of centering, especially of group mean centering, or centering “within context”. Group mean centering in multilevel models is the topic of a recent discussion in the Multilevel Modelling Newsletter, started off by Raudenbush’s (1989a) article: “Centering” predictors in Multilevel Analysis: Choices and Consequences”. This was followed by reactions of Plewis (1989) and Longford (1989) in the next issue of the same newsletter, and by a response of Raudenbush (1989b) to his two critics. The discussion is not new (Cronbach, 1976, Burstein, 1980). The use of centering has a long history in education since Cronbach (1976) advocated such a model for the separation of student effects from school or classroom effects in fixed coefficients models. It is to this history that Raudenbush (1989a, 1989b), among others, refers, in claiming several advantages of centering within context. The high correlation between the random slope and random intercept in raw scores models is eliminated. The first level regression model does not suffer from a misspecification of the second level model if centering within context is applied (Raudenbush, 1989a, p. 10). In contextual models (where the context mean is reintroduced as a second level predictor) this context mean

might have a high correlation with the raw score. Centering within context eliminates this correlation.

Notation and Definitions

For our discussion of the effects of centering we define $X_{\bullet j}$ as the mean for group j , and $X_{\bullet\bullet}$ as the mean of all observations i over all contexts j . Three different centering approaches are used.

- 1) RAS: leaving the micro-predictors in raw score form, the predictor is X_{ij} ;
- 2) CGM: centering around the grand mean, predictor $X_{ij}^* = X_{ij} - X_{\bullet\bullet}$;
- 3) CWC: centering within context, predictor $X_{ij}^{**} = X_{ij} - X_{\bullet j}$.

Two different regression models are fitted:

- 1) a simple regression model with one single first level predictor (RAS_1 , CGM_1 , CWC_1), and
- 2) a “contextual model” with the context mean reintroduced as a second level predictor of the intercept (RAS_2 , CGM_2 , CWC_2).

Throughout this paper we will use these acronyms to distinguish the six different models. We also briefly discuss more elaborate models, which have second level predictors in the equation for the random slopes. It turns out that our algebraic techniques are general enough to cover these more complicated situations as well.

Model Equivalence

Each multilevel model defines a formula for the expected value and the variance covariance matrix of the dependent variable Y_{ij} . If two different models generate the same set of expectations and dispersions, they are equivalent. In our case we only look at expected values and dispersions, because normal distributions are the same if and only if they have the same expected values and dispersions. In non-normal models we can have identical first and second order moments, without having equivalence. Even in normal models, we can have equivalence for the expected values and non-equivalence for the dispersions.

Even if we have equivalent models, the formulas for describing the model may look different, because different parametrizations are used. Parametrizations are just systems of coordinates to describe the expectations and dispersions generated by the model. Some parametrizations can be more simple or more parsimonious than others, even if equivalent models are described. If there is a one-to-one transformation between two different parametrizations, then the models they describe are equivalent.

In Table 1 our results are summarized. The RAS model is in both cases (1) and (2) equivalent to the CGM model, but both are not equivalent to the CWC model. Thus the six combinations give rise to only four different models, which we denote by A, B, C, and D in the table. The numerical results we report can be used to verify equivalence of models. In general, it is not necessarily the case that fitting equivalent models will produce the

same parameter estimates. This also depends on the estimation procedure. Since we will be using maximum likelihood estimation throughout this paper, we can use the invariance property of maximum likelihood estimation to show that the estimates should coincide for equivalent models. Hopefully this result also extends to computer programs that implement the maximum likelihood methods.

Table 1: Six models compared

Centering Mode	Do Not Include Group Means (1)	Include Group Means (2)
RAS	$RAS_1:A$	$RAS_2:C$
CGM	$CGM_1:A$	$CGM_2:C$
CWC	$CWC_1:B$	$CWC_2:D$

Centering in Fixed Coefficients Models

In this paper we explore the effect of centering in random coefficients models. For that purpose we summarize here what we know about the effects of centering in fixed coefficients regression models in similar situations (Cronbach, 1976). We shall see that results for the effect of centering in fixed coefficients models cannot be applied directly to random coefficients models, mainly because the between group variation is defined in a much more complicated way. By making more explicit the implications of centering for random coefficients models we hope to contribute to the understanding of these models.

The reasons given by Raudenbush for using CWC_1 or CWC_2 are based on fixed coefficients considerations. The relationship between the estimates of the coefficients in RAS and CWC fixed linear models is fairly simple. For the time being we suppose both X and Y are in deviations from the grand mean. Define the coefficient for the prediction of Y from X across all data from all contexts as b_T , and make use of the fact that b_T is a composite of the between-group regression b_B estimate and the pooled within-group regression estimate b_W according to $b_T = \eta^2 b_B + (1 - \eta^2) b_W$, where η^2 is the proportion of the variance in X explained by differences between contexts. We now have the following table.

Table 2: Four fixed coefficients models compared

Centering Mode	Do Not Include Group Means (1)	Include Group Means (2)
RAS	$\hat{Y}_{ij} = \{\eta^2 b_B + (1 - \eta^2) b_W\} X_{ij}$	$\hat{Y}_{ij} = b_W X_{ij} + (b_B - b_W) X_{\bullet j}$
CWC	$\hat{Y}_{ij} = (1 - \eta^2) b_W X_{ij}$	$\hat{Y}_{ij} = b_W (X_{ij} - X_{\bullet j}) + b_B X_{\bullet j}$

Relations between coefficients in fixed coefficients models can be defined more properly now (see e.g., Duncan, Curzort, & Duncan, 1966, or Kreft, 1987). In the fixed coefficients CWC_2 model, the effect of $X_{ij} - X_{\bullet j}$, corrected for the effect of $X_{\bullet j}$ is equal to b_W , while the effect of $X_{\bullet j}$ corrected for $X_{ij} - X_{\bullet j}$ is equal to b_B . In RAS_2 the same effect for $X_{\bullet j}$, but corrected for the raw score effect of X_{ij} is equal to $(b_B - b_W)$. This shows that centering around the context mean changes the definition of the context effect (which is

the coefficient of $X_{\bullet j}$ from $b_B - b_W$ to b_B . We also see in Table 2 that RAS_2 and CWC_2 are equivalent, with CWC_2 giving the more natural parametrization. Models RAS_1 and CWC_1 are different. RAS_1 does not take the group structure into account at all, and CWC_1 shrinks b_W by multiplying it with the proportion of variance that is within-group. In fixed coefficients models using CWC_2 does indeed separate the between-group variation from the within-group variation. We shall show below that the same orthogonalization occurs in the fixed part of random coefficient models when CWC_2 is employed, but that the effect on the between-group variation in the random part is much less clear.

In the following paragraphs we will show that for random coefficients models:

- 1) CWC_1 results in a different model than RAS_1 or CGM_1 both in the fixed and the random part (compare the first column of Table 1).
- 2) The CWC_2 model is equivalent to RAS_2 and CGM_2 in the fixed coefficients, but not in the variance-covariance part (compare the second column of Table 1).
- 3) Notions based on fixed coefficients models are not directly applicable to random coefficients models. The random coefficients model is more complicated than the fixed coefficients contextual model, especially in the definition of between-groups variation.

MODELS WITHOUT SECOND-LEVEL PREDICTORS

We first formulate the simplest random coefficients models without the group means reintroduced, and consider the impact of centering on the regression parameters and the variance components in these simple random coefficients models. We discuss both an algebraic demonstration of the effect of centering and a numerical example of centering within the random coefficients model.

As we will show in this section, centering within context yields a different model from the raw score or the grand mean centering model. Parameters of the CWC_1 model are not simple transformations of those of the CGM_1 or RAS_1 model. Parameter estimates change, potentially leading to different inferences concerning the impact of micro and macro-predictors on outcomes, and the definition of the between-group variance.

The Models

We shall follow the convention of underlining random variables in our models (Hemelryk, 1966). This emphasizes the distinction between what is fixed and what is random, which is obviously critical in models of this class.

The micro-equation within context j is

$$\underline{Y}_{ij} = \underline{a}_j + \underline{b}_j X_{ij} + \underline{e}_{ij}. \quad (1.1)$$

The macro-equations, in the simplest case, are

$$\underline{a}_j = \gamma_{00} + \underline{\delta}_{0j}, \quad (1.2)$$

$$\underline{b}_j = \gamma_{10} + \underline{\delta}_{1j}. \quad (1.3)$$

In words, this means that contexts are replications of one another, drawn from a single population of contexts. This population of contexts is characterized by a single grand mean intercept γ_{00} and a single grand mean slope γ_{10} in equations (1.2) and (1.3), respectively. The slopes and intercepts of the various contexts deviate randomly from their expectations γ_{00} and γ_{10} , as reflected in $\underline{\delta}_{0j}$ and $\underline{\delta}_{1j}$. We assume the micro-level disturbance terms \underline{e}_{ij} have expected value equal to zero, variance σ^2 , and are uncorrelated with each other. The micro-level and macro-level disturbances are uncorrelated with each other. The covariance matrix of the macro-level disturbances is

$$\Omega \triangleq \begin{matrix} & \underline{\delta}_{0j} & \underline{\delta}_{1j} \\ \begin{matrix} \underline{\delta}_{0j} \\ \underline{\delta}_{1j} \end{matrix} & \begin{pmatrix} \omega_{00} & \omega_{01} \\ \omega_{10} & \omega_{11} \end{pmatrix} \end{matrix} \quad (1.4)$$

The correlation ρ between the random slopes and the random intercepts is defined by

$$\rho \triangleq \frac{\omega_{01}}{\sqrt{\omega_{00}\omega_{11}}}. \quad (2)$$

It follows from (1.1)-(1.4) that

$$\mathbf{E}(\underline{Y}_{ij}) = \gamma_{00} + \gamma_{10}X_{ij}, \quad (3.1)$$

$$\mathbf{V}(\underline{Y}_{ij}) = \omega_{00} + 2\omega_{10}X_{ij} + \omega_{11}X_{ij}^2 + \sigma^2, \quad (3.2)$$

$$\mathbf{C}(\underline{Y}_{ij}, \underline{Y}_{k\ell}) = \begin{cases} \omega_{00} + \omega_{10}(X_{ij} + X_{kj}) + \omega_{11}X_{ij}X_{kj} & \text{if } i \neq k \text{ and } j = \ell, \\ 0 & \text{otherwise.} \end{cases} \quad (3.3)$$

What we have formulated here is the RAS₁ model, with predictor X_{ij} . If we replace X_{ij} by X_{ij}^* throughout, we have the CGM₁ model, if we use X_{ij}^{**} we have CWC₁.

Algebraic Relationships Between RAS₁, CGM₁, and CWC₁

Let us first investigate equivalence of the expected values for RAS₁ and CGM₁. What we have to show is that for given γ_{00} and γ_{10} we can always find γ_{00}^* and γ_{10}^* such that $\gamma_{00} + \gamma_{10}X_{ij} = \gamma_{00}^* + \gamma_{10}^*X_{ij}^*$ for all i and j . This can be rewritten as $\gamma_{00} + \gamma_{10}X_{ij} = (\gamma_{00}^* - \gamma_{10}^*X_{..}) + \gamma_{10}^*X_{ij}$, and we see that the solution is

$$\gamma_{00} = \gamma_{00}^* - \gamma_{10}^*X_{..}, \quad (4.1)$$

$$\gamma_{10} = \gamma_{10}^*. \quad (4.2)$$

This expresses the parameters of the fixed part of RAS₁ in terms of those of CGM₁. It is easy to invert the relationship, and derive the inverse expressions. They are

$$\gamma_{00}^* = \gamma_{00} + \gamma_{10}X_{..}, \quad (4.3)$$

$$\gamma_{10}^* = \gamma_{10}. \quad (4.4)$$

Thus the fixed part (the expectations) are equivalent.

For dispersions the situation is a bit more complicated. We need to solve $\omega_{00} + 2\omega_{10}X_{ij} + \omega_{11}X_{ij}^2 = \omega_{00}^* + 2\omega_{10}^*X_{ij}^* + \omega_{11}^*(X_{ij}^*)^2$. The right-hand side is equal to $\omega_{00}^* + 2\omega_{10}^*(X_{ij} - X_{\bullet\bullet}) + \omega_{11}^*(X_{ij} - X_{\bullet\bullet})^2$, and by expanding and collecting terms we see this is equal to $(\omega_{00}^* - 2\omega_{10}^*X_{\bullet\bullet}) + 2(\omega_{10}^* - \omega_{11}^*X_{\bullet\bullet})X_{ij} + \omega_{11}^*X_{ij}^2$. This shows that we must have

$$\omega_{00} = \omega_{00}^* - 2\omega_{10}^*X_{\bullet\bullet} + \omega_{11}^*X_{\bullet\bullet}^2, \quad (5.1)$$

$$\omega_{10} = \omega_{10}^* - \omega_{11}^*X_{\bullet\bullet}, \quad (5.2)$$

$$\omega_{11} = \omega_{11}^*. \quad (5.3)$$

Again the relationship expressed in these equations can easily be inverted. For completeness we give the results.

$$\omega_{00}^* = \omega_{00} + 2\omega_{10}X_{\bullet\bullet} + \omega_{11}X_{\bullet\bullet}^2, \quad (5.4)$$

$$\omega_{10}^* = \omega_{10} + \omega_{11}X_{\bullet\bullet}, \quad (5.5)$$

$$\omega_{11}^* = \omega_{11}. \quad (5.6)$$

Thus the random parts (the dispersions) of RAS_1 and CGM_1 are equivalent as well.

We now try to follow the same method of proof to investigate equivalence of RAS_1 and CWC_1 . We need to solve $\gamma_{00} + \gamma_{10}X_{ij} = \gamma_{00}^{**} + \gamma_{10}^{**}X_{ij}^{**} = \gamma_{00}^{**} + \gamma_{10}^{**}X_{ij} - \gamma_{10}^{**}X_{\bullet j}$. This can be written as $(\gamma_{00} - \gamma_{00}^{**}) + (\gamma_{10} - \gamma_{10}^{**})X_{ij} + \gamma_{10}^{**}X_{\bullet j} = 0$. Because in general the constant vector, the vector X_{ij} , and the vector $X_{\bullet j}$ are linearly independent, this means that we must have both $\gamma_{10}^{**} = 0$ and $\gamma_{10} = 0$. Thus, in the general case, the equations cannot be solved, and the fixed parts of RAS_1 and CWC_1 are not equivalent.

For equivalence of the random parts we need to solve $\omega_{00} + 2\omega_{10}X_{ij} + \omega_{11}X_{ij}^2 = \omega_{00}^{**} + 2\omega_{10}^{**}X_{ij}^{**} + \omega_{11}^{**}(X_{ij}^{**})^2$. But again the right-hand side, after expansion, will contain terms such as $X_{\bullet j}$, $X_{\bullet j}^2$, and $X_{ij}X_{\bullet j}$ that are simply missing from the left-hand side. Thus, also for the random part, there can be no equivalence.

Numerical Examples

To illustrate our discussion and formulas we use a specific example (Kreft & De Leeuw, 1994). The data consists of 5241 employees in twelve industries. The relationship explored is that between education and income across the twelve industries. Industries are from both the public and private sectors. Examples of the former are College and University, US Military, Human Services, and Government. Examples of private sector industries are Manufacturing and Construction, Retail or Wholesale, and Commerce, Insurance, Finance or Real Estate. Employees are nested within industry. We use two employee-level (micro-level) measurements, education and income, from this dataset. Educational level is the independent variable X that predicts the income Y . Since the sample consists of a longitudinal study of university sophomores, the educational level has only four categories, from no degree to Ph.D. To correct for skewness, income (in \$ 1000) has been categorized into seven categories with midpoints ranging from 3.5 to 22.5.

We have analyzed this education-income example, using model (1), and the computer program VARCL (Longford, 1990). The solutions, with z-values for the slopes in square brackets, for respectively the RAS₁, the CGM₁, and CWC₁ models, are given in Table 3. Under the heading “Variance Components” we have collected the matrix Ω of (1.4).

Table 3: Solutions of random effects models with RAS, CGM, and CWC predictors

Model	Deviance	Equation	Variance Components
RAS ₁	24009.26	$\hat{Y}_{ij} = 8.37 + 1.01[6.08]X_{ij}$	$\begin{pmatrix} +4.26 & -0.73 \\ -0.73 & +0.22 \end{pmatrix}$
CGM ₁	24009.26	$\hat{Y}_{ij} = 11.23 + 1.01[6.08]X_{ij}^*$	$\begin{pmatrix} +1.83 & -0.12 \\ -0.12 & +0.22 \end{pmatrix}$
CWC ₁	24006.81	$\hat{Y}_{ij} = 11.23 + 1.02[6.13]X_{ij}^{**}$	$\begin{pmatrix} +1.47 & -0.11 \\ -0.11 & +0.22 \end{pmatrix}$

Comparing fixed effects (coefficients) and random effects (variance components) over the models, we see that the difference between RAS₁ and CGM₁ is only in the intercepts and the difference between CGM₁ and CWC₁ is in the slope estimates (and their corresponding z-values). Centering has an effect on the intercept and, as expected, on the variance of the intercept ω_{00} and the covariance between slope and intercept ω_{10} . The results also illustrate the computational advantage of centering. The correlation, using equation (2), between intercept and slope is reduced from -0.75 when RAS₁ is employed to -0.20 when CGM₁ is employed, and to -0.19 for CWC₁. The argument in favor of CWC₁ (Raudenbush, 1989a) is that it removes for a large part (but not totally) the confounding of slope and intercept variance. We see that in this example CGM₁ accomplishes this reduction in multicollinearity already to a large extent.

Table 3 shows that estimates of the parameters change when using data centered within context, since rescaling X_{ij} from RAS₁ from to CWC₁ form is not a simple linear transformation. The context mean varies across contexts; hence centering within context introduces a new variable $X_{ij} - X_{\bullet j}$ into the prediction scheme. In our dataset the estimate of γ_{10} for CWC₁ is 1.02, while for CGM₁ it is 1.01. This is clearly a small difference. We cannot calculate how much difference can “usually” be expected, but we can provide an illustration from Yoon (1993). She analyzed a large dataset, based on an US sample from the IEA Second International Science Study, collected by the International Association for the Evaluation of Educational Achievement (IEA, 1988). The sample consisted of 121 teachers and 2909 students. The dependent variable was science achievement. Predictors were Verbal Score as an indicator for general ability, SES and Number of Books in the home as indicators for family background, and Practical Work which describes how often students work as a group. In her Table 4-6 (p. 66) she compares the solutions given by CWC₁ and CGM₁. No second level or context variables are present in the model. The solutions for the regression coefficients are given in our Table 4.

Table 4: Solutions for CWC₁ and CGM₁ predictors for Yoon data

Model	Verbal Score	SES	# Books	Practical Work
CWC ₁	0.54	0.95	0.50	0.05
CGM ₁	0.56	1.03	0.56	0.17

The respective standard errors of the parameters in Table 4 do not differ over CWC₁ and CGM₁, and significance levels of individual parameters are not affected. The largest difference between CWC₁ and CGM₁ is in the variance of the intercept from 12.33 in the CWC₁ model to 3.54 in the CGM₁ model. In the example in Table 3 we find an intercept variance of 1.83 in CGM₁ and of 1.47 in CWC₁.

CONTEXTUAL MODELS

Contexts may well be expected to differ in arithmetic mean level on the micro-predictors, e.g., in mean educational level from industry to industry. The strategy is sometimes adopted of modeling the differences among context means at the macro-level by creating a macro-level predictor that captures the differences in arithmetic mean levels over contexts. Such models are referred to as contextual models (Burstein, 1980). The macro-predictor we use is simply $X_{\bullet j}$. We could also use $X_{\bullet j} - X_{\bullet\bullet}$, but it is easy to see that this leads to equivalent models, and indeed identical results in the computation, except for the parametrization of the fixed coefficients. In this section we first explore numerically the impact of using the macro-level predictors, in combination with the three scalings of the micro-predictors we have considered.

Contextual models, which use means of micro-level variables as macro-predictors, are more general than models with no macro-predictors. Given a contextual model with the predictor $X_{\bullet j}$ introduced at the macro level, we examine how the choice of the micro predictor as X_{ij} or X_{ij}^* or X_{ij}^{**} affects parameter estimates.

Algebraic Comparison of Contextual Models

The micro-model is

$$\underline{Y}_{ij} = \underline{a}_j + \underline{b}_j X_{ij} + e_{ij} \quad (6.1)$$

as before, but the macro-model is

$$\underline{a}_j = \gamma_{00} + \gamma_{01} X_{\bullet j} + \underline{\delta}_{0j}, \quad (6.2)$$

$$\underline{b}_j = \gamma_{10} + \underline{\delta}_{1j}. \quad (6.3)$$

Of course this is the RAS₂ model, the CGM₂ model and the CWC₂ model are obtained by replacing X_{ij} in (6.1) by X_{ij}^* or X_{ij}^{**} .

For the expected values and dispersions we find

$$\mathbf{E}(\underline{Y}_{ij}) = \gamma_{00} + \gamma_{01} X_{\bullet j} + \gamma_{10} X_{ij}, \quad (7.1)$$

$$\mathbf{V}(\underline{Y}_{ij}) = \omega_{00} + 2\omega_{10}X_{ij} + \omega_{11}X_{ij}^2 + \sigma^2, \quad (7.2)$$

$$\mathbf{C}(\underline{Y}_{ij}, \underline{Y}_{k\ell}) = \begin{cases} \omega_{00} + \omega_{10}(X_{ij} + X_{k\ell}) + \omega_{11}X_{ij}X_{k\ell} & \text{if } i \neq k \text{ and } j = \ell, \\ 0 & \text{otherwise.} \end{cases} \quad (7.3)$$

Expressions (7.2) and (7.3) do not involve the context means $X_{\bullet j}$, and are exactly the same as (3.2) and (3.3). As a consequence the results about the equivalence of models from the previous section apply directly. As far as the variances are concerned, CGM₂ is equivalent to RAS₂, while CWC₂ is not equivalent.

The results for the expected values are different from the previous section. Using the star-notation, we see that for equivalence we must have $\gamma_{00} + \gamma_{01}X_{\bullet j} + \gamma_{10}X_{ij} = \gamma_{00}^* + \gamma_{01}^*X_{\bullet j} + \gamma_{10}^*X_{ij}^* = \gamma_{00}^{**} + \gamma_{01}^{**}X_{\bullet j} + \gamma_{10}^{**}X_{ij}^{**}$. By collecting terms we see that this can be written as $\gamma_{00} + \gamma_{01}X_{\bullet j} + \gamma_{10}X_{ij} = (\gamma_{00}^* + \gamma_{10}^*X_{\bullet\bullet}) + \gamma_{01}^*X_{\bullet j} + \gamma_{10}^*X_{ij} = \gamma_{00}^{**} + (\gamma_{01}^* - \gamma_{10}^{**})X_{\bullet j} + \gamma_{10}^{**}X_{ij}$. Thus the solution is

$$\gamma_{00} = \gamma_{00}^* + \gamma_{10}^*X_{\bullet\bullet} = \gamma_{00}^{**}, \quad (8.1)$$

$$\gamma_{01} = \gamma_{01}^* = \gamma_{01}^{**} - \gamma_{10}^{**}, \quad (8.2)$$

$$\gamma_{10} = \gamma_{10}^* = \gamma_{10}^{**}. \quad (8.3)$$

Expressing one set in terms of one of the other two sets turns out to be easy. Thus the three models are equivalent in terms of the expected values.

Numerical example of Contextual Models

In Table 5 the three models RAS₂, CGM₂ and CWC₂ are compared. Slopes and intercepts vary randomly as in the case previously considered. The coefficient for the context variable is fixed, and is an additional parameter in the model.

Table 5: The Contextual Model with RAS₂, CGM₂, and CWC₂

Model	Deviance	Equation	Variance Components
RAS ₂	23997.36	$\hat{Y}_{ij} = 29.13 + 1.04[6.63]X_{ij} - 7.44[-5.10]X_{\bullet j}$	$\begin{pmatrix} +1.05 & -0.31 \\ -0.31 & +0.18 \end{pmatrix}$
CGM ₂	23997.36	$\hat{Y}_{ij} = 32.07 + 1.04[6.64]X_{ij}^* - 7.44[-5.10]X_{\bullet j}$	$\begin{pmatrix} +0.74 & +0.20 \\ +0.20 & +0.18 \end{pmatrix}$
CWC ₂	23997.08	$\hat{Y}_{ij} = 28.98 + 1.04[6.64]X_{ij}^{**} - 6.34[-4.52]X_{\bullet j}$	$\begin{pmatrix} +0.69 & +0.19 \\ +0.19 & +0.18 \end{pmatrix}$

If we compare RAS₁ in Table 3 and RAS₂ in Table 5, variance component ω_{00} is reduced from 4.26 to 1.05, variance component ω_{11} from 0.22 to 0.18, and covariance component ω_{01} goes from -0.73 to -0.31 , compared to the model (1) without the macro-level variable. The correlation between the macro-level disturbances goes from -0.75 to -0.71 . Again individual education X_{ij} has a positive effect on income within industries. However, we see an opposite effect of mean education per industry on individual income. The negative coefficient of $X_{\bullet j}$ shows that the higher the mean educational level of an industry the lower

the mean income after correcting for individual educational effects. The best educated industry overall (i.e. College and University) has a lower mean income than other less well educated industries. This example highlights the need to differentiate micro-level and macro-level effects in multilevel data. It also shows that the between-group variation is no longer captured in a single between-component. Even in this simple model we have three sources of between-industry variation, the three elements of Ω .

It has been shown that parameter estimates from a CWC₂ model that contains context means as a macro-predictor can be recalculated to the estimates of the RAS₂ model. The equation (8.2) is the analogy of the fixed coefficients equation in Table 2, which says that the coefficient of $X_{\bullet j}$ in RAS₂ is $b_B - b_W$. But γ_{01}^{**} and γ_{10}^{**} are no longer the simple ratios based on the orthogonal variance partitioning that obtains in the fixed coefficient model.

If $X_{\bullet j}$ is the macro predictor, the estimates of the coefficient γ_{01} for this context effect are different when CWC₂ is employed from when RAS₂ or CGM₂ is employed. Models containing raw score predictor X_{ij} or grand mean centered predictor $(X_{ij} - X_{\bullet\bullet})$ show again that they produce identical parameter estimates, with the usual exception for the values related to the intercept. The discrepancy is in the macro level parameters γ_{01} over models, where the value for γ_{01} in the first two models (RS₂ and CGM₂) is -7.44 ($z = 5.10$), while this value in the CWC₂ model is -6.34 ($z = 4.52$).

In the previous section, which examined relationships between the RAS₁, CGM₁, and the CWC₁ model, it was indicated that no simple linear transformation of the RAS₁ model parameters could produce the CWC₁ model parameters. The reason was straightforward: X_{ij} as a raw score micro-predictor contained any variation in context means $X_{\bullet j}$ over contexts, while this source of variation is removed from the micro-predictor $X_{ij} - X_{\bullet j}$ for CWC₁ analysis. In this section the variation among context means $X_{\bullet j}$ is reintroduced at the macro-level, by using $X_{\bullet j}$ either or $X_{\bullet j} - X_{\bullet\bullet}$ as the macro-predictor. The model with $X_{ij} - X_{\bullet j}$ as micro-predictor and $X_{\bullet j}$ or $X_{\bullet j} - X_{\bullet\bullet}$ as the macro-predictor contains all the variation as it is contained in the RAS₂ (or CGM₂) model with context mean, only partitioned differently over the parameters. For these contextual models there are again simple rules to find the RAS₂ estimates of the fixed coefficients from the CWC₂ estimates. But the contextual models, CWC₂, CGM₂ and RAS₂ no longer describe the same family of distributions: both generate the same set of means but not the same set of dispersions.

Second level variables that have no individual component (so called global variables, see Lazarsfeld and Menzel, 1961) are, as a result of the difference in variances and covariances, also estimated differently over CWC₂ versus RAS₂/CGM₂. In the analysis of the IEA data (Yoon, 1993, p. 86) we find that the estimate of global schoollevel variable RURAL is .98 (standard error .64) in CWC₂, while in CGM₂ the same estimate is 1.36 (standard error .63). In CWC₂ the coefficient is not significant, while the coefficient in CGM₂ is significant, which results in different conclusions based on the same dataset, based on the same multilevel model, based on the same software (HLM), but using different centering methods.

Some Additional Algebra

There are some small, but interesting, extensions of our results that we have not discussed so far. We collect them in this section.

First, there is the simplification that obtains if there are no random slopes. Thus we have model (3.1)-(3.3), but $\underline{\delta}_{1j} \equiv 0$. This implies that $\omega_{01} = \omega_{10} = \omega_{11} = 0$, and thus $\mathbf{V}(\underline{Y}_{ij}) = \omega_{00} + \sigma^2$, and $\mathbf{C}(\underline{Y}_{ij}, \underline{Y}_{kj}) = \omega_{00}$. Because all X_{ij} have now disappeared from the random part of the model, it follows immediately that all random parts are equivalent. It does not even matter if we reintroduce the context mean as a predictor of the random intercept, all six models of Table 1 are equivalent as far as the variance component structure is concerned. The results we had earlier about the fixed parts of the models do not change. Thus, for random intercept models, CWC_1 is still not equivalent to RAS_1 and CWC_1 , but CWC_2 is equivalent to RAS_2 and CGM_2 . A model with a random intercept only, but with centering within context, does not seem to appear in the literature. It seems to be general practice among HLM software users to employ CWC only for variables with random slope coefficients (compare Bryk and Raudenbush, 1992, and many of the contributions to Raudenbush and Willms, 1992).

Second, there is the complication that results from introducing the context mean as a predictor of the random slope. Thus (6.3) becomes

$$\underline{b}_j = \gamma_{10} + \gamma_{11}X_{\bullet j} + \underline{\delta}_{1j}. \quad (9)$$

The random part of the model is still given by (7.2)-(7.3), but the fixed part becomes

$$\mathbf{E}(\underline{Y}_{ij}) = \gamma_{00} + \gamma_{01}X_{\bullet j} + \gamma_{10}X_{ij} + \gamma_{11}X_{\bullet j}X_{ij}. \quad (10)$$

Using our familiar method we see that equivalence of RAS and CGM in this model is equivalent to solvability of

$$\gamma_{00} = \gamma_{00}^* - \gamma_{10}^*X_{\bullet\bullet}, \quad (11.1)$$

$$\gamma_{01} = \gamma_{01}^* - \gamma_{11}^*X_{\bullet\bullet}, \quad (11.2)$$

$$\gamma_{10} = \gamma_{10}^*, \quad (11.3)$$

$$\gamma_{11} = \gamma_{11}^*. \quad (11.4)$$

This can easily be inverted to

$$\gamma_{00}^* = \gamma_{00} + \gamma_{10}X_{\bullet\bullet}, \quad (11.5)$$

$$\gamma_{01}^* = \gamma_{01} + \gamma_{11}X_{\bullet\bullet}, \quad (11.6)$$

$$\gamma_{10}^* = \gamma_{10}, \quad (11.7)$$

$$\gamma_{11}^* = \gamma_{11}. \quad (11.8)$$

Thus we continue to have equivalence in this case. On the other hand, if we center within context in this case, we see that $X_{\bullet j}X_{ij}^{**} = X_{\bullet j}X_{ij} - X_{\bullet j}^2$. The quadratic term in the

context means immediately shows that CWC in this case is not equivalent to RAS or CGM.

DISCUSSION

Let us start by summarizing our results on equivalence. Remember that equivalence does not mean that the models produce the same parameter estimates, nor does it necessarily mean that the same number of parameters is estimated. It means that the values of the estimates can be easily recalculated from one centering method to another. We have seen that RAS and CGM are always equivalent, even if we use $X_{\bullet j}$ as the macro-predictor of random slope and intercept. CWC_1 is not equivalent to CGM_1 and RAS_1 , but CWC_2 is the most interesting model in terms of equivalence. It is equivalent in its fixed part to RAS_2 and CGM_2 , but not in its variance component part. If slopes are fixed, then CWC_2 is completely equivalent to RAS_2 and CGM_2 . If $X_{\bullet j}$ is used as the macro-predictor of random slope then CWC_2 is no longer equivalent, neither in its fixed part nor in its random part. These are the facts. What can we suggest to the practical researcher as a consequence of these facts.

Practical Considerations in the Choice of Model

There is no statistically correct choice among RAS, CGM, and CWC, because from the statistical point of view the models are all equally correct. There are some minor differences in parsimony, but basically each of the models we have considered can be applied correctly to empirical data.

From the point of view of estimation the question is mainly one of computational ease and stability. If predictor variables have widely differing scales, e.g. SAT scores with a mean of 500 and a standard deviation of 100 versus a grade point average scale with a range of four, then centering is called for. Since scales in psychology are in the main arbitrary, rescaling predictors to approximately equal locations and variances prior to analysis is often both possible and desirable. Thus, as the computational argument leads equally to the choice of CGM_2 and CWC_2 , the choice among these latter two models must be determined by theory.

Theoretical Considerations in the Choice of Model

In some literature the context is considered as crucially important, as in the Theory of Reasoned Action (Fishbein and Ajzen, 1975) or as in Aptitude \times Treatment Interaction research (Cronbach and Webb, 1975). TRA asserts that intentions of individuals are determined by two components: an individual's attitude towards certain behavior, and an individual's perception of the social pressure and subjective norms to engage in that type of behavior. Many times the group mean is used as a proxy for these social norms or pressure. TRA researchers model the treatment explicitly, either as a group mean or as an interaction between a context effect and an individual characteristic. This seems to lead to RAS_2 forms of modeling.

In Cronbach's ATI theory distinguishing the within-group and between-group variation in regressions is regarded as an important theoretical step. Cronbach and Webb (1975, p. 717) argue that there are comparative, class-level, and individual effects in the regression. The comparative effects are measured by within-group regression, the class-level effects by between-group regression, and the individual effects influence both regressions. There is no way to separate all three types of effects, but contrasting the between-group and within-group regressions does provide a fairly precise analysis. Raudenbush (1989a, p. 10) uses Cronbach's arguments to support of CWC₂, by remarking that not centering this way "led to distorted estimates". The CWC₂ approach has a natural appeal to people using fixed effects analysis of (co)variance models. But it is not so natural in random coefficient models, because even the most simple contextual models are no longer equivalent in the variance-covariance part over different forms of centering. In fixed coefficient models we can do the between-group and within-group analysis independently of each other. This is no longer possible in random coefficient models.

We use another simple example to illustrate theoretically driven choice of model, closely related to Cronbach's ATI. If scores on a standardized test, such as SAT, are used to predict success in college, the effect of SAT can be considered purely as an individual effect, or can be considered a joint effect of the individual SAT and the school average SAT. In this last case the variation in the SAT score is divided in a between- and a within-part, where the between-part represents the effect of the school, and the within-part represents the effect of the individual relative to her peers. In the CWC₂ model that formalizes these notions each student's success is predicted from her being a big or a small frog in that particular pond, and from the size of the pond (known as the "frog-pond" effect). Burstein (1980, p. 200-201) reviews the sociological and social-psychological literature which discusses relative standing in the group as a determinant of behavior. He shows that the frog pond effect theory naturally leads to the CWC₂ model. The contextual model, discussed in the sociological methodological literature in the seventies, assumes that the context has an effect on individual behavior. This leads directly to RAS₂. Of course we can also have frog-pond effects without contextual effects, which leads to CWC₁ models. An example would be that grades are used as the predictor of success in college, where grades are based on the subjective evaluation of a teacher, who grades in relation to the total achievement of the class (who "grades on a curve"). The various issues are discussed in considerable detail in Burstein (1980).

One reason why CWC₁ or CWC₂ appear to be a logical way of handling the data can be found in the history of these models. Multilevel analysis has grown out of the notion of separate models for separate schools (Burstein, Linn and Capell, 1978, and Cronbach, 1976), and the corresponding two-step analysis. In the two-step method separate models are fitted within schools in the first step, and the aggregated first step coefficients are used as dependent variables in the second step. This "Slopes-as-Outcomes" approach naturally led to the emphasis on the two-step formulation of random coefficient models in Raudenbush and Bryk's software and papers. If multilevel analysis is considered as a two-step approach, then it is natural to center within each group, since groups are considered as separate entities, which are only connected in the second step. Also centering in each group

does not even change the numerical values of the within-group slopes, although it does of course change the intercepts. Centering within groups changes the interpretation of the within-group slopes, from the expected performance of an individual with zero scores on all predictors to the expected performance of an individual that performs at the group average on all predictors. This last interpretation is, in many cases, the more natural one, because zero scores on predictors are often even outside the range of the scale. Of course changing the interpretation of first-level intercepts also means changing the interpretation of the second-level regression equations, which predict these first level intercepts. If the emphasis is on slopes, and not on intercepts, then centering within groups seems irrelevant, because it does not even change the numerical values of the within-group slopes. If multilevel analysis, on the other hand, is seen as a one-step procedure, fitting an overall regression model with cross-level interactions, then CWC is no longer such an obvious choice.

It is clear, however, from the sociological literature on contextual models that the theoretical discussion was complicated by the methodological fact that RAS, CGM, and CWC cannot be distinguished on the basis of empirical data or statistical analysis. As we have seen, this fact is no longer true for random coefficient models. Thus the choice has to be made more explicitly, based on theoretical considerations. It is this foundation that is missing in much of the literature, where typically CWC₁ or CWC₂ is used (see Bryk and Raudenbush, 1992, and the contributions of Gamoran, Fitz-Gibbon, Bryk & Frank, Lee & Smith, in Raudenbush and Willms, 1991). One could argue that RAS₁ or RAS₂ users are also in need of a theory, but the practice seems to be that researchers do not need to specify why they ignore certain effects, like second order or third order effects. It is more common practice that researchers specify why they include certain effects, than to explain why they exclude certain effects.

To sum up, using the SAT example again, we would say: if SAT is used to predict success in college, then in order to choose a centering method, it is necessary that the researcher determines her position in the above debate first. The position can be that either success in college is considered an individual effect, or that success in college is considered, at least partly, a school effect. The fact that an individual went to a certain school is used in the model as a purely individual characteristic (highly associated with SES of parents, neighborhood ethnicity, and so on), or as partly a school characteristic. Both definitions can be defended in the light of the research question. In the last case CWC₂ would be a good approach, in the first case RAS₂/CGM₂ would be better. It is up to the researcher to decide which model to use, given her philosophy, her knowledge of the data, and her research question. We don't think using CWC₂ is ever a question of superior formulation, and the choice can certainly not be made by producers of software.

REFERENCES

- [1] L. S. Aiken and S. G. West. *Multiple Regression: Testing and Interpreting Interaction*. Sage Publications, Newbury Park, CA, 1991.
- [2] D.A. Belsley. *Conditioning Diagnostics*. Wiley, New York, NY, 1991.
- [3] R. D. Bock. *Multilevel Analysis of Educational Data*. Academic Press, New York, NY, 1989.
- [4] A. S. Bryk and S. W. Raudenbush. *Hierarchical Linear Models. Applications and Data Analysis Methods*. Sage Publications, Newbury Park, CA, 1992.
- [5] A.S. Bryk, S.W. Raudenbush, M. Seltzer, and R.T. Congdon. *An Introduction to HLM: Computer Program and User's Guide*. University of Chicago, 1988.
- [6] L. Burstein. The Analysis of Multilevel Data in Educational Research and Evaluation. *Review of Research in Education*, 8:158–233, 1980.
- [7] L. Burstein, R.L. Linn, and F.J. Capell. Analyzing Multilevel Data in the Presence of Heterogeneous Within-class Regressions. *Journal of Educational Statistics*, 3:347–383, 1978.
- [8] L. J. Cronbach and N. Webb. Between- and Within-Class Effects in a Reported Aptitude- by-Treatment Interaction: Reanalysis of a Study by G. L. Anderson. *Journal of Educational Psychology*, 6:717–724, 1975.
- [9] L.J. Cronbach. Research in Classrooms and Schools: Formulation of Questions, Designs and Analysis. Occasional paper, Stanford Evaluation Consortium, Stanford, CA, 1976.
- [10] O.D. Duncan, R.P. Curzort, and R.P. Duncan. *Statistical Geography: Problems in Analyzing Areal Data*. Free Press, Glencoe, IL, 1966.
- [11] M. Fishbein and L. Ajzen. *Belief, Attitude, Intention and Behavior*. Addison-Wesley, Reading, MA, 1975.
- [12] H. Goldstein. *Multilevel Models in Educational and Social Research*. Griffin, London, GB, 1987.
- [13] J. Hemelrijk. Underlining Random Variables. *Statistica Neerlandica*, 20:1–7, 1966.
- [14] I.E.A. *Science Achievement in Seventeen Countries*. Pergamon, Oxford, GB, 1988.
- [15] I. G. G. Kreft. *Models and Methods for the Measurement of School Effects*. Elinkwijk, Utrecht, Netherlands, 1987.
- [16] I.G.G. Kreft and J. de Leeuw. The Gender Income Gap. *Sociological Methods and Research*, 23:319–341, 1994.
- [17] I.G.G. Kreft, J. de Leeuw, and K.-S. Kim. Comparing Four Different Statistical Packages for Hierarchical Linear Regression: GENMOD, HLM, ML2, VARCL. Preprint 50, UCLA Statistics, Los Angeles, CA, 1990.

- [18] P. F. Lazarsfeld and H. Menzel. On the Relation between Individual and Collective Properties. In E. Etzioni, editor, *Complex Organizations: A Sociological Reader*. Holt, Rinehart and Winston, New York, NY, 1961.
- [19] N. T. Longford. To Center or Not To Center. *Multilevel Modelling Newsletter*, 1(3):7 and 8 and 11, 1989.
- [20] N.T. Longford. *VARCL. Software for Variance Component Analysis of Data with Nested Random Effects (Maximum Likelihood)*. Educational Testing Service, Princeton, NJ, 1990.
- [21] W.M. Mason. Notes on Centering in Multilevel Analysis. Technical report, Department of Sociology UCLA, Los Angeles, CA, 1991.
- [22] W.M. Mason, A.F. Anderson, and N. Hayat. *Manual for GENMOD*. Department of Sociology UCLA, Los Angeles, CA, 1991.
- [23] I. Plewis. Comment on “Centering” Predictors. *Multilevel Modelling Newsletter*, 1(3):6 and 11, 1989.
- [24] R. Prosser, J. Rabash, and H. Goldstein. *ML3. Software for Three-level Analysis. User’s Guide for V2*. Institute of Education, University of London, London, GB, 1992.
- [25] S. W. Raudenbush and J. D. Willms. *Schools, Classrooms, and Pupils*. Academic Press, San Diego, CA, 1991.
- [26] S.W. Raudenbush. A Response to Longford and Plewis. *Multilevel Modelling Newsletter*, 1(3):8–10, 1989.
- [27] S.W. Raudenbush. “Centering” Predictors in Multilevel Analysis: Choices and Consequences. *Multilevel Modelling Newsletter*, 1(2):10–12, 1989.
- [28] B.-K. Yoon. *A Comparative Examination of the Determinants of Science Achievement in Korea and the United States: Multilevel Perspectives*. PhD thesis, University of California, Los Angeles, 1993.