THE UNIVERSITY of York

CENTRE FOR HEALTH ECONOMICS

MULTILEVEL MODELS WHERE THE RANDOM EFFECTS ARE CORRELATED WITH THE FIXED PREDICTORS: A CONDITIONED ITERATIVE GENERALISED LEAST SQUARES ESTIMATOR (CIGLS)

Nigel Rice
Andrew Jones
Harvey Goldstein

CHE Technical Paper Series 10

Multilevel Models Where the Random Effects are Correlated with the Fixed Predictors: A Conditioned Iterative Generalised Least Squares Estimator (CIGLS)

Nigel Rice
Andrew Jones
Harvey Goldstein

December 1997

CENTRE FOR HEALTH ECONOMICS TECHNICAL PAPER SERIES

The Centre for Health Economics has a well established Discussion Paper series which was originally conceived as a means of circulating ideas for discussion and debate amongst a wide readership that included health economists as well as those working within the NHS and pharmaceutical industry.

The introduction of a Technical Paper Series offers a further means by which the Centre's research can be disseminated. The Technical Paper Series publishes papers that are likely to be of specific interest to a relatively specialist audience, for example papers dealing with complex issues that assume high levels of prior knowledge, or those that make extensive use of sophisticated mathematical or statistical techniques.

The series provides a maximum of 20 pages for authors. The content and its format are entirely the responsibility of the author, and papers published in the Technical Paper series are not subject to peer-review or editorial control, unlike those produced in the Discussion Paper series. Offers of further papers, and requests for information should be directed to Frances Sharp in the Publications Office, Centre for Health Economics, University of York.

Address for Correspondence:

Nigel Rice
Centre for Health Economics
University of York
York YO1 5DD

Price £4.00

Multilevel models where the random effects are correlated with the fixed predictors: a conditioned iterative generalised least squares estimator (CIGLS)

By

Nigel Rice[1], Andrew Jones[2] and Harvey Goldstein[3]

17 December 1997

SUMMARY

For small group sizes, the multilevel iterative generalised least squares (IGLS) estimator is biased and inconsistent where the random effects are correlated with the fixed predictors. By exploiting the iterative nature of the IGLS algorithm we show how unbiased and consistent estimates can be obtained without conditioning on dummy variables or measuring fixed predictors as deviations from group means. The method proposed provides consistent estimation of the regression parameters of interest whilst retaining the properties of random effects models via efficient estimation and full exploration of residual heterogeneity.

Keywords: MULTILEVEL MODELS; ITERATIVE GENERALISED LEAST SQUARES; RANDOM EFFECTS; FIXED EFFECTS; COVARIANCE ESTIMATOR; CONDITIONED ITERATIVE GENERALISED LEAST SQUARES.

---

[1] Address for correspondence: Research Fellow, Centre for Health Economics, University of York, York, Y01 5DD, England. E-mail: nr5@york.ac.uk

[2] Senior Lecturer, Department of Economics and Related Studies, University of York, York, Y01 5DD, England.

[3] Professor of Statistical Methods, Institute of Education, University of London, London, WC1H 0AL, England.

1. INTRODUCTION

This paper considers an extension to the iterative generalised least squares estimator (IGLS) proposed by Goldstein (Goldstein (1995)) to produce consistent estimates of fixed predictor parameters for multilevel models where the random effects are correlated with the fixed predictors and group sample sizes are small. The motivation for this work draws heavily on the econometrics literature on panel data estimators and the debate surrounding the use of fixed versus random effects. The general class of multilevel models can be seen as an extension of panel data models to the case where there are any number of levels in the data hierarchy and the variance function is complex (see Goldstein (1995) for full description). We begin by presenting the issues as they have been discussed in the panel data literature and extend these to the general case of multilevel models in section 3. More complex multilevel models are considered in section 4. Sections 5 and 6 present some simulations and examples.

2. PANEL DATA ESTIMATORS

There has been much debate in the literature on panel data estimators, between the two alternative specifications of fixed and random effects (see for example, Judge et al (1980), Hsiao (1986) and Baltagi (1995)). In its simplest form a variance components or time series cross-sectional model can be specified in the following manner:

$$y_{it} = (\beta_X X)_{it} + \mu_i + e_{it} \qquad i = 1,\ldots,N; \quad t = 1,\ldots,T \qquad (1)$$

where X is a $K \times 1$ vector of exogenous variables with X_{it} representing the *it*th observation on the *K* variables, and β_X is a $K \times 1$ vector of constants. The error term e_{it} is independently identically distributed (i.i.d.) over i and t, with zero mean and constant variance, σ_e^2. Unobserved or unobservable individual specific effects are represented by μ_i which are assumed time invariant. In this formulation, there are T observations per individual and a total of NT observations across all individuals. In model (1) the parameters of interest are typically those associated with the fixed predictors β_X. The specification of the quantities μ_i and e_{it} are central to the debate concerning the relative merits of fixed and random effect models.

In the fixed effects model individual effects, μ_i, are treated as fixed but unknown to the observer. Such models allow the investigator to make inference conditional on the effects that are contained within the sample. In contrast, a random effects specification may be viewed as providing marginal or unconditional inference with respect to the population of all effects. It treats μ_i as being random draws from an i.i.d distribution, typically $\mu_i \sim N(0, \sigma_\mu^2)$. The choice of specification may, in certain circumstances, be clear to the analyst and depend on the manner in which the data were sampled and the context of the investigation. For example, see discussions by Hausman (1978) and Goldstein (1995).

However, as Hausman (1978) shows, estimates derived from fixed and random effects specifications may lead to vastly different estimates of the parameters of the explanatory variables. Hausman compares the two competing specifications when estimating a wage equation using a sample of 629 high school graduates followed over a six year period. Comparisons of the two sets of parameter estimates (relative to standard errors) obtained for explanatory variables of interest showed a marked difference. One case where such differences arise is where the individual specific effects are correlated with one or more components of the explanatory variables X_{it} and within group sample sizes T are small.

Mundlak (1978) argues that the conflicts between random and fixed effects formulations is essentially erroneous, and that the issue is really one of model specification. He considers a random effects specification of model (1) and represents the joint distribution of the explanatory variables X_{it} and the effects μ_i by approximating $E(\mu_i | X_{it})$ by an auxiliary linear regression based on within group means $(\overline{X}_i)$:

$$\mu_i = \overline{X}_i'\eta + w_i, \qquad \overline{X}_i = \frac{1}{T}\sum_t X_{it}, \qquad w_i \sim N(0, \sigma_w^2) \tag{2}$$

Here, Mundlak assumes that the individual effects are a linear function of the averages of all the explantory variables across time and that these effects are uncorrelated with explanatory variables if and only if $\eta = 0$. Substituting (2) into (1) gives:

$$y_{it} = (\beta_X X)_{it} + \overline{X}_i'\eta + w_i + e_{it} \qquad i = 1, \ldots, N; \quad t = 1, \ldots, T \tag{3}$$

where $E(e_{it} + w_i) = 0, \quad E(e_{it} + w_i)(e_{it} + w_i)' = \begin{cases} \sigma_e^2 I_T + \sigma_w^2 J_T, & i = i', \\ 0, & i \neq i', \end{cases}$

and I_T, J_T are respectively the identity matrix and the square matrix of ones, of order T.

Mundlak shows that the generalised least squares (GLS) estimator of the β_X in (3) is identical to the fixed effects estimator achieved by applying OLS to the model

$$y_{it} = (\beta_X X)_{it} + \sum_{i=1}^{N-1} \alpha_i d_i + e_{it}^* \qquad i = 1, \ldots, N; \quad t = 1, \ldots, T \tag{4}$$

where $\{d_i\}$is a set of N-1 dummy variable indicators representing individual effects (chosing one as a 'base'). By using this formulation, Mundlak maintains that the difference between the fixed and random effects approach is based on incorrect specification and that only when the assumption that $corr(X_{it}, \mu_i) = 0$ holds, does (3) reduce to (1) such that GLS estimation of β_X in (1) is equivalent to OLS estimation of β_X in (4). Hausman and

Taylor (1981) consider a generalisation of the model proposed by Mundlak and partition X into those variables which are correlated with μ_i and those which are not.

In practical situations, consideration of possible correlations between the set of explanatory variables X_{it} and the individual effects may be compromised in order to preserve degrees of freedom. For fixed T, random effects estimation of β_X by GLS ($\hat{\beta}_{GLS}$) is asymptotically efficient whilst the least squares dummy variable (LSDV) fixed effects estimator ($\hat{\beta}_{FE}$) is unbiased and consistent (consistent estimation of the dummy variable coefficients is only obtained for $T \rightarrow \infty$). However, LSDV will be expensive in terms of the loss of degrees of freedom where the number of individuals, N, is large compared to the total number of observations NT. The trade off between efficiency and consistency may well be a deciding factor in the practical choice between specifications.

Hausman (1978) proposes testing the assumption of independence of μ_i and X_{it} by considering the contrast between the two estimators, $\hat{q} = \hat{\beta}_{FE} - \hat{\beta}_{GLS}$. When μ_i and X_{it} are orthogonal, $\hat{q}$ will be near zero. Therefore, whilst $\hat{\beta}_{GLS}$ is a weighted average of $\hat{\beta}_{FE}$, when the specification is correct, the two estimators should approximate to the same. This is equivalent to testing the hypothesis, $H_0 : \eta = 0$ in model (3).

In this paper, we add to the debate on the relative merits of fixed and random effects by considering the general case of multilevel models. We describe an extension to the iterative generalised least squares estimator (IGLS (Goldstein (1995)) that provides consistent and efficient estimation of explanatory variable parameters in situations where some or all of the explanatory variables are correlated with the random effects and within group sample sizes are small. The procedure is general and may be extended to include variables measured at the group level, more than two levels of the data hierarchy as well as random-coefficients and variable within group sample sizes.

3. MULTILEVEL MODELS

3.1. Variance components model

Model (1) above can be regarded as a simple case of the general multilevel model proposed by Goldstein (1986) and may be termed a two-level variance components multilevel model. To maintain consistency with the literature on multilevel models we change the notation used in model (1) to the following:

$$y_{ij} = (\beta_X X)_{ij} + u_j + e_{ij} \qquad i = 1,\ldots,N; \quad j = 1,\ldots,M \qquad (5)$$

Again, X is a $K \times 1$ vector of exogenous variables, and β_X a $K \times 1$ vector of constants. We assume here that there are M level 2 units or groups (analogous to individuals in model (1)) and N observations in total (and hence a total of N level 1 observations).

Group sample sizes n_j are not required to be constant across the M groups. The components u_j and e_{ij} are residuals at level 2 and level 1 respectively, assumed to be i.i.d. with zero mean and constant variance:

$$\text{cov}(u_j, u_{j'}) = \text{cov}(e_{ij}, e_{i'j}) = 0,$$
$$\text{cov}(u_j, e_{ij}) = 0,$$
$$E(u_j) = E(e_{ij}) = 0,$$
$$\text{var}(u_j) = \sigma_u^2,$$
$$\text{var}(e_{ij}) = \sigma_e^2.$$

The quantities of interest in (5) are the estimated parameters $\hat{\beta}_X$ (termed the fixed part parameters) and the estimated random components $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$ (termed the random part parameters).

In the absence of correlation between the components of X and the level 2 random effects u_j, the iterative GLS estimator of (5) produces both efficient and consistent estimates of the fixed and random part parameters for fixed n_j (Goldstein, 1986). However, as discussed above where correlations between the level 2 random effects u_j and components of X exist, although IGLS estimation is efficient it is inconsistent as $M \to \infty$ when group sample sizes n_j are small (for example, see Blundell and Windmeijer (1997) for a full discussion relating to multilevel models).

3.2. Consistent estimation

As noted earlier, consistent estimation of β_X in (5) can be achieved by specifying the model as a fixed effects model analogous to (4) and estimating by OLS. Alternatively, if we pre-multiply (4) (whilst retaining the multilevel notation) by the idempotent matrix

$$Q = \{Q_j\}, \quad Q_j = I_{n_j} - J_{n_j} / n_j \tag{6}$$

where $\{\ \}$ denotes a matrix, and I_{n_j}, J_{n_j} are respectively the identity matrix and the square matrix of ones, of order n_j, we have (in matrix notation):

$$QY = QX\beta_X + QE^* \tag{7}$$

Applying OLS to (7) leads to consistent estimates of β_X. The estimator $(X^T QX)^{-1}(X^T QY)$ is known as the within groups or covariance estimator (CV) (for example, see Hsiao (1995)).

We now consider the following alternative conditioned iterative estimation procedure (CIGLS). IGLS estimation may be viewed as a two step procedure for each iteration. In the first step we re-express model (5) in matrix notation as:

$$\begin{aligned} Y &= X\beta_X + S\beta_S + E \\ &= Z\beta_Z + E \end{aligned} \tag{8}$$

where $E = \{e_{ij}\}$ and

$$S = \{S_j\},\quad S_j^T = (s_1,\ldots,s_M),\quad s_j = \overset{*}{s}_j \times \iota_{n_j},\quad \overset{*}{s}_j = \sum_{i=1}^{n_j} \hat{w}_{ij} / n_j, \\ \hat{W} = \{\hat{w}_{ij}\} = Y - X\hat{\beta}_X^* \tag{9}$$

where ι_{n_j} is a vector of ones of length n_j, $\hat{\beta}_X^*$ is the current estimate of β_X and S_j is obtained by stacking the vectors s_1 to s_M and is of length $\sum_{j=1}^{M} n_j = N$. In other words, the vector S consists of the group means of the estimated residuals from the previous iteration. Once S is constructed, updated estimates of β_X are then obtained through GLS estimation of (8).

In the second step, we condition on $\hat{\beta}_X$ and form the matrix $\overset{*}{Y} = \hat{W}\hat{W}'$. By stacking the columns of $\overset{*}{Y}$ these are regressed on the random parameter design matrix and GLS estimation produces the parameters of interest; $\hat{\sigma}_e^2$ and $\hat{\sigma}_u^2$ (for a variance components model, the random parameter design matrix is the block diagonal matrix leading to the covariance matrix V with elements $\sigma_e^2 I_{n_j} + \sigma_u^2 J_{n_j}$ for each group or block (for a full discussion, see Goldstein, 1995)).

Suitable starting values for $\hat{\beta}_X^*$ may be obtained by OLS estimation of (5). Iteration of the two steps proceeds to convergence defined by a pre-assigned tolerence for $(\hat{\beta}_X - \hat{\beta}_X^*)$.

3.3. Convergence

We can re-express S in (8) as

$$S = (I - Q)(Y - X\hat{\beta}_X^*) \tag{10}$$

where Q is defined in (6) and I is the identity matrix. It then follows directly form (8) that

$$Y = X\beta_X + ((I - Q)(Y - X\hat{\beta}_X^*))\beta_S + E \tag{11}$$

If at convergence we have $\hat{\beta}_X = \hat{\beta}^*_X$, and $\hat{\beta}_S = 1.0$, (8) reduces to

$$QY = QX\beta_X + E \tag{12}$$

which is equivalent to the within groups specification (7) with $E = \overset{*}{E}Q.$.

It follows immediately therefore that the GLS estimator for the full set of fixed coefficients $\beta_Z = \begin{pmatrix} \beta_X \\ \beta_S \end{pmatrix}$ in (9), namely $(Z^T V_E^{-1} Z)^{-1}(Z^T V_E^{-1} Y)$, where V_E is the block diagonal covariance matrix $\left(= \begin{cases} \sigma_e^2 I_{n_j} + \sigma_u^2 J_{n_j}, & i = i', \\ 0, & i \neq i', \end{cases} \right)$, provides both the efficient and *consistent* (maximum likelihood under Normality) estimator of β_X.

Note, that in comparison to the GLS estimator, the OLS estimator ignores the lack of independence induced by premultiplying E in the equivalent multilevel fixed effects specification of (4) or (8) by Q. Also, if at convergence we obtain a value of $\hat{\beta}_S$ substantially different from 1.0 this may indicate misspecification in either the fixed or random parts of the model and could form a basis for diagnostic specification checks.

4. EXTENSIONS

4.1. Group level predictors

An obvious extension to (5) is to consider predictor variables measured at the group level, such that we have:

$$y_{ij} = (\beta_X X)_{ij} + (\beta_G G)_j + \overset{*}{u}_j + e_{ij} \qquad i = 1,\ldots,N; \quad j = 1,\ldots,M \tag{13}$$

In the formulation adopted in model (1), G would represent a matrix of time invariant regressors. The corresponding re-expression of (13), substituting dummy variable fixed effects for the random components $\overset{*}{u}_j$ cannot be estimated using LSDV because of perfect multicollinearity between the group level predictors and the fixed effects. Without imposing restrictions on the parameter estimates, consistent estimation is not feasible. Similarly, applying a covariance estimator leads to a lack of identification of the parameter estimates $\hat{\beta}_G$ since calculating deviations from group means will transform variables G to zero.

If model (13) is true and group level variables G are orthogonal to u_j^*, estimation using CIGLS proceeds in a straight forward manner and we construct S as in (9), except now $\hat{W}=\{\hat{w}_{ij}\}=Y-(X\hat{\beta}^*_X+G\hat{\beta}^*_G)$, where $\hat{\beta}^*_X$ and $\hat{\beta}^*_G$ represent current estimates. Where correlations between group level variables G and u_j^* exist, estimation using CIGLS will result in biased and inconsistent estimates of β_G; this being an obvious extension of omitted variable bias. In such instances, variables that will purge G and u_j^* of such correlation should be sought and included in the model specification.

4.2. Group level effects and their interpretation

Suits (1984) and Kennedy (1986) discuss the difficulties in interpreting model (4) based on the LSDV approach since identification often involves constraining one of the dummy variable coefficients to zero (in the presence of a constant term) and estimating the effects of other group membership relative to this 'baseline' group. Suits (1984) suggests that to aid interpretation of the coefficient estimates derived they should first be transformed such that estimates attached to all dummy variable groups are shown together with the corresponding adjustment to the constant regression term, β_0. In the case where all groups have equivalent populations, β_0 may be interpreted as the population average. Where within group population sizes differ, β_0 represents a weighted population average (Kennedy (1986)).

The estimation of S in (8) above does not rely on identification restrictions and as such S consists of group effect estimates for all groups, not $M-1$ as for the LSDV estimator. The corresponding estimate of the 'intercept' term, β_0 represents a weighted average across all groups (weighted proportional to the within group sample sizes n_j). As argued by Suits and Kennedy this representation of group effects lends itself more readily to interpretation.

4.3. Random coefficient models

4.3.1. Complex level 1 variation

The first extension is to consider complex level 1 variation (for a discussion, see Goldstein (1995), chapter. 3). The specification of random coefficients at this level can be viewed as explicitly modelling heteroskedasticity but may also be of substantive importance to the analyst. The basic results outlined above still hold, but now we no longer have the equivalence between OLS and GLS in the case where the variables are measured from their group means, and OLS may be much less efficient than GLS.

4.3.2. Random coefficients at level 2

Where we have random coefficients at level 2 the algorithm is modified as follows. As before, calculate the quantities $W = \{\hat{w}_{ij}\}$ and regress these on the level 2 random part explanatory variables. Thus, if we have an 'intercept' (S_0) and a 'slope' (S_1) at level 2, such that in model (5); $u_j = v_j + x_{ij}\lambda_j$, where effects v_j represent the intercept and λ_j the slope, then we estimate the coefficients in the following OLS model for each level 2 unit (or combining into a single OLS analysis with dummy variables for the groups).

$$\hat{w}_{ij} = s_0 + s_1 x_{ij} + e_{ij}^{**} \tag{14}$$

Once we have obtained the estimates $\hat{s}_0$ and $\hat{s}_1$ for each group j, we can construct the vectors S_0 and S_1 by multiplying the vectors ι_{n_j} by $\hat{s}_0$ and $\hat{s}_1$ and stacking the resulting vectors. We then carry out GLS estimation for the model

$$Y = X\beta_X + S_0\beta_{S_0} + \overset{*}{S_1}\beta_{S_1} + E^{**} \tag{15}$$

where $\overset{*}{S_1}$ is the $N \times 1$ vector: $x_{ij}S_1$.

If we re-express (14) as an OLS regression across all groups we have:

$$\hat{w}_{ij} = \sum_{j=1}^{M} \alpha_j d_j + \sum_{j=1}^{M} \xi_j \psi_j + e_{ij}^{**} \tag{16}$$

where $\psi_j = x_{ij}d_j$ and $\{d_j\}$ is a set of M dummy variables. We can retrieve the constructed vector S_0 by multiplying the vectors ι_{n_j} by the estimated coefficients α_j and stacking the resulting vectors. $\overset{*}{S_1}$ can be retrieved in a similar way by stacking the vectors obtained by multiplying the vectors ι_{n_j} by the set of coefficients ξ_j.

In matrix form (and using the multilevel notation) the equivalent model formed by combining equation (16) with the set of fixed part predictor variables of interest may be written as:

$$Y = X\beta_X + D\alpha + D'\xi + E^{**} \tag{17}$$

where $Y \to N \times 1$, $X \to N \times K$, $\beta \to K \times 1$, $D \to N \times M$, $\alpha \to M \times 1$, $\xi \to M \times 1$, and $E^{**} \to N \times 1$. D' is an $N \times M$ matrix formed by multiplying the dummy variable matrix D by the random coefficient vector X^T.

Model (17) is overidentified and cannot be estimated in a single step. In the presence of a constant term, the usual restriction is to re-specify D and D' to be matrices of order $N \times (M-1)$. The resulting (consistent) estimator is then LSDV.

If some of the level 2 random coefficients are uncorrelated with any of the explanatory variables, then these may be taken out and estimated in the usual way. To do this (14) will need to be modified to include only the correlated random coefficients and a GLS regression carried out for each level 2 unit, with the appropriate random coefficient contributing to the variance structure.

4.4. Higher levels

IGLS is not restricted to the simple case of a two level hierachy and similarly, this procedure can be extended to any number of levels but not to the case where a level 1 random effect (e^*_{ij}) is correlated with an explanatory variable. The reason for this is essentially the same as the case discussed in 4.1. and applies generally where explanatory variables are correlated with residuals (random effects) at the same level.

5. SIMULATIONS

5.1. Simulation 1: variance components model

We simulate a multilevel model and estimate its parameters using the consistent covariance estimator (CV), the multilevel GLS estimator (IGLS) and the conditioned multilevel GLS estimator described above (CIGLS). The following model was simulated

$$y_{ij} = 1 + 1x_{1ij} + 1.5x_{2ij} + u_j + e_{ij}$$

where

$$n_j = 5 \quad \forall j, \quad j = 1,\ldots,30$$

$$x_{1ij} \sim N(0,1), \quad x_{2ij} \sim N(0,1), \quad u_j \sim N(0,1), \quad e_{ij} \sim N(0,1.5)$$

$$\sigma_{x_{1ij}x_{2ij}} = 0, \quad \sigma_{x_{1ij},u_j} = 0, \quad \sigma_{x_{2ij}u_j} = 0.75, \quad \sigma_{e^*_{ij}u_j} = 0.$$

The results of 500 simulations of the above are presented in Table 1. We describe the distribution of the estimated parameters (including random effects) by the mean, standard deviation and mean squared error. The results show equivalence of CV and CIGLS estimates of β_X and improved estimates of β_2 and σ^2_u using CIGLS compared to IGLS.

5.2. Simulation 2: group level variable

We simulated two models that include a predictor variable measured at the group level (x_{2j}). Both have the following specification, but different assumptions concerning the correlation between the group level predictor variable and the group specific effects u_j.

$$y_{ij} = 1 + 1x_{1ij} + 1.5x_{2j} + u_j + e_{ij}$$

where

$$n_j = 5 \quad \forall j, \quad j = 1,\ldots,30$$

$$x_{1ij} \sim N(0,1), \quad x_{2j} \sim N(0,1), \quad u_j \sim N(0,1), \quad e_{ij} \sim N(0,1.5)$$

$$\sigma_{x_{1ij}x_{2ij}} = 0, \quad \sigma_{x_{1ij},u_j} = 0.75, \quad \sigma_{e_{ij}^{\bullet}u_j} = 0, \quad \rho_{x_{1ij}u_j} = 0.75.$$

and

1. $\sigma_{x_{2j}u_j} = 0,$
2. $\sigma_{x_{2j}u_j} = 0.25,$

Table 2 presents the results of simulating the above model with the assumption; $\sigma_{x_{2j}u_j} = 0$. As expected, both IGLS and CIGLS produce unbiased estimates of the parameter attached to x_{2j} (β_2); however, CIGLS also produces unbiased estimates of β_1. Results of the simulation setting $\sigma_{x_{2j}u_j} = 0.25$ are given in Table3. Although CIGLS produces improved estimates for β_1 (corresponding to CV estimates and those given in Table 2) as expected a biased estimate, equivalent to IGLS, of β_2 is obtained.

5.3. Simulation 3: random coefficient model

We now consider the simualtion of a random coefficient at level 2, such that:

$$y_{ij} = 1 + 1x_{1ij} + 1.5x_{2ij} + v_j + \lambda_j x_{2ij} + e_{ij}$$

where

$$n_j = 5 \quad \forall j, \quad j = 1,\ldots,30$$

$$x_{1ij} \sim N(0,1), \quad x_{2ij} \sim N(0,1), \quad v_j \sim N(0,1), \quad \lambda_j \sim N(0,1.25), \quad e_{ij} \sim N(0,1.5)$$

$$\sigma_{x_{1ij}x_{2ij}} = 0, \quad \sigma_{x_{1ij},v_j} = 0.75, \quad \sigma_{x_{2ij},v_j} = 0, \quad \sigma_{e_{ij}^{\bullet}v_j} = 0, \quad \sigma_{x_{1ij},\lambda_j} = 0.671, \quad \sigma_{\lambda_j,v_j} = 0.563.$$

The results of simulating the above model are presented in Table 3. Again we obtain improved estimates using CIGLS over IGLS, particularly for β_1 which corresponds to the LSDV estimate. The difference between the estimated constant and β_2 derived through LSDV and CIGLS is due to the different assumptions concerning the 'baseline group' adopted. For CIGLS the constant represents a weighted average over all level 2 units, as does the estimate for β_2. In contrast, LSDV estimates are made relative to a chosen 'baseline group' and as such can only be interpreted respective to a particular level 2 unit.

6. EXAMPLES

6.1. Example 1

We take as an example the following gasoline demand equation considered by Baltagi and Griffin (1983) and reproduced in Baltagi (1995):

$$\ln\left(\frac{Gas}{Car}\right)_{ij} = constant + \beta_1 \ln\left(\frac{Y}{N}\right)_{ij} + \beta_2 \ln\left(\frac{P_{MG}}{P_{GDP}}\right)_{ij} + \beta_3 \ln\left(\frac{Car}{N}\right)_{ij} + u_j + e_{ij} \qquad (18)$$

Gas/Car represents gasoline consumption per car, Y/N is real per capita income, P_{MG}/P_{GDP} is real price of gasoline and Car/N is the stock of cars per capita.The data consists of a panel of observations across 18 OECD countries, during the period 1960-1978. Years are indexed i, and countries j.

Various regression results for this model are reported in Baltagi (1995). Table 5 presents the results of an OLS specification of (18) together with a fixed effects model (LSDV estimator), IGLS and CIGLS. Clearly the parameter estimates derived from IGLS are not consistent (a Hausman test of fixed versus random effects specification (Hausman 1978) indicates that fixed effects are appropriate ($\chi_3^2 = 306, p < 0.01.$). The explanatory parameter estimates produced by CIGLS are the same as the LSDV estimates but the standard errors indicate the improved efficiency of CIGLS.

6.2. Example 2

In this example we consider a a multilevel analysis of school examination results described by Goldstein et. al. (1992). The data consist of General Certificate of Secondary Examination (GCSE) results from 5748 students in 66 schools in six Inner London Education Authorities. A full description of the examination data and the scoring system used is given in Nuttal et. al. (1989).

Students had scores on a common reading test taken at age 11 years; the London Reading Test (LRT) and were graded into three categories on the basis of a verbal reasoning (VR) test (VR1; band 1 (top 25%), VR2; band 2 (middle 50%), VR3; band 3 (bottom 25%)). Data on gender were also available. The response variable was transformed using normal scoring and the predictor variable LRT was standardised prior to fitting the following random coefficient model:

$$\begin{aligned} Score_{ij} = constant + \beta_1 LRT_{ij} + \beta_2 LRT_{ij}^{\ 2} + \sum_{h=1}^{2} \beta_h VR_{hij} + \beta_5 Gender_{ij} \\ + u_j + \lambda_j LRT + \gamma_j VR + e_{ij} \end{aligned} \qquad (19)$$

In model (19) the parameters of interest are the estimated coefficients; $\hat{\beta}_1 - \hat{\beta}_5$, together with the estimated variance components representing the within school variance: $\hat{\sigma}^2_e$, and the between school variance terms: $\hat{\sigma}^2_u$, $\hat{\sigma}^2_{(LRT)}$, $\hat{\sigma}^2_{(VRQ)}$, together with the respective between school covariance terms: $\hat{\sigma}_{(LRT,u)}$, $\hat{\sigma}_{(VRQ,u)}$, $\hat{\sigma}_{(LRT,VRQ)}$.

The results obtained from applying IGLS and CIGLS are presented in Table 6 In this example, CIGLS has little effect over IGLS which is not surprising given that the group sample sizes are relatively large (range 29 to 219). However, the method illustrates the use of CIGLS in the presence of random coefficients (note that the lack of change in the random structure is a direct consequence of lack of change in the estimates of the fixed predictors).

7. CONCLUSION

The iterative generalised least squares estimator conditioning on the mean level 2 effects (CIGLS) provides both efficient and consistent estimates of β_X when the random effects are correlated with one or more of the fixed predictors and group sample sizes, n_j, are small. Modifications to the standard IGLS estimation routine are trivial and computationally undemanding in the case where variance components models are considered. More elaborate estimation is required where a random coefficient is also correlated with a fixed predictor, but again this can be handled adequately using existing software (MLn, Rasbash, J. et al (1995)). In all cases, the procedure avoids the use of dummy variables (as in the standard LSDV estimator) and hence the associated loss in degrees of freedom, and the requirement to transform data to represent deviations from group means.

REFERENCES

Baltagi, B.H. (1995). *Economic analysis of panel data*. New York, John Wiley & Sons.

Baltagi. B.H. and Griffin, J.M. (1983). Gasoline demand in the OECD: an application of pooling and testing procedures. *European Economic Review;* **22:** 117-137.

Blundell, R. and Windmeijer, F. (1997). Correlated cluster effects and simultaneity in multilevel models. *Health Economics*; **1**: 6-13.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalised least squares. *Biometrika*, **73**: 43-56.

Goldstein, H. (1995). *Multilevel Statistical Models (second edition).* London: Edward Arnold.

Goldstein, H. Rasbash, J. Yang, M. Woodhouse, G. Pan, H. Nuttall, D. Thomas, S. A *Multilevel analysis of school examination results*. Presented to the European Conference on Educational Research, University of Twente, June 1992.

Hausman, J.A. (1978). Specification tests in econometrics, *Econometrica* **46**: 1251-71.

Hausman, J.A. Taylor, W.E. (1981). Panel data and unobservable individual effects. *Econometrica*, **49**: 1377-98.

Hsiao, C. (1995). *Analysis of Panel Data*. Econometric Society Monographs; Cambridge: Cambridge University Press.

Judge, G.G. Griffiths, W.E. Hill, R.C. Lee, T-C. (1980*). The theory and practice of econometrics*, New York, John Wiley & Sons.

Mundlak, Y. (1978). On the pooling of time series and cross section data, *Econometrica* **46**: 69-81.

Nuttall, D,N. Goldstein, H. Prosser, R. Rasbash, J. (1989). Differential school effectivenesss. *International Journal of Educational Research*; **13**: 769-776.

Rasbash, J. Yang, M. Woodhouse, G. and Goldstein, H. (1995*): MLn: command reference guide*. London: Institute of Education.

	OLS			Multilevel			Multilevel		
	CV			IGLS			CIGLS		
	Mean	SD	MSE	Mean	SD	MSE	Mean	SD	MSE
Random effects									
Level 2 - σ^2_u				0.947	0.344	0.121	0.974	0.348	0.121
Level 1 - $\sigma^2_{e^*}$	1.484	0.195	0.038	1.493	0.196	0.038	1.460	0.190	0.038
Fixed predictors									
Constant				0.996	0.213	0.038	0.996	0.218	0.040
β_1	1.005	0.116	1.003	1.005	0.113	1.002	1.005	0.116	1.003
β_2	1.494	0.111	2.281	1.588	0.113	2.006	1.494	0.111	2.281

Table 1. Simulation 1: Variance components model.

	OLS			Multilevel			Multilevel		
	CV			IGLS			CIGLS		
	Mean	SD	MSE	Mean	SD	MSE	Mean	SD	MSE
Random effects									
Level 2 - σ^2_u				0.960	0.349	0.123	1.024	0.363	0.132
Level 1 - $\sigma^2_{e^*}$	1.510	0.195	0.038	1.492	0.194	0.038	1.484	0.192	0.037
Fixed predictors									
Constant				1.007	0.204	1.055	1.007	0.208	1.058
β_1	1.005	0.116	0.003	1.095	0.118	0.832	1.005	0.116	1.003
β_2				1.498	0.214	2.302	1.498	0.219	2.304

Table 2. Simulation 2a: Group level predictor; $\sigma_{x_{2j}u_j} = 0$.

	OLS CV			Multilevel IGLS			Multilevel CIGLS		
	Mean	SD	MSE	Mean	SD	MSE	Mean	SD	MSE
Random effects									
Level 2 - σ_u^2				0.894	0.346	0.121	0.961	0.346	0.121
Level 1 - σ_{e}^2	1.510	0.195	0.038	1.493	0.195	0.038	1.484	0.192	0.037
Fixed predictors									
Constant				1.007	0.198	1.052	1.007	0.202	1.055
β_1	1.005	0.116	0.003	1.100	0.118	0.824	1.005	0.116	1.003
β_2				1.748	0.208	1.612	1.748	0.214	1.614

Table 3. Simulation 2b: Group level predictor; $\sigma_{x_{2j}u_j} = 0.25$.

	OLS LSDV			Multilevel IGLS			Multilevel CIGLS		
	Mean	SD	MSE	Mean	SD	MSE	Mean	SD	MSE
Random effects									
σ_v^2				0.966	0.373	0.140	1.029	0.389	0.152
Level 2 - σ_λ^2				1.243	0.448	0.266	1.253	0.449	0.262
$\sigma_{(\lambda, v)}$				0.537	0.304	0.380	0.564	0.312	0.415
Level 1 - σ_{e}^2	2.251	0.342	0.180	1.502	0.227	0.299	1.494	0.225	0.307
Fixed predictors									
Constant	1.026	1.297	5.575	1.002	0.211	4.035	1.003	0.215	4.035
β_1	1.007	0.133	1.004	1.093	0.129	0.839	1.007	0.133	1.003
β_2	1.494	1.567	4.719	1.507	0.235	2.285	1.506	0.234	2.285

Table 4. Simulation 3: Random coefficient model.

Dependent Ln(Gas/Car)	OLS		LSDV		Multilevel IGLS		Multilevel CIGLS	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE
Random components								
σ_e^2			0.009		0.009	0.001	0.009	0.001
σ_u^2					0.094	0.031	0.123	0.041
Fixed predictors								
Constant	2.391	0.117	2.403	0.225	2.152	0.209	2.403	0.224
Ln(Y/N)	0.889	0.036	0.662	0.073	0.592	0.065	0.662	0.068
Ln(P_{MG}/P_{GDP})	-0.891	0.030	-0.322	0.044	-0.374	0.041	-0.322	0.043
Ln(Car/N)	-0.763	0.019	-0.641	0.030	-0.618	0.027	-0.641	0.028
S							1	0.251

Table 5
Note: Data reported in Baltagi (1995), Appendix 6.

Dependent Pupil examination score	Multilevel IGLS†		Multilevel CIGLS	
	Coef	SE	Coef	SE
Random components				
Pupil level				
σ^2_e	0.519	0.010	0.519	0.010
School level				
σ^2_u	0.069	0.014	0.070	0.014
$\sigma_{(LRT,u)}$	0.013	0.004	0.013	0.004
$\sigma^2_{(LRT)}$	0.003	0.002	0.003	0.002
$\sigma_{(VRQ,u)}$	0.005	0.010	0.005	0.010
$\sigma_{(VRQ,LRT)}$	0.007	0.004	0.007	0.004
$\sigma^2_{(VRQ)}$	0.024	0.013	0.024	0.013
Fixed predictors				
Constant	-0.454	0.046	-0.448	0.045
LRT	0.359	0.016	0.360	0.016
LRT Squared	0.043	0.007	0.043	0.007
VRQ2§	0.702	0.047	0.700	0.047
VRQ3§	0.332	0.031	0.329	0.031
Gender (Female)	0.134	0.026	0.125	0.026
S_{0j}			1	0.109
S_{LRTj}			1	0.101
S_{VRQj}			1	0.116

Table 6

Note: § 'Baseline - VRQ1.

† Likelihood ratio test statistics for random part parameters:

I. Joint test of significance of VRQ: $\sigma^2_{(VRQ)}, \sigma_{(VRQ,u)}, \sigma_{(VQR,LRT)}$; $\chi^2_3 = 9.01, p = 0.029$.

II. Joint test of significance of LRT: $\sigma^2_{(LRT)}, \sigma_{(LRT,u)}\ \sigma_{(LRT,VRQ)}$; $\chi^2_3 = 15.67, p = 0.001$.