# MULTILEVEL MODELS WHERE THE RANDOM EFFECTS ARE CORRELATED WITH THE FIXED PREDICTORS

Nigel Rice

Centre for Health Economics University of York Heslington York Y01 5DD England and Institute of Education University of London Andrew M Jones Department of Economics and Related Studies University of York Harvey Goldstein

Institute of Education University of London

Key words: MULTILEVEL MODELS; ITERATIVE GENERALISED LEAST SQUARES; RANDOM EFFECTS; FIXED EFFECTS; INSTRUMENTAL VARIABLES; CONDITIONED ITERATIVE GENERALISED LEAST SQUARES.

JEL classifications: C19, C23, C87, I21

## ABSTRACT

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. Consistent estimates of the parameters of endogenous variables may be obtained using instrumental variables or conditioning on group level effects. In this paper we review various approaches to ensure consistency in panel data models and extend these to the general class of multilevel models. Further, by exploiting the iterative nature of the IGLS algorithm we derive an unbiased and consistent estimator based on conditioning on estimated group effects. The method proposed provides consistent estimation of the endogenous regression parameters of interest whilst retaining the properties of multilevel models via efficient estimation and full exploration of residual heterogeneity. The proposed estimator is termed conditioned iterative generalised least squares (CIGLS).

### 1. INTRODUCTION

For small group sizes, the multilevel iterative generalised least squares (IGLS) estimator is biased and inconsistent where the random effects are correlated with one or more fixed predictors. Such fixed predictor variables are termed endogenous and consistent estimators have been proposed in the literature on panel data models by, for example, taking deviations from group means, or employing instrumental variables estimators. Multilevel models are extensions of the random effects panel data models to the case where there are any number of levels in the data hierarchy and the residual variance function is complex and includes random coefficients at any level of the data hierarchy. In this paper we review consistent estimators proposed for use with panel data and show how these may be extended to the general class of multilevel model. Further, by exploiting the iterative nature of the IGLS algorithm we derive a consistent estimator based on conditioning on estimated group effects. The method proposed provides unbiased and consistent estimation of groupvarying endogenous regression parameters as well as variance and covariance parameters associated with random coefficients. The proposed estimator, termed CIGLS, is compared with various alternative estimators in an analysis of the income returns to schooling experience first considered by Cornwell and Rupert (1988) and subsequently by Baltagi and Khanti-Akom (1990). The potential efficiency gains are shown. The finite sample biases of these methods are also discussed.

#### 2. MULTILEVEL MODELS

#### 2.1. Two level variance components model

Consider the following two-level variance components multilevel model:

$$y_{ii} = X_{ii}^{T} \beta + Z_{i}^{T} \gamma + u_{i} + e_{ii}, \qquad i = 1, ..., N; \quad j = 1, ..., M,$$
(1)

where *i* indexes an individual observation (level 1) and *j* group membership (level 2).  $X_{ij}$  is a  $(P \times I)$  (level 1) vector of explanatory variables varying over both *i* and *j*;  $Z_j$  is a  $(G \times I)$  vector of group level explanatory variables (invariant within groups);  $\beta$  and  $\gamma$  are comformably dimensioned parameter vectors. We assume  $u_j \sim N(0, \sigma_u^2)$ and similarly  $e_{ij} \sim N(0, \sigma_e^2)$ . Model (1) can be regarded as a simple case of the general class of multilevel models described by Goldstein (1995). We assume here that there are *M* level 2 units or groups and *N* observations in total (and hence a total of *N* level 1 observations). Group sample sizes  $n_j$  may vary across the *M* groups. The components  $u_j$  and  $e_{ij}$  are disturbances at level 2 and level 1 respectively with;  $cov(u_j, u_{j'}) = cov(e_{ij}, e_{i'j}) = 0$ , and  $cov(u_j, e_{ij}) = 0$ . The quantities of interest in (1) are the estimated fixed parameters  $\hat{\beta}$  and  $\hat{\gamma}$  and the estimated random components  $\hat{\sigma}_u^2$  and  $\hat{\sigma}_e^2$ .

## 2.2 IGLS Estimation

Combining all *N* observations we can re-express (1) in matrix notation:

$$y = X\beta + Z\gamma + E$$
  
=  $W\delta + E$  (2)

where *E* is a block-diagonal error term such that  $E = Z_u u + Z_e e$  where  $Z_u$  and  $Z_e$  denote  $(N \times I)$  dimensioned vectors. The dimensions of *y*, *X*, and *Z* are  $(N \times I)$ ,  $(N \times P)$  and  $(N \times G)$  respectively.

For known  $V = cov(Y|W\delta) = cov(E)$ , estimation of (2) proceeds via the usual generalized least squares (GLS) estimators:

$$\hat{\delta} = \left(W^T V^{-1} W\right)^{-1} W^T V^{-1} Y, \quad cov \left(\hat{\delta}\right) = \left(W^T V^{-1} W\right)^{-1} \tag{3}$$

If  $\delta$  is known but V is unknown, then we can obtain estimators  $\hat{\vartheta}$  of the parameters of V using GLS as

$$\hat{\vartheta} = \left(D^T V^{*-1} D\right)^{-1} D^T V^{*-1} Y^* \tag{4}$$

here  $Y^*$  is a vector of the upper triangle elements of  $(Y - W\delta)(Y - W\delta)^T$ , that is, the squares and products of the residuals rearranged by stacking one column on top of the other, termed  $vec(\tilde{Y}\tilde{Y}^T)$ .  $V^*$  is the covariance matrix of  $vec(\tilde{Y}\tilde{Y}^T)$ ; assuming normality, this is  $2(V^{-1} \otimes V^{-1})$  where  $\otimes$  is the kronecker product. *D* represents the design matrix linking  $Y^*$  to *V* in the regression of  $Y^*$  on *D*.

The estimation procedure commences from an initial estimate of V which is used to obtain estimates  $\hat{\delta}$ , from which an updated estimate of V is obtained. This procedure is repeated iteratively until convergence is achieved (defined by a suitable choice of tolerance); hence the estimator is iterative GLS or IGLS. An initial consistent estimator of  $\hat{\delta}$  is obtained via OLS assuming  $V = \sigma^2 I$ . Assuming the existence of finite moments up to the forth moment, at convergence we obtain consistent estimators  $\hat{\delta}$  and  $\hat{\vartheta}$ , which are asymptotically efficient (see Goldstein (1986) for full discussion). Efficient computational procedures for the estimation of the parameters in the general class of multilevel models are provided by Goldstein and Rasbash (1992).

Under multivariate normality the above estimates are maximum likelihood estimates. Goldstein (1989) shows how unbiased estimates, equivalent to restricted maximum likelihood estimates under multivariate normality, can be obtained using what is termed restricted iterative generalised least squares estimation.

## 2.3 Extensions

Model (1) specifies a simple two-level variance components model representing the most basic form of a multilevel model. Extensions to this model include the addition of further levels in the data hierarchy, random coefficient models as well as cross-classified hierarchical structures. All rely on the same principles of estimation iterating between (3) and (4) above. For example, for a three-level multilevel model, the design matrix, D, in (4) will contain an additional term representing the third level in the data hierarchy. Similarly, although, V will retain a block diagonal structure, within each block representing the third level will be nested further block diagonal sub-matrices representing the clustering of level-1 units with level-2 units.

4

Random coefficients at any level of the hierarchy may be incorporated as follows. Thus at level 2 we may have:

$$y_{ij} = X_{ij}^{T} \beta + Z^{T} \gamma + u_{0j} + u_{1j} x_{1ij} + e_{ij}, \qquad i = 1, \dots, N; \quad j = 1, \dots, M, \quad (5)$$

where  $u_{0j}$  represents a random intercept to be estimated by  $\sigma_{u0}^2$ , and  $u_{1j}$  a random slope with variance,  $\sigma_{u1}^2$ , and a covariance parameter (between slope and intercept)  $\sigma_{u01}$ . In matrix notation, defining the matrix  $\Omega_i$  as the covariance matrix for the set of level 1 random coefficients, and  $\Omega_2$  as the corresponding matrix for the level 2 random coefficients, the covariance matrix of responses for model (5) can be defined as:

$$X_{I}\Omega_{2}X_{I}^{T}+\Omega_{I}$$

where, for example, for two level 1 units in each level 2 unit, we have:

$$X_{l} = \begin{pmatrix} l & x_{lj} \\ l & x_{2j} \end{pmatrix}, \quad \boldsymbol{\Omega}_{2} = \begin{pmatrix} \boldsymbol{\sigma}_{u0}^{2} & \boldsymbol{\sigma}_{u0l} \\ \boldsymbol{\sigma}_{u0l} & \boldsymbol{\sigma}_{ul}^{2} \end{pmatrix}, \quad \boldsymbol{\Omega}_{l} = \begin{pmatrix} \boldsymbol{\sigma}_{e}^{2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\sigma}_{e}^{2} \end{pmatrix}$$

Goldstein (1995) describes fully the estimation of more complex multilevel models.

## 3. FIXED VERSUS RANDOM EFFECTS

Assuming, for the time being, that variables measured at a particular level are uncorrelated with the residuals at the same level, then in the absence of correlations between the *p* components of *X* and the level 2 random effects  $u_j$ , that is  $E(X_{pij}u_j) = 0$ ,  $\forall p$ , the IGLS estimator of (1) is an efficient and consistent estimator of the fixed and random part parameters for fixed  $n_j$  (Goldstein, 1986). However, where correlations between the level 2 random effects  $u_j$  and components

of X are non-zero for some p, although IGLS estimation is efficient it is inconsistent for  $\beta$  as  $M \rightarrow \infty$  when group sample sizes  $n_j$  are small (for example, see Blundell and Windmeijer (1997) for a full discussion relating to multilevel models). The inconsistency of the estimated parameters  $\hat{\beta}$ , will also be reflected in the IGLS estimator of the random parameters used to construct  $\hat{V}$ . Given that the majority of applied analyses using multilevel models concentrates heavily on the random quantities, the biases introduced here may be of more concern. Throughout this paper, we regard correlated effects biases in both the fixed and random part of the multilevel model as misspecification biases due to the exclusion of relevant terms on the righthand-side of the regression model.

In static panel data models the standard procedure used to assess misspecification biases of this sort is to apply the Hausman test (Hausman (1978)). This is based on a straight-forward comparison between the estimated parameters from a fixed effects regression (using the least squares dummy variable estimator (LSDV) or the covariance estimator (CV) and assuming correct model specification) and those obtained through GLS. Significant differences suggest that misspecification biases are present in the GLS estimation and that a fixed effects specification should be used. The same test can be applied directly to the multilevel model depicted as (1). We now discuss these estimators in detail.

# 4. CONSISTENT ESTIMATION OF MULTILEVEL MODELS USING FIXED EFFECTS

## 4.1. LSDV estimator

For model (1) if  $E(X_{pij}u_j) \neq 0$ , for all, or some p, GLS estimation yields biased and inconsistent estimates of the parameters  $(\beta_p, \sigma_u^2, \sigma_e^2)$ . Consistent estimation of  $\beta_p$  in (1) (as either N or  $M \to \infty$ ) can be achieved by re-specifying (1) and (2) as a fixed effects model, removing group variables and specifying dummy variables for group membership and estimating by OLS:

$$Y = X\beta + U\alpha + E^* \tag{6}$$

where U is a  $(N \times M)$  dimensioned matrix of dummy variables. However, LSDV estimation of  $\beta$  is not without problems; group level variables cannot be estimated since they are not now identifiable and estimation is not fully efficient compared to a random effects model.

## 4.2. Within groups (CV) estimator

Alternatively, for any matrix A, define the projection matrix  $P_A$  to be the projection onto the column space of A such that  $P_A = A(A^T A)^{-I} A^T$  (A is of full column rank). Also define  $Q_A = I - P_A$  to be the projection onto the space orthogonal to A. Let U be an  $(N \times M)$  matrix of group level dummy variables, so that  $P_u$  converts an  $(N \times I)$  vector such as y into a vector of group means, while  $Q_u$  converts it into a vector of deviations from group means. If we pre-multiply (6) by the idempotent matrix  $Q_u$ , we have:

$$Q_{\mu}Y = Q_{\mu}X\beta + Q_{\mu}E^{*}$$
<sup>(7)</sup>

Applying OLS to (7) provides a consistent estimator of  $\beta$ , termed  $\hat{\beta}_{W}$ . The estimator  $(X^{T}Q_{u}X)^{-l}(X^{T}Q_{u}Y)$  is known as the within groups or covariance estimator (CV).

As noted, both the LSDV and within estimator are consistent estimators of  $\beta$ . For both estimators group level coefficients,  $\gamma$ , are not identifiable. However, these may be retrieved in a two-step process. For example, Hausman and Taylor (1981) consider the estimation of a two-level variance components model using as a first step the within estimator from (7). From this, they compute residuals:

$$\hat{d}_{j} = Y_{j} - X_{j}\hat{\beta}_{W}$$
(8)

and regress these on the group level variables:

$$\hat{d}_j = Z\gamma + u_j \tag{9}$$

where  $u_j$  is a mean zero disturbances. If all elements of Z are uncorrelated with  $u_j$ , OLS will be consistent for  $\gamma$ . If, however, columns of Z are correlated with  $u_j$ , instruments may be found for Z, and estimation of (9) can proceed through two stage least squares (2SLS). If we have consistent estimates of  $\beta$  and  $\gamma$  then we can obtain consistent estimates of the variance components  $\sigma_u^2$  and  $\sigma_e^2$ . However, since  $\hat{d}_j$  in (9) is calculated from the residuals from a within regression, suggests that if the parameters  $\hat{\beta}_w$  are not fully efficient, then nor are the parameters  $\hat{\gamma}$ , unless a weighted estimation is performed.

# 5. CONSISTENT ESTIMATION OF MULTILEVEL MODELS USING INSTRUMENTAL VARIABLES

5.1. Instrumental variables estimators for two-level variance components model For random effects models where correlations with group level effects are known to exist, a commonly adopted approach to ensure consistent estimation of  $\hat{\beta}$  and  $\hat{\gamma}$  is IV estimation of (1) or (2). This involves finding suitable instruments for those columns of X and Z which are correlated with  $u_j$  and applying the IV estimators which have the following general form:

$$\hat{\delta} = \left[ W^{T} V^{-\frac{1}{2}} P_{A} V^{-\frac{1}{2}} W \right]^{-1} W^{T} V^{-\frac{1}{2}} P_{A} V^{-\frac{1}{2}} y$$

$$cov(\hat{\delta}) = \left( W^{T} V^{-\frac{1}{2}} P_{A} V^{-\frac{1}{2}} W \right)^{-1}$$
(10)

where  $P_A$  is the orthogonal projection operator onto the column space of the instrument set. In general though, it may be difficult to find appropriate instruments legitimately excluded from equation (1) which are good predictors of the endogenous

columns of X and Z, but uncorrelated with  $u_j$ . However, because  $u_j$ , the component of the error term assumed correlated with the columns of X and Z is invariant within groups, any vector which is orthogonal to it can be used as a potential instrument. In this respect the within estimator (7) can be represented as an IV estimator of  $\beta$ , calculated by projecting (6) onto the null space of U by the matrix  $Q_u$  (that is defining the instrument set as  $(Q_uX)$ . Unfortunately,  $Q_u$  is orthogonal not only with U but with all group level variables Z. Consequently, in the absence of further information concerning the relationships between the columns of X and Z and  $u_j$ , the parameters  $\gamma$  are not identifiable.

Given prior information on which columns of X and Z are uncorrelated with  $u_j$ (denoted  $(X_1, Z_1)$ ) and which columns are correlated with  $u_j$  (denoted  $(X_2, Z_2)$ ), (or by investigation - using the Hausman test for example), various instrumental variable estimators based on (10) have been proposed which are potentially more efficient that the within estimator.

Instrumental variable estimators for variance components panel data models have been proposed by Hausman and Taylor (1981) (*HT*), Amemiya and MacCurdy (1986) and Breusch, Mizon and Schmidt (1986) (*BMS*). A general framework for such estimators is provided by Arellano and Bover (1995). Cornwell and Rupert (1988) and Baltagi and Khanti-Akom (1990) provide comparisons of the relative efficiency gains of IV estimators applied to a returns to schooling example. Below, we summarise briefly these estimators adopting the notation used by Cornwell and Rupert (1988).

*HT* propose estimation of (10) where the list of instruments for the set of endogenous variables  $(X_2, Z_2)$  is defined as  $A = (Q_u, X_1, Z_1)$ , which *BMS* show to be equivalent to the following instrument sets;  $A_{HT} = (Q_u X_1, Q_u X_2, X_1, Z_1)$ , or  $A_{HT} = (Q_u X_1, Q_u X_2, P_u X_1, Z_1)$  (they span the same column space as *A*). When the parameters of (1) are identified by means of this set of instruments, the *HT* estimator

provides a consistent and asymptotically efficient estimator for  $(\beta, \gamma)$  (see *HT* for details).

Identification of the *HT* estimator relies on the relative number of exogenous within group varying variables  $k_1$ ; and the number of endogenous group level (therefore invariant within groups) variables  $g_2$ . If the model is just-identified ( $k_1 = g_2$ ), then  $\hat{\beta}^* = \hat{\beta}_W$ , and  $\hat{\gamma}^* = \hat{\gamma}_W$  (note that  $\hat{\gamma}_W$  is defined by  $\hat{\gamma}_W = (Z_j^T P_A Z_j)^{-1} Z_j^T P_A \hat{d}_j$  where  $\hat{d}_j$  are residuals from a within regression (see equation (9)). If the model is over-identified ( $k_1 > g_2$ ), then more efficient estimates ( $\hat{\beta}^*, \hat{\gamma}^*$ ) may be obtained.

Assuming a balanced design  $(n_j = n_l = n, \forall j \neq l)$ , *AM* define a potentially more efficient estimator than *HT* defining the instrument set  $A_{AM} = (Q_u X_1, Q_u X_2, X_1^*, Z_1)$ . Here  $X_1^*$  is an  $N \times nK$  matrix where each column contains values of  $X_{1ij}$  for each within-group observation. For example, for the *p*th observation within each group, the constructed column is given by  $X_{1p}^* = (X_{11p}, ..., X_{11p}, ..., X_{1Mp}, ..., X_{1Mp})$ . Whilst *HT* use each  $X_1$  variable as two instruments, *AM* use each of these variables as (p+1) instruments  $(Q_u X_1, and X_1^*)$ . For identification,  $pk_1 \ge g_2$ .

The exogeneity assumptions for the AM estimator are stronger than those for the HT estimator. Whilst the HT estimator only requires the means of the  $X_1$  variables to be uncorrelated with the effects, the AM estimator requires uncorrelatedness for each individual observation.

In a direct extension of *AM*'s treatment of the  $X_1$  variables, to the  $X_2$  variables, *BMS* derive a potentially more efficient estimator by defining the following instrument set:  $A_{BMS} = (Q_u X_1, Q_u X_2, P_u X_1, (Q_u X_1)^*, (Q_u X_2)^*, Z_1)$ . The potential efficiency gain from the *BMS* procedure depends on whether the  $(Q_u X_2)^*$  are legitimate instruments.

They are valid if the variables in  $X_2$  are correlated with the effects only through a within-groups component which would be removed by the transformation  $Q_u X_2$ .

Therefore, for known V, the application of asymptotically efficient estimators is straightforward. For unknown V, variance components can by obtained through some consistent estimator, for example, the approach adopted by Hausman and Taylor (1981) described above.

### 5.2. Instrumental variables estimators for a general class of multilevel models

The *IV* methods described above can readily be extended to the general class of multilevel models. For a two-level variance components model, the generalisation of the methods is trivial and consistent estimation of both the fixed predictors and the random components can be achieved through iterating between (10) and (4). Starting values may be obtained from *IV* estimation of (10) assuming  $V = \sigma^2 I$ . For random coefficient models, where a parameter of an explanatory variable is assumed to vary randomly at, say, level 2, application of *IV* will depend on whether the variable assumed random is exogenously defined. If it is, then *IV* of (10) can be performed followed by estimation of *V* by (4), by imposing an appropriate design matrix linking  $Y^*$  to *V*. Where the random coefficient is endogenous, iterative IV of both (10) and (4) (again after imposing an appropriate design matrix) will be necessary.

The panel data *IV* approaches proposed by *HT*, *AMS* and *BMS* are easily extended to variance components multilevel models where there are more than two levels in the data hierarchy. For example, consider the following three-level multilevel model:

$$y_{ijk} = X_{ijk}^T \beta + Z_{jk}^{\prime T} \gamma + v_k + u_{jk} + e_{ijk} , \quad i = 1, ..., N; \ j = 1, ..., M; \ k = 1, ..., L$$
(11)

where, the terms  $e_{ijk}$ ,  $u_{jk}$ , and  $v_k$  represent residuals at level 1, 2 and 3 respectively, with  $e_{ijk} \sim N(0, \sigma_e^2)$ ,  $u_{jk} \sim N(0, \sigma_u^2)$ , and  $v_k \sim N(0, \sigma_v^2)$ . Following Hausman and Taylor we can partition both X and Z into columns uncorrelated with these random effects; these can be denoted  $(X_1, Z_1)$ . We can also define columns of X and Z correlated with  $u_{jk}$  by  $(X_2^u, Z_2^u)$ , and those correlated with  $v_k$  by  $(X_2^v, Z_2^v)$ . Generalising the *HT* instrument set to this example, we can define the following set of instruments:  $IV = (Q_u X_1, Q_u X_2^u, Q_v X_1, Q_v X_2^v, Q_v Z_2^v, X_1, Z_1)$ . Again, by defining  $P_A$  as the projection matrix onto the column space of the instrument set *IV*, estimation proceeds through iterating between (10) and (4). Potentially more efficient estimation could be obtained by applying the principles of *AM* and *BMS* to *IV* of model (11).

# 6. CONSISTENT ESTIMATION OF MULTILEVEL MODELS BY CONDITIONING ON GROUP VARIABLES

### 6.1. Mundlak's formulation

Mundlak (1978) considers a random effects specification of model (1) and represents the joint distribution of the explanatory variables  $X_{ij}$  and the effects  $\mu_j$  by approximating  $E(\mu_j|X_{ij})$  by an auxiliary linear regression based on within group means  $(\overline{X}_i)$ :

$$\mu_j = \overline{X}_j^T \eta + w_j, \qquad \overline{X}_j = \frac{1}{n_j} \sum_i X_{ij}, \qquad w_j \sim N(0, \sigma_w^2) \qquad (12)$$

Here, Mundlak assumes that the group or level 2 effects are a linear function of the averages of all the explanatory variables across each level 2 unit. Substituting (12) into (1) gives:

$$y_{ij} = X_{ij}^{T} \beta + \overline{X}_{j}^{T} \eta + w_{j} + e_{ij} \qquad i = 1, ..., N; \quad j = 1, ..., M$$
(13)

where 
$$E(e_{ij} + w_j) = 0$$
,  $E(e_{ij} + w_j)(e_{ij} + w_j)^T = \begin{cases} \sigma_e^2 I_{n_j} + \sigma_w^2 J_{n_j}, & j = j', \\ 0, & j \neq j' \end{cases}$ 

and  $I_{n_j}$ ,  $J_{j_j}$  are respectively the identity matrix and the square matrix of ones, of order  $n_j$ .

Mundlak shows that the generalized least squares (GLS) estimator of the  $\beta$  in (13) is identical to the fixed effects estimator (6). 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  $corr(X_{ij}, \mu_j) = 0$ , leading to  $\eta = 0$ , does (13) reduce to (1) such that GLS estimation of  $\beta$  in (1) is equivalent to OLS estimation of  $\beta$  in (6).

# 7. CONSISTENT ESTIMATION OF MULTILEVEL MODELS BY CONDITIONING ON GROUP EFFECTS

## 7.1. Conditioned iterative generalized least squares (CIGLS)

As noted earlier, consistent estimation of  $\beta$  can be achieved by specifying the model as a fixed effects model (6) and estimating by OLS. By conditioning on group variables, estimation is purged of correlation between explanatory variables and group effects. By exploiting the iterative nature of the IGLS estimator, we now show how a constructed variable, representing group level effects can be used to condition on whilst estimating the parameters,  $\beta$ . By including this constructed variable in the estimation of the fixed part parameters using (3), but removing it from the estimation of random parameters using (4), we derive an estimator which is both consistent and asymptotically efficient and has the advantage of being readily generalizable to the general class of multilevel models.

Consider re-expressing (1) excluding group level variables:

$$y_{ij} = X_{ij}^{T} \beta + u_j + e_{ij} \qquad i = 1, \dots, N; \quad j = 1, \dots, M \qquad (14)$$

or in matrix notation:

$$Y = X\beta + E \tag{15}$$

where, once again, where *E* is a block-diagonal error term such that  $E = Z_u u + Z_e e$ with the dimensions of the vectors  $Z_u$  and  $Z_e$  being  $(N \times I)$ . We now consider an alternative estimation procedure based upon a conditioned version of IGLS estimation, and term this CIGLS. For each iteration, IGLS estimation may be viewed as a two step procedure estimating the fixed and random parameters through (3) and (4) respectively. In the first step we re-express model (6) in matrix notation as:

$$Y = X\beta_{X} + S + E$$
where  $E = \{e_{ij}\}$  and (16)

$$S = \{S_{j}\}, \quad S_{j}^{T} = (s_{1}, \dots, s_{M}), \quad s_{j} = s_{j}^{*} \times t_{n_{j}}, \quad s_{j}^{*} = \sum_{i=1}^{n_{j}} \hat{y}_{ij} / n_{j},$$
  

$$\hat{Y} = \{\hat{y}_{ij}\} = Y - X\hat{\beta}_{X}^{*}$$
(17)

where  $t_{n_j}$  is a vector of ones of length  $n_j$ ,  $\hat{\beta}_X^*$  is the current estimate of  $\beta_X$  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 raw residuals from the previous iteration. Once *S* is constructed, updated estimates of  $\beta_X$  are then obtained through GLS estimation using (3). Note that this is equivalent to treating *S* as an offset in (16) such that the transformed response is (Y - S).

In the second step, we condition on  $\hat{\beta}_X$  and form the matrix  $Y^*$ , the upper trianglular elements of  $(Y - X\hat{\beta}_X)(Y - X\hat{\beta}_X)^T$ . By stacking the columns of  $Y^*$  these are regressed on the random parameter design matrix and applying GLS in (4), estimates of the random parameters,  $\sigma_e^2$  and  $\sigma_u^2$ , can be obtained.

Suitable starting values for  $\hat{\beta}_{X}^{*}$  may be obtained by OLS or IGLS estimation of (14). Iteration of the two steps proceeds to convergence defined by a pre-assigned tolerance for  $(\hat{\beta}_{X} - \hat{\beta}_{X}^{*})$ .

#### 7.2. Equivalence of CIGLS and CV estimation

We can re-express S in (16) as

$$S = (I - Q_u)(Y - X\hat{\beta}_X^*)$$
(18)

where  $Q_u$  is defined in section 4.2 and *I* is the identity matrix. It then follows directly from (16) that

$$Y = X\beta_{X} + ((I - Q_{\mu})(Y - X\hat{\beta}_{X}^{*})) + E$$
(19)

If at convergence we have  $\hat{\beta}_{X} = \hat{\beta}_{X}^{*}$ , (16) reduces to

$$Q_u Y = Q_u X \beta_X + E \tag{20}$$

which is equivalent to the within groups or CV specification (7) with  $E = Q_u E^*$ 

It follows immediately therefore that the GLS estimator of the parameters of (16), namely  $(X^T V_E^{-1} X)^{-1} (X^T V_E^{-1} Y)$ , where  $V_E$  is the block diagonal covariance matrix, provides both an efficient and *consistent* (maximum likelihood under Normality) estimator of  $\beta_X$ .

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 (6) or (16) by  $Q_u$ .

## 7.3. Extensions using CIGLS

### 7.3.1. Group level variables

An obvious extension to (14) is to consider explanatory variables measured at the group level, such as model (1). Where model (1) is true and group level variables *Z* are orthogonal to  $u_j^*$ , estimation using CIGLS proceeds in a straightforward manner and we construct S as in (17), except now  $\hat{Y} = \{\hat{y}_{ij}\} = Y - (X\hat{\beta}_X^* + Z\hat{\beta}_Z^*)$ , where  $\hat{\beta}_X^*$ 

and  $\hat{\beta}_{Z}^{*}$  represent current estimates. Where correlations between group level variables Z and  $u_{j}^{*}$  exist, estimation using CIGLS will result in biased and inconsistent estimates of  $\beta_{Z}$ . This is an example of omitted variable bias and IV estimation techniques will be required. Section 8 of this paper considers an example of the income returns to schooling where the response of interest is a measure of wage rate and the data consist of a panel of repeated yearly observations on individuals. A problem with this analysis is that individual ability and years spent in education are thought, a priori, to be correlated with one another and also with wage rates. This leads to a potential problem of correlation between a group level variable (educational achievement) and the group level error term (individual specific variability) described above. The solution offered is to instrument for educational achievement.

# 7.3.2. Group level effects and their interpretation

Suits (1984) and Kennedy (1986) discuss the difficulties in interpreting model (6) 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,  $\beta_0$ . In the case where all groups have equivalent populations,  $\beta_0$  may be interpreted as the population average. Where within group population sizes differ,  $\beta_0$  represents a weighted population average (Kennedy (1986)). The estimation of S in (16) 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.

# 7.3.3. Random coefficient models

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 which may 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.

The second extension is where we have random coefficients at level 2 and the algorithm is modified as follows. As before, we 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 (14);  $u_j = v_j + x_{ij}\lambda_j$ , where effects  $v_j$  represent an intercept parameter and  $\lambda_j$  a slope parameter, 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):

$$\hat{w}_{ij} = s_0 + s_1 x_{ij} + e_{ij}^{**} \tag{21}$$

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  $t_{n_j}$  by  $\hat{s}_0$  and  $\hat{s}_1$  and stacking the resulting vectors. We then carry out GLS estimation for the model

$$Y = X\beta_{x} + S_{0} + S_{1}^{*} + E^{**}$$
(22)

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

If we re-express (21) as an OLS regression across all groups we have:

$$\hat{w}_{ij} = \sum_{j=1}^{M} \alpha_j d_j + \sum_{j=1}^{M} \xi_j \psi_j + e_{ij}^{**}$$
(23)

where  $\psi_j = x_{ij}d_j$  and  $\{d_j\}$  is a set of *M* dummy variables. The corresponding coefficients are represented by  $\xi_j$ . We can retrieve the constructed vector  $S_0$  by multiplying the vectors  $t_{n_j}$  by the estimated coefficients  $\alpha_j$  and stacking the resulting vectors.  $S_i^*$  can be retrieved in a similar way by stacking the vectors obtained by

multiplying the vectors  $t_{n_j}$  by the set of coefficients  $\xi_j$ . The equivalent LSDV estimator is obtained as follows.

In matrix form (and using the multilevel notation) the equivalent model formed by combining equation (23) with the set of fixed part predictor variables of interest may be written as:

$$Y = X\beta_X + D\alpha + D^*\xi + E^{**}$$
(24)

where Y and  $E^{**}$  are  $(N \times I)$  dimensioned vectors,  $\beta$  is a  $(P \times I)$  vector and both  $\alpha$  and  $\xi$  are vectors of dimension  $(M \times I)$ . X and D are  $(N \times P)$  and  $(N \times M)$  dimensioned matrices respectively, and  $D^*$  is an  $(N \times M)$  matrix formed by multiplying the dummy variable matrix D by the random coefficient vector  $X^T$ , that is  $D^* = X^T D$ .

Model (24) 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 (21) 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.

## 7.3.4. Higher levels

CIGLS is not restricted to the simple case of a two level hierarchy and the 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.

### 8. EXAMPLE

# 8.1. Two-level variance components model: Returns to schooling example of Cornwell and Rupert (1988).

The following example is based on data used by Cornwell and Rupert (1988) and Baltagi and Khanti-Akom (1990) in their respective analyses of future income returns to schooling and represents a subset of the Panel Study of Income Dynamics (PSID) consisting of a panel of 595 individuals observed over the years 1976 to 1982.

The analyses of Cornwell and Rupert and Baltagi and Khanti-Akom were primarily concerned with an investigation into the relationship between income returns from employment and schooling experience where the response was a measure of wage rate. Whilst some of the potential predictors of income were time varying, others were not and a random effects specification was adopted. As the authors highlight, a potential problem with such an analysis is that individual ability or achievement and years spent in education are likely to be highly correlated with one another and ability may also be highly correlated with wage rates. If conditioning on years spent in education does not remove the relationship between ability and wage rates then we may obtain an inconsistent estimator of the years spent in education coefficient. By instrumenting for this variable, and applying the estimators described in section 5, the authors were able to obtain consistent parameter estimates and assess the relative efficiency gains of the AM and BMS estimators over the HT estimator. Although primary interest was with the estimated parameter of the years spent in education variable, various time varying variables were also specified as endogenous and instrumental variables applied accordingly.

Full details of the data used can be found in Cornwell and Rupert. The response variable is log wage and the set of explanatory variables are as follows: years spent in education (ED); years of full-time work (EXP); weeks worked (WKS); occupation (OCC = 1, if blue collar occupation, = 0 otherwise); industry (IND = 1, if individual works in manufactoring industry, = 0 otherwise); residence (SOUTH = 1, SMSA = 1,

if individual resides in the south or metropolitan areas respectively, = 0 otherwise); marital status (MS = 1 if individual is married, = 0 otherwise); union coverage (UNION = 1,if wage is set by union contract, = 0 otherwise); gender (FEM = 1 if female; 0 otherwise); race (BLK = 1 if black ethnic origin, = 0 otherwise). FEM, BLK and ED are time invariant, all other variables are time varying.

We adopt the model specification of Baltagi and Khanti-Akom such that OCC, SOUTH, SMSA and IND are assumed exogenous time varying regressors and FEM and BLK are assumed exogenous time-invariant regressors. To estimate the proposed models we apply the multilevel IGLS estimator, the within estimator, the *HT*, *AM* and *BMS* instrument set applied to the IGLS estimator, the CIGLS estimator and the *HT* instrument set for the endogenous ED variable applied to the CIGLS estimator. All models included time dummies to capture productivity and price level effects (the corresponding parameter estimates are not shown). Mundlak's specification was not included since aggregating the time dummies would lead to collinearity.

The results for the above estimators are given in Table 1. The IGLS estimates differ from the GLS results reported by Baltagi and Khanti-Akom. This is due to differences in estimated variance components. Baltagi and Khanti-Akom use  $\hat{\sigma}_{\varepsilon}^3 = 0.023$ , and  $\hat{\sigma}_{\nu}^3 = 0.256$  throughout. We estimate both the fixed part parameters and the variance components iteratively and the inconsistency of the fixed part parameter estimates will be fed into the estimates of the variance components. The main differences are observed in the estimates for OCC, SOUTH, SMSA, and MS. These are respectively, 1.5, 1.7, 5.3 and 0.6 times the size reported by Baltagi and Khanti-Akom.

The estimates of the endogenous education variable ED change quite dramatically moving across the results of Table 1. The inconsistent IGLS estimator of ED is 0.0663, the HT-IGLS estimate is 0.2241, the AM-IGLS estimate is 0.1472 and the BMS-IGLS estimate is 0.0878. The corresponding standard errors also decrease sharply moving from HT-IGLS to BMS-IGLS. The CIGLS estimates coincide with the Within estimates but the efficiency gains are apparent from the decreased standard errors. The increased efficiency of CIGLS over the instrumental variables estimators are also notable for the coefficient of EXP. CIGLS is a consistent estimator of the

time-varying variables, but remains an inconsistent estimator of the time-invariant ED variable.

The final three models of Table 1 combine the IV estimators of *HT*, *AM* and *BMS* with CIGLS by instrumenting for ED alone and conditioning on group level effects using CIGLS. The time-varying parameter estimates of the HT-CIGLS, AM-CIGLS and BMS-CIGLS estimators coincide with those obtained using CIGLS, however, in general, there are slight efficiency gains moving from HT-CIGLS to BMS-CIGLS albeit very modest. The instrumental variable estimates of ED are 0.0973, 0.09316 and 0.0742 respectively. For *HT* and *AM*, these differ from the results of HT-IGLS and AM-IGLS and are closer to the results obtained by CIGLS. Again, gains in efficiency can be observed moving from HT-CIGLS to BMS-CIGLS. These gains are also observed comparing IGLS with instrumental variables and the corresponding CIGLS with instrumental variables. As observed by Baltagi and Khanti-Akom the gains in efficiency are not limited to the ED variable, but can also be seen in the EXP estimate, although this trend does not persist beyond CIGLS.

Cornwell and Rupert and Hausman and Taylor (1981) note that the estimated schooling coefficient increases when using instruments over the estimate obtained when not controlling for the correlation between the explanatory variables and the individual effects. This is also true here, where all IV estimators using IGLS provide larger coefficient estimates for ED, and similarly IV with CIGLS provides increased estimates over CIGLS alone. CIGLS also produces larger estimates than IGLS.

## 9. DISCUSSION

In this paper we have shown how consistent estimation of the coefficients of endogenous variables can be obtained for a general class of multilevel models. Standard methods for random effects panel data models rely heavily on instrumental variables techniques and their application to multilevel models are also described. By conditioning on estimated group means, a conditioned estimator of parameters (CIGLS) is derived. The efficiency gains and finite sample biases of these approaches are compared in an example of the income returns to schooling analysed by Cornwell

21

and Rupert (1988) and subsequently Baltagi and Khanti-Akom (1990). The efficiency gains using CIGLS are not limited to the estimated coefficients of the group-invariant variables, but are also present in group-varying parameter estimates. Efficiency gains are also noted using CIGLS estimation over instrumental variables estimation. For group-varying variables, CIGLS estimates coincide with the parameter estimates obtained using the Within estimator. The methods described are applicable to the full range of multilevel models, including many levels in the data hierarchy and complex residual heterogeneity. Macros enabling CIGLS estimation within software specifically designed for multilevel modelling (*MLwiN*: Rasbash, J., Browne, W., Goldstein, H., Yang, M., et al. (1999)) are available from the authors. At present the method is only applicable to models with linear link functions, although research into the applicability of the procedures to non-linear models is underway.

	CIGLS	5.226	(0.0822)	-0.0192	(0.0129)	0.0031	(0.0216)	-0.0419	(0.0161)	0.0208	(0.0135)	0.0267	(0.0025)	-0.0004	(0.00005)	0.0007	(0.0006)	-0.0286	(0.0177)	0.0295	(0.0133)	-0.4267	(0.0427)	-0.1469	(0.0489)	0.07223	(0.0048)
ABLE I	BMS* -IGLS	4.855	(0.1350)	-0.0272	(0.0135)	-0.0408	(0.0219)	0.0251	(0.0161)	0.0284	(0.0138)	0.0340	(0.0042)	-0.0004	(0.00005)	0.0009	(0.0006)	-0.0232	(0.0180)	0.0467	(0.0133)	-0.4031	(0.0435)	-0.1464	(0.0495)	0.0878	(0.0075)
Ξ	AM -IGLS	4.059	(0.2850)	-0.0215	(0.0137)	-0.0139	(0.0252)	-0.0109	(0.0180)	0.0219	(0.0153)	0.0370	(0.0057)	-0.0004	(0.00005)	0.0009	(0.0006)	-0.0243	(0.0184)	0.0279	(0.0147)	-0.4061	(0.0535)	-0.0734	(0.0663)	0.1472	(0.0204)
	HT - IGLS	3.000	(0.5024)	-0.0192	(0.0137)	0.0119	(0.0292)	-0.0394	(0.0193)	0.0211	(0.0154)	0.0415	(0.0109)	-0.0004	(0.00005)	0.0008	(0.0006)	-0.0258	(0.0188)	0.0267	(0.0148)	-0.4043	(0.0796)	0.0166	(0.1070)	0.2241	(0.0404)
	Within	4.667	(0.0488)	-0.0192	(0.0137)	0.0031	(0.0342)	-0.0419	(0.0194)	0.0208	(0.0154)			-0.0004	(0.00005)	0.0007	(0.0006)	-0.0286	(0.0189)	0.0295	(0.0149)						
	IGLS	5.249	(0.0791)	-0.0426	(0.0128)	-0.0581	(0.0208)	-0.0419	(0.0155)	0.0280	(0.0133)	0.0277	(0.0024)	-0.0004	(0.00005)	0.0009	(0.0006)	-0.0165	(0.0177)	0.0429	(0.0131)	-0.4243	(0.0407)	-0.1509	(0.0462)	0.0663	(0.0046)
		Constant		000		SOUTH		SMSA		DNI		EXP		EXP2		WKS		MS		UNION		FEM		BLK		ED	

BMS\* -CIGLS

AM -CIGLS

HT - CIGLS

4.900

5.200 (0.1090) -0.0193

 $\begin{array}{c} (0.0135)\\ 0.0028\\ 0.0028\\ (0.0218)\\ -0.0418\\ (0.0162)\\ 0.0207\\ 0.0267\\ 0.0267\\ (0.0136)\\ -0.0005\\ 0.0005\\ 0.0007\\ 0.0006\\ 0.0177\\ (0.0006)\\ -0.0286\\ (0.0177)\\ 0.0006\\ 0.0177\\ 0.0043\\ (0.0173)\\ -0.1441\\ (0.0433)\\ 0.073\\ (0.0073)\\ 0.073\\ (0.0073)\\ \end{array}$ 

4.957 -0.0191 (0.2309) -0.0127) 0.0033 0.0228 -0.0420 0.0174) 0.0174) 0.0256 (0.0174) 0.0256 -0.0005 0.0265 -0.0005 0.0006 -0.0005 0.0285 -0.0005 0.0006 0.0285 -0.0005 0.0006 0.0285 (0.0178) 0.0006 0.0285 (0.0178) 0.0006 0.0285 (0.0178) 0.0006 0.0285 (0.0178) 0.0006 0.0285 (0.0178) 0.0006 0.0285 (0.0178) 0.0006 0.0285 (0.0178) 0.0006 0.0285 (0.0178) 0.0006 0.0285 (0.0178) 0.0006 0.0285 (0.00178) 0.0006 (0.0178) 0.0006 (0.0178) 0.0006 (0.0007 (0.0007 (0.0005) 0.0006 (0.0007 (0.0005) 0.0006 (0.0007 (0.0005) 0.0006 (0.0007 (0.0005) 0.0006 (0.0007 (0.0005) (0.0007 (0.0005) (0.0005) (0.0006 (0.0005) (0.0006 (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0006 (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0005) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055) (0.0055)

 $\begin{array}{c} (0.3559)\\ -0.0191\\ (0.0126)\\ 0.0041\\ (0.0242)\\ -0.0425\\ -0.0425\\ (0.0135)\\ 0.0265\\ (0.00135)\\ -0.0006\\ -0.0284\\ (0.0006)\\ -0.0284\\ (0.0006)\\ -0.0284\\ (0.00133)\\ -0.0113\\ 0.0255\\ (0.0025)\\ 0.0073\\ (0.0013)\\ -0.0113\\ (0.0622)\\ 0.0973\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.0275)\\ (0.027$ 

 $X_1 = (WKS, SOUTH, SMSA, MS); Z_1 = (FEM, BLK)$ Notes:

0.0850 0.0228

0.0883 0.0228

0.0897 0.0228

0.0850 0.0228

0.1284 0.0229

0.2560 0.0228

0.0756 0.0230

 $\sigma_{0}^{7}$  $\sigma_{E}^{7}$ 

0.0230

0.0229 0.0841

\* Removes EXP and EXP2 from the set of additional instruments allowed by the BMS procedure. <sup> $\perp$ </sup> Includes HT instruments for endogenous group level education variable ED.

All regressions include dummy variables to represent time effects.

23

### ACKNOWLEDGEMENTS

Nigel Rice was supported by the ESRC Visiting Fellowship Scheme for the ALCD Programme 1997/98: award reference number: H519 44 5003 97. We are grateful to Professor Christopher Cornwell for allowing us access to the Returns to Schooling data.

## **REFERENCES**

- Amemiya, T., MacCurdy, T.E., Instrumental variables estimation of an error components model, *Econometrica*, 1986; **54**: 869-880.
- Arellano, M., Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, **68**; 29-51.
- Baltagi, B.H., Khanti-Akom, S. (1990). On efficient estimation with panel data: An empirical comparison of instrumental variables estimators. *Journal of Applied Econometrics*, 5: 401-06.
- Blundell, R. and Windmeijer, F. (1997). Correlated cluster effects and simultaneity in multilevel models. *Health Economics*; 1: 6-13.
- Breusch, T.S., Mizon, G.E., Schmidt, P. (1989). Efficient estimation using panel data. *Econometrica* **57**: 695-700.
- Cornwell, C., Rupert, P. (1988). Efficient estimation with panel data: An empirical comparison of instrumental variables estimators. *Journal of Applied Econometrics*, **3**: 149-55.
- Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalised least squares. *Biometrika*, **73**: 43-56.
- Goldstein, H. (1989). Restricted unbiased iterative generalized least squares estimation. *Biometrika*, **76**: 622-3.
- Goldstein, H., Rasbash, J. (1992). Efficient computational procedures for the estimation of parameters in multilevel models based on iterative generalized least squares. *Computational Statistics and Data Analysis*, **13**: 63-71.
- Goldstein, H. (1995). *Multilevel Statistical Models (second edition)*. London: Edward Arnold.
- 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.
- Kennedy, P. (1986). Interpreting dummy variables. *The Review of Economics and Statistics*, **68**: 174-5.
- Mundlak, Y. (1978). On the pooling of time series and cross section data, *Econometrica* **46**: 69-81.
- Rasbash, J., Browne, W., Goldstein, H., Yang, M., et al. (1999). A user's guide to *MLwiN* (Second Edition). London, Institute of Education.
- Suits, D.B. (1984). Dummy variables: mechanics v. interpretation. *The Review of Economics and Statistics*, **66**: 177-80.