Estimating Models for Panel Survey Data under Complex Sampling

Marcel D. T. Vieira

Universidade Federal de Juiz de Fora
Departamento de Estatística, Juiz de Fora, 36036-330, MG, Brazil
and

Chris J. Skinner

University of Southampton
Southampton Statistical Sciences Research Institute, Southampton, SO17 1BJ, United Kingdom

Abstract. Complex designs are often used to select the sample which is followed over time in a panel survey. We consider some parametric models for panel data and discuss methods of estimating the model parameters which allow for complex schemes. We incorporate survey weights into alternative point estimation procedures. These procedures include pseudo maximum likelihood (PML) and various forms of generalized least squares (GLS). We also consider variance estimation using linearization methods to allow for complex sampling. The behaviour of the proposed inference procedures are assessed in a simulation study, based upon data from the British Household Panel Survey. The point estimators have broadly similar performance, with few significant gains from GLS estimation over PML estimation. The need to allow for clustering in variance estimation methods is demonstrated. Linearization variance estimation performs better, in terms of bias, for the PML estimator compared to a GLS estimator.

Key words: longitudinal survey; covariance structure; multistage sampling; stratification; weighting.

Acknowledgments: This work was partially supported by the Brazilian National Council for Scientific and Technological Development (CNPq) grant 200286/01.3. We are grateful to the Associate Editor and five referees for comments which helped to improve the paper.

1. Introduction

A broad class of 'regression-type' models has found a wide range of useful applications with panel survey data (e.g. Wooldridge, 2001; Diggle *et al*., 2002). Such data often consist of repeated observations on the same variables for the same individuals across equally spaced waves of data collection. The 'regression-type' models considered here are broadly concerned with representing the relationship between one of the variables, treated as dependent, and a number of the other variables, treated as covariates. A typical example of the kind of panel survey considered here is the British Household Panel Survey (BHPS), in which a sample of households was selected at wave one and then individuals in this sample were followed up repeatedly at annual intervals.

It is common for the selection of the initial panel sample at wave one to involve a complex sampling scheme. For example, stratification and multistage sampling were employed in the selection of the initial BHPS sample. In addition, sample individuals are often selected with unequal probabilities and weights are constructed to compensate for these unequal probabilities as well as for different forms of wave nonresponse and other complexities (Kalton and Brick, 2000). In the mainstream panel data modelling literature there is little consideration of such sampling schemes other than through extensions of models to capture clustering effects (e.g. Wooldridge, 2001).

A number of methods has been developed in the survey sampling literature to take account of complex sampling schemes in the regression analysis of cross section survey data. See Chambers and Skinner (2003) for references. One broad approach which has increasingly been implemented in statistical software packages is pseudo maximum likelihood estimation (Skinner, 1989), where maximum likelihood point estimators are adapted using survey weights and the variances of these point estimators are estimated using survey sampling methods, such as Taylor series linearization.

In this paper we shall extend this broad approach to the estimation of panel data model parameters, allowing for complex sampling designs. We shall discuss methods of statistical inference for models with parametric assumptions about the covariance structure of errors over time. We shall incorporate survey weights into alternative point estimation procedures, including

maximum likelihood, generalized least squares and asymptotically distribution free (ADF) approaches. We shall also consider standard error estimation approaches using linearization methods to allow for complex sampling, and indicate connections with some established ADF methods. We shall adopt an aggregate modelling strategy (Skinner, Holt and Smith, 1989) rather than a multilevel covariance modelling approach. For developments of the latter approach see Müthén and Satorra (1995, Section 5).

Some previous work on estimation for panel data models under complex designs has been undertaken by Feder, Nathan and Pfeffermann (2000), who propose combining multilevel modelling, time series modelling and survey sampling methods; Sutradhar and Kovacevic (2000), where a generalised estimating equations approach is developed by considering an autocorrelation structure in a multivariate polytomous longitudinal survey data context; Skinner and Holmes (2003), who study two approaches for dealing with sampling effects, either considering the repeated observations as multivariate outcomes and adopting weighted estimators that account for the correlation structure, or considering a two-level longitudinal model and to modify weighting strategy proposed by Pfeffermann *et al.* (1998); and Skinner and Vieira (2007), who presented some empirical evidence that the variance-inflating impacts of complex sampling schemes can be higher for longitudinal analyses than for corresponding cross-sectional analyses.

This paper is organized as follows. The basic structure of the data and sample are described in Section 2. The models are given in Section 3. Point estimation methods, including weighted estimation of covariance matrices are reviewed in Section 4. Estimation of model parameters using least squares methods and pseudo maximum likelihood estimation are also considered. The paper proceeds in Section 5 to consider variance estimation methods, by adopting linearization methods to allow for complex sampling and also considering ADF variance estimation techniques. Two simulation studies, based upon data from the British Household Panel Survey, will be presented in Section 6 to assess the behaviour of the different estimation procedures. We make brief remarks in the concluding discussion in Section 7.

2. Sampling and Data

We suppose that the data consist of the values y_{it} of an outcome variable and $1\times q$ vectors of values $\mathbf{x}_{it}$ of covariates for each individual i in a sample, denoted s, and each wave of data collection $t=1,\ldots,T$. We shall sometimes write $s=\{1,...,n\}$, without loss of generality, where n is the sample size. The sample is assumed to be selected from a specified finite population at wave 1 according to a (without replacement) probability design for which the inclusion probability π_i of each individual i in s is known and the sample and the population are fixed thereafter. The design may be complex, for example involving stratification and multi-stage sampling. We suppose that sampling weights w_i are available for estimation. In the absence of nonresponse, these may be design weights, i.e. the reciprocals of the sample inclusion probablities π_i. In practice, nonresponse will occur at each wave, especially as a result of attrition. In this case, we suppose that s denotes the set of individuals providing values y_{it} and $\mathbf{x}_{it}$ at each of the T waves of data collection and we suppose that weights w_i are available that adjust not only for the sampling but also for the nonresponse.

3. Models

We consider standard kinds of models for the repeated measurements (Diggle *et al*., 2002, Chapters 4 and 5) in which the y_{it} obey the (superpopulation) linear model:

$$E(y_{it})=\mathbf{x}_{it}\boldsymbol{\beta}, \qquad (1)$$

in the population, where $\mathbf{x}_{it}$ is treated as fixed (or conditioned upon), $\boldsymbol{\beta}$ is a $q\times 1$ vector of unknown parameters (and we make no distinction between the realised y_{it} and the underlying random variables). We allow for serial correlation in the measurements by writing the repeated measurements for individual i as the $T\times 1$ vector $\mathbf{y}_i=(y_{i1},\ldots,y_{iT})'$ and allowing for non-zero off-diagonal elements of the covariance matrix Σ of this vector:

$$\Sigma=\mathrm{cov}(\mathbf{y}_i)=E\{[\mathbf{y}_i-X_i\boldsymbol{\beta}][\mathbf{y}_i-X_i\boldsymbol{\beta}]'\}, \qquad (2)$$

where $X_i = (\mathbf{x}_{i1}', \ldots, \mathbf{x}_{iT}')'$ is the $T \times q$ matrix of covariate values.

We consider two possible structures for the matrix Σ. The first is referred to as the uniform correlation model (UCM), where all the off-diagonal elements of Σ are σ_u^2 and all the diagonal elements are $\sigma_u^2 + \sigma_v^2$. This corresponds to the multilevel model:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + u_i + v_{it} \tag{3}$$

where u_i and v_{it} are random effects with zero means and variances σ_u^2 and σ_v^2 respectively, which are uncorrelated over time. In this case the correlation between y_{it} and $y_{it'}$ for any two occasions t and t' for $t \neq t'$ is given by $\rho = \sigma_u^2 / (\sigma_u^2 + \sigma_v^2)$.

In our second structure, referred to as the AR1 model, the correlation is allowed to decay over time. We again assume that all diagonal elements are $\sigma_u^2 + \sigma_v^2$ but now suppose that the covariance between y_{it} and $y_{it'}$ for occasions t and t' takes the form $\mathrm{cov}(y_{it}, y_{it'}) = \sigma_u^2 + \gamma^{|t-t'|}\sigma_v^2$, where γ is an additional parameter ($|\gamma| < 1$). This model corresponds to the following first-order autoregressive process for the v_{it}:

$$v_{it} = \gamma v_{it-1} + \varepsilon_{it}, \tag{4}$$

where the ε_{it} are mutually independent disturbances with zero mean and variance $\sigma_\varepsilon^2 = (1-\gamma)^2\sigma_v^2$. Note that in both models it is assumed that Σ does not depend upon *i*.

To emphasise the fact that the covariance matrix Σ takes a particular parametric structure for each model, we write $\Sigma = \Sigma(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a $b \times 1$ parameter vector. In particular, $\boldsymbol{\theta} = (\sigma_u^2, \sigma_v^2, \gamma)'$ for the AR1 model and $\boldsymbol{\theta} = (\sigma_u^2, \sigma_v^2)'$ for the UCM model. Note that the UCM model is a special case of the AR1 model where $\gamma = 0$.

We have so far only made assumptions about the correlation of the y_{it} between different time points *t* but not between different individuals *i*. We shall, indeed, assume that the parameter vector $\boldsymbol{\theta}$ governing the inter-temporal covariance matrix $\Sigma(\boldsymbol{\theta})$ is of scientific interest, but that any

correlation between values of y_{it} for different individuals is a 'nuisance'. In the UCM and AR1 models we shall assume that the correlation between y_{it} and $y_{i't'}$ is zero for any two distinct individuals i and i' and any two occasions t and t'. We shall also consider a UCM(C) model, where C denotes cluster, for which this correlation is given by a fixed quantity, τ, for any distinct individuals i and i' in the same cluster and any two occasions t and t' and zero otherwise, where the inter-temporal covariance structure $\Sigma(\boldsymbol{\theta})$ is the same as for the UCM model.

4. Point Estimation

We shall suppose that $\boldsymbol{\beta}$ is estimated following an established approach for repeated survey observations, as implemented for example in the software SUDAAN (Shah et al. 1997), by:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i \in s} w_i X_i' V^{-1} X_i \right)^{-1} \sum_{i \in s} w_i X_i' V^{-1} \mathbf{y}_i \qquad (5)$$

where V is a specified 'working' covariance matrix of $\mathbf{y}_i$ (Diggle et al. 2002, p.70) and the w_i are the survey weights introduced in section 2. Provided (a) the linear model in (1) holds, (b) the weights w_i have the property that weighted sample moments are consistent for population moments with respect to the joint sampling/nonresponse probability distribution, i.e. $\sum_s w_i z_i / \sum_s w_i$ is consistent for the finite population mean of z_i (an arbitrary variable) and (c) V is constant, $\hat{\boldsymbol{\beta}}$ will be consistent for $\boldsymbol{\beta}$ with respect to the joint model/sampling/nonresponse distribution as the sample size *n* increases (c.f. Fuller, 1975; Isaki and Fuller, 1982; Liang and Zeger, 1986).

Note that this result allows for the possibility that the sampling/nonresponse scheme is 'informative' with respect to the model, in the sense that the selection of individuals into the sample *s* is dependent upon y_{it} conditional on the $\mathbf{x}_{it}$. In this case, weighting by w_i (e.g. if they are inversely proportional to the probabilities of inclusion in *s*) in (5) may adjust for bias arising from such selection. In contrast, the omission of the weights from (5) could lead to bias in large samples in the presence of such selection.

In practice, condition (c) that V is constant will not hold. In the simulation study we shall suppose that V is estimated using the UCM model as the working model. This just requires estimating the intra-individual correlation ρ since $\sigma^2 = \sigma_u^2 + \sigma_v^2$ cancels out of the two places where V appears in (5). We shall estimate the correlation ρ by iterating between GLS estimation of $\boldsymbol{\beta}$ and survey-weighted moment-based estimation of the intra-individual correlation (Liang and Zeger, 1986; Shah *et al.*, 1997). Following standard large sample arguments (Liang and Zeger, 1986) $\hat{\boldsymbol{\beta}}$ will remain consistent for $\boldsymbol{\beta}$ when V is estimated in this way, even though there may be a loss of efficiency if the model underlying V is not well specified.

As in section 3, let $\boldsymbol{\theta}$ denote the $b \times 1$ vector of parameters of interest which determine the covariance structure $\Sigma = \Sigma(\boldsymbol{\theta})$ of $\mathbf{y}_i$, as given in (2). In order to define a class of estimators $\boldsymbol{\theta}$, we first define the weighted residual covariance matrix:

$$S_w = \hat{N}^{-1} \sum_{i \in s} w_i \left(\mathbf{y}_i - X_i \hat{\boldsymbol{\beta}}\right)\left(\mathbf{y}_i - X_i \hat{\boldsymbol{\beta}}\right)' \qquad (6)$$

where $\hat{N} = \sum_{i=1}^{n} w_i$ estimates the population size, N. The matrix S_w is a consistent estimator of Σ with respect to the joint model/sampling/nonresponse distribution, provided that the model assumptions in (1) and (2) hold and that the weights enable consistent estimation of population moments (condition (b) under equation (5)). Having defined S_w, we now define the class of estimators $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ to be considered, as those that minimise different measures of 'distance' between S_w and $\Sigma(\hat{\boldsymbol{\theta}})$ (Jöreskog and Goldberger, 1972). More precisely, if $F(S_w, \Sigma)$ denotes the fitting function, which measures the distance between S_w and Σ, then $\hat{\boldsymbol{\theta}}$ is defined as the value of $\boldsymbol{\theta}$ which minimises $F(S_w, \Sigma(\boldsymbol{\theta}))$ across values of $\boldsymbol{\theta}$ in a specified b-dimensional parameter space.

The simplest example of a fitting function is the *unweighted least squares* (ULS) function:

$$F_{ULS}(S, \Sigma) = \frac{1}{2} \cdot tr\{[S - \Sigma]^2\}. \qquad (7)$$

The resulting ULS estimator $\hat{\boldsymbol{\theta}}_{ULS}$ is uniquely defined and is consistent for $\boldsymbol{\theta}$, given that S (S_w in our setting) is consistent for Σ (Browne, 1982; Browne, 1984). However, $\hat{\boldsymbol{\theta}}_{ULS}$ is not in general an asymptotically efficient estimator of $\boldsymbol{\theta}$. Moreover, it is not scale invariant (Jöreskog and Goldberger, 1972) although this does not seem a serious problem when the elements of $\mathbf{y}_i$ are repeated measurements of the same variable. With the aim of improving efficiency, we consider also a class of *generalised least squares* fitting functions:

$$F_{GLS}(S,\Sigma) = \{vech(S) - vech(\Sigma)\}' \, \mathrm{U}^{-1} \{vech(S) - vech(\Sigma)\}, \qquad (8)$$

where *vech* is the vector of distinct elements of a symmetric matrix (Fuller, 1987). For the $T \times T$ matrices considered here, *vech* is of dimension $k \times 1$, where $k = T(T+1)/2$. The 'weight' matrix U remains to be specified. For efficient estimation, we should like U to correspond to (approximately) to the covariance matrix of $vech(S)$, for the relevant matrix S, which is S_w in our setting. A traditional approach to the specification of U, which ignores the complex sampling scheme and is motivated by a working assumption of normality and independent and identically distributed observations, is:

$$U = 2K(W \otimes W)K', \qquad (9)$$

where K is the so-called 'elimination' matrix, W is any consistent estimator of Σ, and $\otimes$ is the Kronecker product operator (Muthén and Satorra, 1995). Expression (9) may alternatively be written elementwise as (Joreskog and Goldberger, 1972):

$$U_{tt',t''t'''} = W_{tt''}W_{t't'''} + W_{tt'''}W_{t't''}, \qquad (10)$$

where $U_{tt',t''t'''}$ and $W_{tt'}$ represent typical elements respectively of U and W.

Expressions (8) and (9) imply (Browne, 1982) that $F_{GLS}(S,\Sigma)$ takes the form:

$$F_{GLS-NORM}(S,\Sigma) = \left(\frac{1}{2}\right) \cdot tr\{[(S-\Sigma)W^{-1}]^2\}, \qquad (11)$$

where GLS-NORM indicates that this choice of fitting function is based upon an underlying normality assumption. There are two natural choices of *W*. The first is given by *S*, since this (S_w in our setting) is assumed consistent for Σ. In this case we may write:

$$F_{GLS-NORM1}(S,\Sigma)=\left(\frac{1}{2}\right)\cdot tr\{[(S-\Sigma)S^{-1}]^2\}=\left(\frac{1}{2}\right)\cdot tr\{[I-\Sigma S^{-1}]^2\}. \qquad (12)$$

An alternative choice is to set *W* equal to Σ, leading to:

$$F_{GLS-NORM2}(S,\Sigma)=\left(\frac{1}{2}\right)\cdot tr\{[S\Sigma^{-1}-I]^2\}. \qquad (13)$$

We denote the resulting estimators of $\boldsymbol{\theta}$ as $\hat{\boldsymbol{\theta}}_{GLS-NORM1}$ and $\hat{\boldsymbol{\theta}}_{GLS-NORM2}$. An alternative approach, not based on the working assumption of normality, is to set *U* equal to an estimator of the asymptotic covariance matrix of $vech(S)$, making no assumption about the underlying distribution. Such an approach is often called *asymptotically distribution free* (ADF). See e.g. Browne (1982, 1984). We shall consider the use of linearization methods of variance estimation for this purpose in the next section, following some earlier applications of this idea in Skinner (1989), Satorra (1992), and Muthén and Satorra (1995).

Another approach to estimation is achieved by adopting the pseudo-maximum likelihood (PML) approach (Skinner, 1989) in which a census log-likelihood (assuming independent and identically distributed observations) is replaced by a weighted log-likelihood given by (ignoring constants):

$$-N\log|\Sigma(\boldsymbol{\theta})|-\sum_{i\in s}w_i[\mathbf{y}_i-X_i\boldsymbol{\beta}]'\Sigma(\boldsymbol{\theta})^{-1}[\mathbf{y}_i-X_i\boldsymbol{\beta}] \qquad (14)$$

If this weighted likelihood is first 'concentrated' by replacing $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$, maximising expression (14) becomes equivalent to minimising the value of the following fitting function:

$$F_{PML}(S,\Sigma)=tr\left[S\Sigma^{-1}\right]-\log\left|S\Sigma^{-1}\right|-T, \qquad (15)$$

with *S* evaluated at S_w to take account of the complex design and nonresponse. Alternatively, if this initial concentration does not take place, $\boldsymbol{\theta}$ could be estimated simultaneously with $\boldsymbol{\beta}$ by maximising expression (14). If *N* is unknown, it might be replaced in (14) by $\hat{N}=\sum_{i=1}^{n}w_i$.

The properties of the GLS-NORM1 and PML approaches may be compared by noting first that (12) may be alternatively expressed as (see Fuller, 1987, p. 334)

$$F_{GLS-NORM1}(S_w, \Sigma) = \frac{1}{2}(n-1)\sum_{t=1}^{T}(\lambda_t - 1)^2 ,$$

where $\lambda_1,\ldots,\lambda_t$ are the eigenvalues of $S_w^{-1/2}\Sigma S_w^{-1/2}$. Similarly, (15) may alternatively be expressed as

$$F_{PML}(S_w, \Sigma) = \sum_{t=1}^{T}(\log \lambda_t + \lambda_t^{-1}) .$$

Moreover if the model holds, i.e. if $\Sigma = \Sigma(\boldsymbol{\theta})$, both GLS-NORM1 and PML estimators are obtained by minimizing (see Fuller, 1987, p. 335) $\sum_{t=1}^{T}(\lambda_t - 1)^2$. Thus the GLS-NORM1 and PML estimators may be considered asymptotic equivalent.

5. Variance estimation

In this section, we consider variance estimation for two purposes: first, to determine possible matrices U to use in the generalised least squares fitting function in (8) and, secondly, for the purpose of estimating standard errors of the estimators of $\boldsymbol{\theta}$ considered in the previous section.

As a preliminary step, we consider estimation of the variances and covariances of the elements of S_w, i.e. we seek to estimate the asymptotic covariance matrix of the vector $vech(S_w)$. To establish the asymptotic covariance matrix with respect to the sampling design, nonresponse and the underlying model requires defining a sequence of populations, sampling designs/nonresponse mechanisms and samples. We suppose that this sequence is such that there exists a non-negative definite matrix C such that the limiting distribution of $\sqrt{n}\{vech(S_w) - vech(\Sigma)\}$ is normal with a mean vector consisting of zeros and covariance matrix, C (c.f. Isaki and Fuller, 1982), i.e.

$$\sqrt{n}\{vech(S_w) - vech(\Sigma)\} \rightarrow_L \mathrm{N}(0, C) . \qquad (16)$$

We seek an estimator of the asymptotic covariance matrix $n^{-1}C$. From (6), we may write

$$vech[S_w] = \left(\sum_{i=1}^{n} w_i\right)^{-1} \sum_{i=1}^{n} w_i \hat{\mathbf{c}}_i , \qquad (17)$$

where $\hat{\mathbf{c}}_i = vech(\hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i')$ and $\hat{\boldsymbol{\varepsilon}}_i = \mathbf{y}_i - X_i \hat{\boldsymbol{\beta}}$. In order to employ the linearization method of variance estimation, we first linearize expression (17) to obtain:

$$vech(S_w) \doteq \boldsymbol{\mu}_z / \mu_w + n^{-1} \sum_{i=1}^{n} \mathbf{u}_i \,, \tag{18}$$

where $\mathbf{u}_i = \mu_w^{-1} w_i (\mathbf{c}_i - \boldsymbol{\mu}_z / \mu_w)$, $\mathbf{c}_i = vech(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i')$, $\boldsymbol{\varepsilon}_i = \mathbf{y}_i - X_i \tilde{\boldsymbol{\beta}}$, $\boldsymbol{\mu}_z = E(n^{-1} \sum_{i=1}^{n} w_i \mathbf{c}_i)$, $\mu_w = E(n^{-1} \sum_{i=1}^{n} w_i)$ and $\tilde{\boldsymbol{\beta}} = p\lim(\hat{\boldsymbol{\beta}})$. A linearization estimator of the asymptotic covariance matrix of $vech(S_w)$ may then be obtained (Wolter, 2007) by constructing an estimator of the covariance matrix of the linear statistic $n^{-1} \sum_{i=1}^{n} \mathbf{u}_i$, allowing for the complex design, and then replacing $\mathbf{u}_i$ by $\hat{\mathbf{u}}_i = \overline{w}^{-1} w_i (\hat{\mathbf{c}}_i - \overline{\mathbf{z}} / \overline{w})$ where $\overline{w} = n^{-1} \sum_{i=1}^{n} w_i$ and $\overline{\mathbf{z}} = n^{-1} \sum_{i=1}^{n} w_i \mathbf{c}_i$.

Any feature of a complex design could, in principle, be handled in this linearization approach. Here, however, we only consider the case of a multistage stratified sampling scheme, where primary sampling units (PSUs) are sampled with replacement at the first stage within *H* strata independently, and sampling with or without replacement is used at subsequent stages. In this case, we rewrite $n^{-1} \sum_{i=1}^{n} \mathbf{u}_i$ as $n^{-1} \sum_{h=1}^{H} \sum_{j=1}^{m_h} \sum_{i=1}^{n_{hj}} \mathbf{u}_{hji}$, where the triple suffix refers to elements within PSUs within strata, m_h is the sample number of PSUs in stratum *h*, n_{hj} is the sample number of elements in PSU *j* in stratum *h*, and $\mathbf{u}_{hji}$ is the $k \times 1$ vector for element *i* in PSU *j* in stratum *h*. A standard estimator for the covariance matrix of $n^{-1} \sum_{h=1}^{H} \sum_{j=1}^{m_h} \sum_{i=1}^{n_{hj}} \mathbf{u}_{hji}$ under this sampling scheme, assuming the $\mathbf{u}_{hji}$ are observed and ignoring finite population corrections, is given by (Shah *et al.*, 1995)

$$\mathrm{v}_L \left[n^{-1} \sum_{h=1}^{H} \sum_{j=1}^{m_h} \sum_{i=1}^{n_{hj}} \mathbf{u}_{hji} \right]_{v,l} = n^{-2} \sum_{h=1}^{H} m_h \left\{ \left[\sum_{j=1}^{m_h} \left(\mathbf{u}_{hj+,v} - \overline{\mathbf{u}}_{h,v} \right)' \left(\mathbf{u}_{hj+,l} - \overline{\mathbf{u}}_{h,l} \right) \right] \Big/ (m_h - 1) \right\}, \tag{19}$$

where $\mathbf{u}_{hj+} = \sum_{i=1}^{n_{hj}} \mathbf{u}_{hji}$, $\bar{\mathbf{u}}_h = m_h^{-1} \sum_{j=1}^{m_h} \mathbf{u}_{hj+}$ and the subscripts v and l denote respectively $v = (t, t')$ and $l = (t'', t''')$. Finally, to obtain a linearization estimator $\mathrm{v}_L\{vech(S_w)\}$ of $\mathrm{var}\{vech[S_w]\}$, the values $\mathbf{u}_{hji}$ in (19) need to be replaced by values $\hat{\mathbf{u}}_{hji}$, defined in the same way that $\hat{\mathbf{u}}_i$ was defined above in terms of $\mathbf{u}_i$. The asymptotic validity of this variance estimator depends on each m_h being large if H is regarded as fixed.

In the special case when the population consists of only one stratum and each individual i is a PSU, we rewrite (19) as

$$\mathrm{v}_L\left[n^{-1} \sum_{i=1}^{n} \mathbf{u}_i \right]_{v,l} = \left[\sum_{i=1}^{n} \left(\mathbf{u}_{i,v} - \bar{\mathbf{u}}_v \right)' \left(\mathbf{u}_{i,l} - \bar{\mathbf{u}}_l \right) \right] \Big/ [n(n-1)]$$

where $\bar{\mathbf{u}} = n^{-1} \sum_{i=1}^{n} \mathbf{u}_i$. When $\mathbf{u}_i$ is replaced by $\hat{\mathbf{u}}_i$, we find $\bar{\mathbf{u}}$ reduces to zero and the linearization estimator of $\mathrm{var}\{vech[S_w]\}$ is:

$$\mathrm{v}_L\{vech(S_w)\} = \frac{n}{(n-1)\left(\sum_{i=1}^{n} w_i\right)^2} \sum_{i=1}^{n} w_i^2 \left(\hat{\varepsilon}_{it} \hat{\varepsilon}_{it'} - S_{w\,tt'} \right) \left(\hat{\varepsilon}_{it''} \hat{\varepsilon}_{it'''} - S_{w\,t''t'''} \right), \qquad (20)$$

corresponding to the estimator proposed by Browne (1984) when the sampling weights are constant.

Replacing U by $\mathrm{v}_L\{vech(S_w)\}$ in (8) gives a fitting function and a point estimator which we denote $F_{GLS-L}(S, \Sigma)$ and $\hat{\boldsymbol{\theta}}_{GLS-L}$ respectively. In the classical setting of independent and identically distributed observations the latter estimator is usually referred to as the ADF estimator. The estimator may allow for the complex design both through weighting in S_w and through the choice of linearization variance estimator $\mathrm{v}_L\{vech(S_w)\}$.

We now turn to the estimation of the variance of GLS estimators of $\boldsymbol{\theta}$. Assuming (16) and using linearization again (Skinner and Holmes, 2003), the asymptotic variance of the GLS estimator based upon the fitting function in (8) with a specified matrix U is:

$$\mathrm{var}(\hat{\boldsymbol{\theta}}) = n^{-1} \left(\Delta' U^{-1} \Delta \right)^{-1} \Delta' U^{-1} C U^{-1} \Delta \left(\Delta' U^{-1} \Delta \right)^{-1}, \qquad (21)$$

where $\Delta = \dfrac{\partial\{vech[\Sigma(\boldsymbol{\theta})]\}}{\partial\boldsymbol{\theta}}$.

The linearization estimator of this variance is then obtained by replacing Δ in (21) by $\hat{\Delta}$, defined as Δ evaluated at $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}$, and by replacing $n^{-1}C$ by a variance estimator $\mathrm{v}_L\{vech(S_w)\}$ as discussed above. When there are no covariates, this approach corresponds to estimation methods proposed by Skinner (1989), Satorra (1992), Muthén and Satorra (1995) and Skinner and Holmes (2003).

If U is chosen to be consistent for $n^{-1}C$, expression (21) reduces in the limit to:

$$\mathrm{var}(\hat{\boldsymbol{\theta}}) = n^{-1}\left(\Delta' U^{-1}\Delta\right)^{-1} . \tag{22}$$

Let us now consider estimation of the asymptotic covariance matrix of the PML point estimator $\hat{\boldsymbol{\theta}}_{PML}$. Following Binder (1983), we may write this asymptotic covariance matrix as:

$$\mathrm{var}(\hat{\boldsymbol{\theta}}_{PML}) = [I(\boldsymbol{\theta})]^{-1}\,\mathrm{var}[\boldsymbol{\varphi}(\boldsymbol{\theta})][I(\boldsymbol{\theta})]^{-1}, \tag{23}$$

where $\boldsymbol{\varphi}(\boldsymbol{\theta})$ is the $b\times 1$ pseudo-score function with j$^{\text{th}}$ element given by:

$$\phi_{\mathrm{j}}(\boldsymbol{\theta}) = \frac{\partial F_{PML}(\boldsymbol{\theta})}{\partial\theta_{\mathrm{j}}} = tr\left\{\Sigma(\boldsymbol{\theta})^{-1}[\Sigma(\boldsymbol{\theta})-S_w]\Sigma(\boldsymbol{\theta})^{-1}\frac{\partial\Sigma(\boldsymbol{\theta})}{\partial\theta_{\mathrm{j}}}\right\}, \tag{24}$$

using (14), and $I(\boldsymbol{\theta})$ is the $b\times b$ pseudo information matrix $I(\boldsymbol{\theta}) = -\partial\boldsymbol{\varphi}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. To estimate the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_{PML}$ it is therefore necessary to estimate the covariance matrix of $\boldsymbol{\varphi}(\boldsymbol{\theta})$. We may write:

$$\varphi_{\mathrm{j}}(\boldsymbol{\theta}) = tr\left[\Sigma(\boldsymbol{\theta})^{-1}\frac{\partial\Sigma(\boldsymbol{\theta})}{\partial\theta_{\mathrm{j}}}\right] + \frac{\sum_{i=1}^{n} w_i z_{ij}}{\sum_{i=1}^{n} w_i}, \tag{25}$$

where $z_{\mathrm{j}i} = -\boldsymbol{\varepsilon}_i'\Sigma(\boldsymbol{\theta})^{-1}\dfrac{\partial\Sigma(\boldsymbol{\theta})}{\partial\theta_{\mathrm{j}}}\Sigma(\boldsymbol{\theta})^{-1}\boldsymbol{\varepsilon}_i$. (26)

Linearizing the ratio in (25) gives:

$$\phi_{\mathrm{j}}(\boldsymbol{\theta}) = tr\left[\Sigma(\boldsymbol{\theta})^{-1}\frac{\partial\Sigma(\boldsymbol{\theta})}{\partial\theta_{\mathrm{j}}}\right] + \frac{\mu_{a\,\mathrm{j}}}{\mu_w} + \frac{1}{n}\sum_{i=1}^{n}\left(a_{\mathrm{j}i} - \frac{\mu_{a\,\mathrm{j}}}{\mu_w}\right)\frac{1}{\mu_w}$$

where $a_{ji} = w_i z_{ji}$, $\mu_{aj} = E(\bar{a}_j)$ and $\bar{a}_j = n^{-1}\sum_{i=1}^{n} a_{ji}$.

The covariance matrix of $\boldsymbol{\varphi}(\boldsymbol{\theta})$ may thus be approximated by

$$\text{var}\{\boldsymbol{\varphi}(\boldsymbol{\theta})\} \doteq \text{var}\left(n^{-1}\sum_{i=1}^{n} \mathbf{u}_i \right),$$

where $\mathbf{u}_i$ is the $b \times 1$ vector with j[th] element given by:

$$\frac{1}{\mu_w} \cdot \left(a_{ji} - \frac{\mu_{aj}}{\mu_w} \right). \tag{27}$$

This covariance matrix may be estimated for a complex design as above, for example using (19), where $\mathbf{u}_i$ is, as above, replaced by $\hat{\mathbf{u}}_i$, which is obtained by replacing $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\varepsilon}_i$ by $\hat{\boldsymbol{\varepsilon}}_i$ in (26) to give $\hat{z}_{ij}$, setting $\hat{a}_{ji} = w_i \hat{z}_{ji}$ and replacing a_{ji}, μ_{aj} and μ_w in (27) by $\hat{a}_{ji}$, $n^{-1}\sum_{i=1}^{n} \hat{a}_{ji}$ and $\bar{w}$ respectively. The linearization estimator of the variance of $\hat{\boldsymbol{\theta}}_{PML}$ is then obtained from (23) by replacing $\text{var}[\boldsymbol{\varphi}(\boldsymbol{\theta})]$ by this estimator and by replacing $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$ in $I(\boldsymbol{\theta})$.

Notice that the evaluation of the information matrix $I(\boldsymbol{\theta})$ requires differentiating $F_{PML}(\boldsymbol{\theta})$ and hence $\Sigma(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ twice. Some simplification is achieved by assuming that the model is correct, i.e. that $E[S_w] = \Sigma(\boldsymbol{\theta})$. If we then replace the information matrix in (23) by

$$\tilde{I}(\boldsymbol{\theta}) = E\left[-\frac{\partial \boldsymbol{\varphi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right],$$

which is asymptotically equivalent, we find from (24) that the jk[th] element of $\tilde{I}(\boldsymbol{\theta})$ may be expressed as:

$$\tilde{I}(\boldsymbol{\theta})_{jk} = tr\left[\Sigma(\boldsymbol{\theta})^{-1} \frac{\partial \Sigma(\boldsymbol{\theta})}{\partial \theta_j} \Sigma(\boldsymbol{\theta})^{-1} \frac{\partial \Sigma(\boldsymbol{\theta})}{\partial \theta_k} \right],$$

and we only need to differentiate $\Sigma(\boldsymbol{\theta})$ once.

6. Simulation with British Household Panel Survey data

In this section we shall assess the properties of the point and variance estimation procedures of sections 3 and 4 using a simulation study. In order to simulate from a realistic model, we shall base our study upon a regression analysis undertaken by Berrington (2002), with individual women as units of primary analytic interest and a measure of attitude to gender roles as the outcome variable, *y*.

The data come from waves 1, 3, 5, 7 and 9 (collected biannually between 1991 and 1999) of the British Household Panel Survey (BHPS) and these waves will be coded $t = 1,...,T = 5$ respectively. Respondents were asked whether they 'strongly agreed', 'agreed', 'neither agreed nor disagreed', 'disagreed' or 'strongly disagreed' with a series of statements concerning the family, women's roles, and work out of the household. Responses were scored from 1 to 5. Factor analysis was used to assess which statements could be combined into a gender role attitude measure. The attitude score, y_{it}, considered here is the total score for six selected statements for woman *i* at wave *t*. Higher scores signify more egalitarian gender role attitudes. Covariates for the regression analysis were selected on the basis of discussion in Berrington (2002) and include economic activity, which distinguishes in particular between women who are at home looking after children (denoted 'family care') and women following other forms of activity in relation to the labour market. Variables reflecting age and education are also included since these have often been found to be strongly related to gender role attitudes (e.g. Fan and Marini, 2000). All these covariates may change values between waves. A year variable (scored 1, 3, …, 9) is also included. This may reflect both historical change and the general ageing of the women in the sample.

The BHPS is a household panel survey of individuals in private domiciles in Great Britain (Taylor *et al.*, 2001). The sample was selected by a stratified multistage design, with individuals being selected with (approximately) equal inclusion probabilities. Given the interest in whether women's primary labour market activity is 'caring for a family', our study population is defined as women aged 16-39 in 1991. This results in a subset of data on *n* = 1340 women. This subset

consists of the longitudinal sample of women in the eligible age range for whom full interview outcomes were obtained in all five waves.

The simulation study consisted of simulating *D* replicate samples. Two approaches to generating the replicate samples were considered. The first involved both drawing a sample and generating y_{it} from a specified model, independently for each replicate. The second only involved the latter part, i.e. generating y_{it}. Since we found the results virtually identical for the two approaches we only report the results for the second. For simplicity, we ignored stratification and survey weights. We considered only two sampling schemes: simple random sampling of individuals and two-stage sampling, consisting of simple random sampling of m^{sim} primary sampling units (PSUs), followed by simple random sampling of n_j^{sim} individuals within each sampled PSU *j*. The population PSUs were defined to be 47 geographically contiguous clusters, formed by aggregating the original PSUs which consisted of 248 postcode sectors. The 1340 women were spread fairly evenly across these 47 PSUs. This aggregation was undertaken to strengthen the potential impact of clustering for the methodological purposes of this study, as in Skinner and Vieira (2007).

In the first approach, the values $\mathbf{x}_{it}$ for the 1340 women were held fixed and subsamples of specified size $n^{sim} = 100, 200, 500$ were drawn from these 1340 women according to the sampling scheme. The y_{it} were then simulated from specified models, independently for each replicate given these $\mathbf{x}_{it}$ values. The distribution across replicate samples may then be interpreted as joint with respect to both the sampling design and the model (for y_{it} conditional on $\mathbf{x}_{it}$). In the second approach a single sample was drawn in the same way, but then retained across all replicates. The distribution across replicate samples may then be interpreted as being with respect only to the model. The fact that the two approaches gave virtually the same results appears to be because the $\mathrm{x}_{it}\beta$ term in the model is assumed to be correctly specified and therefore the choice of sample has little impact on the distribution (with respect to the model) of S_w and hence of $\hat{\theta}$ for the sample sizes considered.

The y_{it} were generated from either the UCM model or the UCM(C) model from Section 2, with parameters set at the values obtained from fitting these models to the BHPS subset and errors following either the normal distribution or a *t* distribution. When simulating from the UCM(C) model, the clusters consisted of the actual 47 PSUs above so that the clustering displayed by the $\mathbf{x}_{it}$ values corresponded to that in the actual BHPS data, whereas the clustering in the y_{it} values was generated from the (fitted) UCM(C) model.

The implementation of the estimation procedures in sections 3 and 4 generally required iterative numerical methods, although explicit expressions for computation could be obtained in some special cases. The numerical minimisation of the fitting functions or maximisation of the pseudo likelihood was generally achieved through the numerical solution of equations obtained by differentiating the fitting functions. Several alternative methods for performing the numerical solution were considered. We eventually adopted an iterative Newton type algorithm, similar to that suggested by Pourahmadi (1999), and available in the function *nlm* of the statistical computer software *R* (R Development Core Team, 2003). The use of several other alternatives for performing the necessary numerical minimizations was also considered, but their performance was either the same or worse than the Newton type algorithm. In a small number of cases for the $\hat{\boldsymbol{\theta}}_{GLS-L}$ and $\hat{\boldsymbol{\theta}}_{GLS-NORM2}$ estimators, the iterative algorithms failed to converge. The non-convergence rates for the $\hat{\boldsymbol{\theta}}_{GLS-L}$ estimator varied across simulation set-ups between 0.1% and 1.0% of the replicate samples, while these rates varied from 0.1% to 0.3% for the $\hat{\boldsymbol{\theta}}_{GLS-NORM2}$ estimator. For all the remaining estimation methods convergence was always achieved. The cases of non-convergence are omitted from the tables presented below.

6.1 Point estimators

In this subsection, we aim to present results based on $D = 1000$ replicate samples, derived as set out above. Five point estimators were considered: ULS, GLS-NORM1, GLS-NORM2 and PML,

defined in (7), (12), (13) and (15) respectively, and GLS-L, defined by (8) with U given by the estimator in (20). It was in fact found that the ULS and PML estimation methods produced virtually identical results for the UCM model and similar results for other models, a finding corresponding to that of Bollen (1989, p. 112). We therefore do not present the ULS results and focus instead on the remaining four estimators, assessing their properties in terms of relative bias and coefficient of variation (*cv*), estimated from across the replicate samples.

Table 1 presents results produced when the UCM model with normal errors is used both to generate the y_{it} values and as a basis for model fitting. The parameter vector $\boldsymbol{\theta} = \left(\sigma_u^2, \sigma_v^2\right)'$ contains two parameters of interest. In this case, we might expect the estimators $\hat{\boldsymbol{\theta}}_{GLS-NORM1}$, $\hat{\boldsymbol{\theta}}_{GLS-NORM2}$ and $\hat{\boldsymbol{\theta}}_{PML}$ which exploit the normality to outperform the estimator $\hat{\boldsymbol{\theta}}_{GLS-L}$ which does not. In fact we observe little difference between the performance of this estimator and that of $\hat{\boldsymbol{\theta}}_{GLS-NORM1}$. We do observe that $\hat{\boldsymbol{\theta}}_{GLS-NORM2}$ performs consistently better than $\hat{\boldsymbol{\theta}}_{GLS-NORM1}$ (though sometimes only slightly) with respect to relative bias and to a lesser extent with respect to coefficient of variation. The estimator $\hat{\boldsymbol{\theta}}_{PML}$ has a similar performance to $\hat{\boldsymbol{\theta}}_{GLS-NORM2}$ with respect to coefficient of variation but displays different patterns of relative bias, being worse for σ_u^2 but slightly better for σ_v^2 . We repeated the simulation in Table 1 using the AR1 model and found similar results, which are not reported here.

In Table 2, we consider the impact of clustering, with the data now generated from the UCM-C model. The UCM model continues to be the fitted model. We considered both normal and t-distributed errors and present the results for t-distributed errors in Table 2. We expected the main difference between Table 2 and Table 1 to be an increase in *cv* from the clustering, but we also noticed an appreciable if not entirely consistent increase in relative bias. We again find that $\hat{\boldsymbol{\theta}}_{GLS-NORM2}$ performs consistently better than $\hat{\boldsymbol{\theta}}_{GLS-NORM1}$ with respect to relative bias, but this is now not necessarily the case with respect to *cv*. As the sample sizes increase, we note that again

$\hat{\boldsymbol{\theta}}_{GLS-NORM\,2}$ and $\hat{\boldsymbol{\theta}}_{PML}$ appear to be the preferred methods with respect to relative bias. In particular, $\hat{\boldsymbol{\theta}}_{PML}$ performs especially well for the relative bias of $\hat{\sigma}_{\mathrm{v}}^2$. There does not appear to be a great difference between all four methods with respect to *cv*, but there was a slight tendency for $\hat{\boldsymbol{\theta}}_{GLS-NORM\,2}$ to be outperformed by the other three methods. Simulation results produced for AR1 model fitting in the current situation, which are not presented again, generally agreed with results presented in Table 2.

We focus on the impact of clustering in Table 3, where the inflation of mean squared error (MSE) arising from the incorporation of cluster effects in the data generation process is considered, in the case when $n^{sim} = 100$ and the errors are t distributed. There are no major differences between the estimation methods in terms of the MSE inflation, although the impact appears to be least for the GLS-L method.

Overall, these simulation results produced for the ADF method GLS-L generally agree with the covariance structure modelling literature (e.g. Bollen, 1989, p. 432; Satorra, 1992), where it is recommended that those methods should be adopted only in situations with large sample sizes (1000 or more), for dealing with situations where departures from normality conditions are evident. We may emphasize that ADF methods have in several situations had good general performance, even though these methods have not shown 'good' levels of bias. PML point estimators have in general produced very good performance in terms of bias and variance, particularly the former. The good performance of PML is particularly marked for the relative bias of $\hat{\sigma}_{\mathrm{v}}^2$.

6.2 Variance estimators

We now consider the properties of the linearization variance estimators denoted v_L in section 4. We restrict attention to their use in the estimation of the variance of the two point estimators: $\hat{\boldsymbol{\theta}}_{GLS-NORM\,1}$ and $\hat{\boldsymbol{\theta}}_{PML}$. To provide benchmarks for comparison, we also consider the variance estimator, $\mathrm{var}_{\mathrm{n}}(.)$, which is based upon the assumption of both normality and independent and identically distributed

observations, and the estimator $\text{var}_{\text{df}}(.)$ which allows for non-normality but still assumes independent and identically distributed observations. The subscript n denotes naïve. In the case of $\hat{\boldsymbol{\theta}}_{GLS-NORM1}$, $\text{var}_{\text{n}}(.)$ and $\text{var}_{\text{df}}(.)$ are obtained from (22) and (21) respectively, with U given by (10) and $W = S_w$. In the case of $\hat{\boldsymbol{\theta}}_{PML}$, $\text{var}_{\text{n}}(.)$ is given by $\left[I(\boldsymbol{\theta})\right]^{-1}$.

To evaluate the properties of these variance estimators, the replicate samples were obtained from two-stage sampling, as described earlier. The number of sampled PSUs, m^{sim}, was set to be 15, 20 or 47. The number of individuals sampled in the j^{th} selected PSU is denoted n_j^{sim}. The UCM-C model was used to generate the values of y_{ijt} now using $D = 10{,}000$ replicates. The parameters of the UCM-C model were the same as in the simulations in section 5.1. , except that there were some different choices for σ_η^2: $\sigma_\eta^{2\ sim,\text{C}} \cong 0.15$, $\sigma_\eta^{2\ sim,\text{C}} \cong 0.45$, and $\sigma_\eta^{2\ sim,\text{C}} \cong 0.75$; to enable the evaluation of effects of different impacts of clustering on the variance estimation procedures. The fitted model was taken as the UCM model.

Table 4 displays results produced when considering $m^{sim} = 47$ and $n_j^{sim} = 15$. The first three variance estimators do not take the clustering into account and, as anticipated, clearly underestimate the variance. The degree of underestimation increases with σ_η^2, i.e. the more clustering the more downward relative bias.

Both methods that allow for clustering have improved properties in terms of relative bias, compared to the first three methods in Table 4. They still tend to be biased downwards, however, corresponding to other findings for linearization variance estimation (Wolter, 2007, Chapter 8; Kott, 1991). Furthermore, these two methods had larger variances than the first three methods, as expected as a result of the reduced degrees of freedom for variance estimation.

Table 5 includes results that were produced when considering $m^{sim} = 20$ and $n_j^{sim} = 15$, i.e. 300 cases. Under this situation, the linearization variance estimators which allow for the complex sampling again led to noticeable improvements in terms of relative bias when compared to methods that ignored the sampling scheme. The smaller number of sample clusters does, however, seem to

have led to some increases in relative bias, although these are still smaller than the *cv*s. Neither the relative bias nor the *cv* were found to vary greatly with σ_η^2.

Table 6 includes results that were produced when $m^{sim} = 15$ and $n_j^{sim} = 10$, i.e the number of SSUs selected per cluster was further reduced, and the sample size was diminished to 150. Further increases in relative bias were observed although again the relative biases were smaller than the *cv*s. As in Table 5 there was no strong relationship between either the relative bias or the *cv* with σ_η^2.

In summary, the linearization method which allows for clustering appears to perform reasonably well for both the PML and the GLS-NORM1 point estimators for a range of possible clustering effects, although there is a tendency for the variance to be seriously underestimated if the number of sampled clusters is small, say twenty or below.

7. Conclusion

This paper has proposed some methods for making inference about parameters in panel data models, allowing for complex sampling schemes. Methods have been evaluated using a simulation study based upon data from the British Household Panel Survey. The study indicated that: (*i*) overall, most of the proposed point estimation methods perform satisfactorily; (*ii*) the asymptotically distribution free point estimator performed reasonably but did not show significant improvements on the other methods and did occasionally suffer from lack of convergence (*iii*) pseudo maximum likelihood (PML) estimators produced satisfactory performance in terms of bias and variance, even when the normality assumption was violated.

Linearization methods for variance estimation for GLS and PML point estimators were considered. The results of the simulation study suggested that: (*iv*) methods that do not take the sampling scheme into account underestimate the variance, in some situations very gravely; (*v*) underestimation tends to increase rapidly with inflation in the impacts of clustering; (*vi*) the linearization estimator of the variance of the PML point estimator has an evidently better performance in terms of bias than the linearization estimator of the variance of the GLS estimator.

Overall, the most satisfactory results in the simulation study were obtained from the combination of the PML point estimator (defined via expression (15)) and the associated linearization variance estimator (defined below expression (23)). The advantages of this combined approach were that: computation did not lead to problems of convergence; the point estimator had good relative performance in terms of both bias and variance, particularly the former; the bias performance of this variance estimator was more favourable than that for the GLS estimator.

References

Berrington, A. (2002) Exploring Relationships between Entry into Parenthood and Gender Role Attitudes: Evidence from the British Household Panel Study. In Lesthaeghe, R. ed *Meaning and Choice: Value Orientations and Life Course Decisions*. Brussels, NIDI.

Binder, D. A. (1983) On the Variances of Asymptotically Normal Estimators from Complex Surveys. *International Statistical Review*, 51, 279-292.

Bollen, K. A. (1989) *Structural Equations with Latent Variables*. New York, John Wiley & Sons.

Browne, M. W. (1982) Covariance Structures. In Hawkins, D. M. eds. *Topics in Applied Multivariate Analysis*. Cambridge, Cambridge University Press.

Browne, M. W. (1984) Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62-83.

Chambers, R. L. and Skinner, C. J. eds. (2003) *Analysis of Survey Data*. Chichester: Wiley.

Diggle, P. J., Heagerty, P., Liang, K. & Zeger, S. L. (2002) *Analysis of Longitudinal Data*. 2nd ed. Oxford, Oxford University Press.

Fan, P. -L. and Marini, M. M. (2000) Influences on Gender-role Attitudes during the Transition to Adulthood. *Social Science Research*, 29, 258-283.

Feder, M., Nathan, G. and Pfeffermann, D. (2000) Multilevel Modelling of Complex Survey Longitudinal Data with Time Varying Random Effects. *Survey Methodology*, 26, 53-65.

Fuller, W. A. (1975) Regression Analysis for Sample Surveys. *Sankhya*. 37, Series C, 117-132.

Fuller, W. A. (1987) *Measurement Error Models*. New York: Wiley.

Isaki, C. T. and Fuller, W. A. (1982) Survey Design under the Regression Superpopulation Model. *Journal of the American Statistical Association,* 77, 89-96.

Jöreskog, K. G. and Goldberger, A. S. (1972) Factor Analysis by Generalized Least Squares. *Psychometrika*, 37, 243-260.

Kalton, G. and Brick, M. (2000) Weighting in household panel surveys. In Rose, D. ed. *Researching Social and Economic Change: the Uses of Household Panel Studies*. London, Routledge.

Kott, P. S. (1991) A Model-Based Look at Linear Regression with Survey Data. *The American Statistician*, 45, 107-112.

Liang, K. and Zeger, S. L. (1986) Longitudinal Data Analysis Using Generalised Linear Models. *Biometrika*, 73, 13-22.

Muthén, B. O. and Satorra, A. (1995) Complex Sample Data in Structural Equation Modelling. *Sociological Methodology*, 25, 267-316.

Pfefferman, D., Skinner, C. J., Holmes, D. J., Goldstein, H., & Rashash, J. (1998). Weighting for Unequal Selection Probabilities in Multilevel Models. *Journal of the Royal Statistical Society*, *Series B*, 60, 23–40.

Pourahmadi, M. (1999) Joint Mean-covariance with Applications to Longitudinal Data: Unconstrained Parameterisation. *Biometrika*, Vol. 86, n. 3, 677-690.

R Development Team (2003) *The R Environment for Statistical Computing and Graphics – Reference Index*, Version 1.8.1, The R Foundation for Statistical Computing.

Satorra, A. (1992) Asymptotic Robust Inferences in the Analysis of Mean and Covariance Structures. *Sociological Methodology*, Vol. 22, 249-278.

Shah, B. V., Barnwell, B. G., and Bieler, G. S. (1997) *SUDAAN User' Manual, Release 7.5*. vol. 1 e 2, Research Triangle Park, NC: Research Triangle Institute.

Shah, B. V., Folsom, R. E., LaVange, L. M., Wheeless, S. C., Boyle, K. E., and Williams, R. L. (1995) *Statistical Methods and Mathematical Algorithms Used in SUDAAN*. Research Triangle Park, NC: Research Triangle Institute.

Skinner, C. J. (1989) Domain Means, Regression and Multivariate Analysis. In Skinner, C. J., Holt, D. and Smith, T. M. F. eds. *Analysis of Complex Surveys*. Chichester: Wiley.

Skinner, C. J. and Holmes, D. (2003) Random Effects Models for Longitudinal Survey Data. In Chambers, R. L. and Skinner, C. J. eds. *Analysis of Survey Data*. Chichester: Wiley.

Skinner, C. J., Holt, D. and Smith, T. M. F. eds. (1989) *Analysis of Complex Surveys*. Chichester: Wiley.

Skinner, C. J. and Vieira, M. D. T. (2007) Variance Estimation in the Analysis of Clustered Longitudinal Survey Data. *Survey Methodology* , 33, 3-12.

Sutradhar, B. C. and Kovacevic, M. (2000) Analysing ordinal longitudinal survey data: Generalised estimating equations approach. *Biometrika*, 87, 837-848.

Taylor, M. F. ed, with Brice J., Buck, N. and Prentice-Lane E. (2001) *British Household Panel Survey - User Manual - Volume A: Introduction, Technical Report and Appendices*. Colchester, University of Essex.

Wolter, K. M. (2007) *Introduction to Variance Estimation*. 2nd Ed. New York, Springer.

Wooldridge, J. M. (2001) *Econometric Analysis of Cross Section and Panel Data*. 1st ed. Cambridge, MA, MIT Press.

Estimator		$n=100$		$n=200$		$n=500$		$n=1340$	
		rel bias	*cv*	rel bias	*cv*	rel bias	*cv*	rel bias	*cv*
$\hat{\theta}_{GLS-NORM1}$	$\hat{\sigma}_u^2$	-16.76%	17.77%	-9.21%	12.14%	-3.40%	7.16%	-1.42%	4.29%
	$\hat{\sigma}_v^2$	-9.70%	8.41%	-4.68%	5.56%	-1.74%	3.39%	-0.74%	1.90%
$\hat{\theta}_{GLS-NORM2}$	$\hat{\sigma}_u^2$	-6.43%	17.69%	-3.77%	11.77%	-1.18%	7.12%	-0.60%	4.27%
	$\hat{\sigma}_v^2$	6.41%	7.19%	3.51%	5.20%	1.59%	3.27%	0.47%	1.88%
$\hat{\theta}_{GLS-L}$	$\hat{\sigma}_u^2$	-15.79%	19.44%	-9.23%	12.76%	-3.41%	7.19%	-1.46%	4.33%
	$\hat{\sigma}_v^2$	-9.89%	9.04%	-4.60%	5.83%	-1.72%	3.44%	-0.74%	1.93%
$\hat{\theta}_{PML}$	$\hat{\sigma}_u^2$	-9.94%	17.18%	-5.61%	11.68%	-1.92%	7.08%	-0.88%	4.26%
	$\hat{\sigma}_v^2$	0.89%	6.84%	0.74%	5.09%	0.47%	3.25%	0.06%	1.87%

Table 1 – Properties of point estimators when both fitted model and true model are UCM.

Estimator		$n=100$		$n=200$		$n=500$		$n=1340$	
		rel bias	*cv*	rel bias	*cv*	rel bias	*cv*	rel bias	*cv*
$\hat{\theta}_{GLS-NORM1}$	$\hat{\sigma}_u^2$	-16.73%	29.27%	-8.75%	22.07%	-4.05%	12.10%	-1.63%	7.54%
	$\hat{\sigma}_v^2$	-12.30%	10.98%	-7.13%	8.08%	-2.65%	5.23%	-1.02%	3.28%
$\hat{\theta}_{GLS-NORM2}$	$\hat{\sigma}_u^2$	-7.11%	29.26%	-3.32%	22.28%	-1.78%	12.17%	-0.76%	7.53%
	$\hat{\sigma}_v^2$	9.45%	14.00%	4.83%	9.92%	2.18%	6.08%	0.92%	3.66%
$\hat{\theta}_{GLS-L}$	$\hat{\sigma}_u^2$	-21.82%	29.11%	-13.00%	18.55%	-6.16%	11.72%	-2.56%	7.44%
	$\hat{\sigma}_v^2$	-17.18%	11.74%	-11.54%	8.23%	-5.58%	5.16%	-2.75%	3.21%
$\hat{\theta}_{PML}$	$\hat{\sigma}_u^2$	-10.33%	28.91%	-5.16%	22.00%	-2.54%	12.10%	-1.05%	7.53%
	$\hat{\sigma}_v^2$	1.56%	10.84%	0.62%	8.62%	0.51%	5.55%	0.26%	3.47%

Table 2 – Properties of point estimators when fitted model is UCM and true model is UCM-C with t distributed errors

Estimator		UCM model	AR1 model
$\hat{\theta}_{ULS}$	$\hat{\sigma}_u^2$	1.44	1.46
	$\hat{\sigma}_v^2$	0.89	0.93
	$\hat{\gamma}$	-	1.01
$\hat{\theta}_{GLS-NORM1}$	$\hat{\sigma}_u^2$	1.27	1.27
	$\hat{\sigma}_v^2$	0.93	0.92
	$\hat{\gamma}$	-	1.01
$\hat{\theta}_{GLS-NORM2}$	$\hat{\sigma}_u^2$	1.52	1.53
	$\hat{\sigma}_v^2$	0.95	1.06
	$\hat{\gamma}$	-	1.10
$\hat{\theta}_{GLS-L}$	$\hat{\sigma}_u^2$	1.22	1.23
	$\hat{\sigma}_v^2$	0.86	0.89
	$\hat{\gamma}$	-	0.82
$\hat{\theta}_{PML}$	$\hat{\sigma}_u^2$	1.44	1.45
	$\hat{\sigma}_v^2$	0.89	0.99
	$\hat{\gamma}$	-	1.04

Table 3 – Ratios of MSEs of estimators with data generated from UCM-C model (numerator) and from UCM model (denominator) (n=100 and t-distributed errors).

Variance Estimator		rel bias			$cv\left(\text{var}\left(\hat{\theta}\right)\right)$		
		$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$	$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$
$\text{var}_n\left(\hat{\theta}_{PML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-0.39%	-7.75%	-11.43%	14.07%	14.27%	14.54%
	$\text{var}(\hat{\sigma}_v^2)$	1.78%	-2.44%	-0.30%	8.54%	8.54%	8.59%
$\text{var}_n\left(\hat{\theta}_{GLS-NORM1}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-1.54%	-8.96%	-12.47%	10.71%	11.14%	11.37%
	$\text{var}(\hat{\sigma}_v^2)$	-5.18%	-10.25%	-7.14%	5.39%	5.54%	5.47%
$\text{var}_{df}\left(\hat{\theta}_{GLS-NORM1}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-1.51%	-9.07%	-12.60%	14.13%	14.34%	14.61%
	$\text{var}(\hat{\sigma}_v^2)$	-4.14%	-9.20%	-6.01%	8.62%	8.70%	8.69%
$v_L\left(\hat{\theta}_{PML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	0.27%	-4.58%	-3.55%	24.65%	25.41%	26.85%
	$\text{var}(\hat{\sigma}_v^2)$	2.53%	-2.35%	0.99%	22.01%	21.86%	21.98%
$\text{v}_L\left(\hat{\theta}_{GLS-NORM1}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-0.85%	-6.02%	-4.91%	24.78%	25.51%	27.00%
	$\text{var}(\hat{\sigma}_v^2)$	-3.48%	-9.13%	-4.80%	22.33%	22.24%	22.43%

Table 4 – Properties of variance estimators, when UCM is fitted model, UCM-C is true model, $m^{sim} = 47$ and $n_j^{sim} = 15$.

Variance Estimator		rel bias			$cv\left(\text{var}\left(\hat{\theta}\right)\right)$		
		$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$	$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$
$\text{v}_L\left(\hat{\theta}_{PML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-5.17%	-5.25%	-4.69%	38.07%	39.03%	40.75%
	$\text{var}(\hat{\sigma}_v^2)$	-1.54%	-0.69%	-0.49%	33.55%	33.79%	34.44%
$\text{v}_L\left(\hat{\theta}_{GLS-NORM1}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-7.31%	-7.60%	-6.55%	38.42%	39.17%	40.83%
	$\text{var}(\hat{\sigma}_v^2)$	-14.17%	-12.87%	-12.23%	34.26%	34.39%	35.00%

Table 5 – Properties of variance estimators, when UCM is fitted model, UCM-C is true model, $m^{sim} = 20$ and $n_j^{sim} = 15$.

Variance Estimator		rel bias			$cv\left(\text{var}\left(\hat{\theta}\right)\right)$		
		$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$	$\sigma_\eta^2 = 0.15$	$\sigma_\eta^2 = 0.45$	$\sigma_\eta^2 = 0.75$
$\text{v}_L\left(\hat{\theta}_{PML}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-5.48%	-6.11%	-4.87%	47.86%	47.80%	50.19%
	$\text{var}(\hat{\sigma}_v^2)$	-3.41%	-2.68%	-1.38%	41.05%	40.43%	40.87%
$\text{v}_L\left(\hat{\theta}_{GLS-NORM1}\right)$	$\text{var}(\hat{\sigma}_u^2)$	-9.26%	-9.63%	-8.64%	48.57%	48.09%	50.85%
	$\text{var}(\hat{\sigma}_v^2)$	-23.34%	-24.21%	-21.92%	42.07%	41.22%	41.86%

Table 6 – Properties of variance estimators, when UCM is fitted model, UCM-C is true model, $m^{sim} = 15$ and $n_j^{sim} = 10$.